The Technical University of Łódź

Faculty of Electrical, Electronic, Computer and Control Engineering

Engineering Thesis

A novel method of mental stress assessment based on heart rate signal analysis

Sławomir Kosowski

Student's number: 143377

Supervised by

prof. Jean-Yves Fourniols

Pawel Strumillo, MSc, PhD, DSc

Łódź, 2011/02/08

TECHNICAL UNIVERSITY OF LODZ

FACULTY OF ELECTRICAL, ELECTRONIC, COMPUTER AND CONTROL ENGINEERING

Sławomir Kosowski

BSc THESIS

"A novel method of mental stress assessment based on heart rate signal analysis"

Lodz, 2011

Supervisor: prof. Jean-Yves Fourniols

Auxiliary supervisor: Pawel Strumillo, MSc, PhD, DSc

Abstract

Research problem:
Mental stress is becoming a serious problem in nowadays developed societies. There does not exist any clinically proved simple method of mental stress assessment up to today. There are several factors that can be observed looking for the symptoms of mental stress. The non-invasive ones are: breath, blood pressure, sweat or ECG monitoring. These factors are more difficult to classify, than the invasive methods, such as cortisol monitoring. Advantages of the invasive methods are better accuracy and predictive values.
Heart Rate Variability (HRV) is a physiological phenomenon of variations of heart beat-to-beat time intervals. This is also known as RR variability, where R is a peak of QRS complex of the ECG signal. HRV in signal analysis is a set of methods that makes possible analyzing beat-to-beat time variations. Moreover HRV is one of the most promising markers of sympathovagal balance in humans. Despite the fact that this set of methods is vast, there are few that permit to retain the sequentiality of the signal in time. Moreover they are mathematically complex and give rather qualitative results than a quantitative value for mental stress assessment.
Methods:
A novel algorithm that bases on Gaussian curves representing the probability of occurrences of particular heart beat frequencies is proposed. Mathematical tools were derived that enable us to measure the Gaussian distributions quantitatively: **Rgauss** and **Wgauss**. Furthermore, a predictive value of more than 20 different algorithms, including neural networks, cluster analysis, Partial Least Squares and others were also apprised. Finally, the algorithm based on the Gaussian curves analysis was chosen as the one, which has potentially the largest predictive value.
Results:
The results gave us partitioning of patients into 4 general groups that can be easily distinguished. The parameters defining these groups vary between each other signi-

ficantly. Moreover there was no correlation found between Gaussian-based algorithm and any other.
At this moment this method is consulted with cardiologists from one of the main cardiology clinic in France.
Finally, in order to easily process immense number of data-files, two graphical toolboxes in MATLAB® environment were developed. These toolboxes are available freely on the Internet for personal use. This will be an invitation for other cardiologists and researchers to test this method in practice and establish threshold values of the parameters proposed for mental stress assessment.

POLITECHNIKA ŁÓDZKA

WYDZIAŁ ELEKTROTECHNIKI, ELEKTRONIKI, INFORMATYKI I AUTOMATYKI

Sławomir Kosowski

PRACA DYPLOMOWA INŻYNIERSKA

“Nowy algorytm oceny stresu przewlekłego na podstawie zmienności rytmu serca”

Łódź, 2011 r.

Opiekun: prof. Jean-Yves Fourniols

Opiekun dodatkowy: dr hab. inż. Paweł Strumiłło

Streszczenie

Problem badawczy:
Stres przewlekły staje się coraz poważniejszym problemem w społeczeństwach rozwiniętych. Obecnie nie istnieje żadna klinicznie potwierdzona, prosta metoda oceny stresu przewlekłego. Istnieje kilka czynników które mogą podlegać ocenie. Metody nieinwazyjne to badanie na przykład: oddechu, ciśnienia krwi, potu oraz EKG. Zależność między tymi czynnikami, a stresem przewlekłym jest trudniejsza do oceny, niż w przypadku metod inwazyjnych, takich jak pomiar hormonu stresu - kortyzolu. Zaletą metod inwazyjnych jest lepsza dokładność oraz przewidywalność oceny.
Zmienność rytmu serca (ang. Heart Rate Variability, HRV) jest zjawiskiem fizjologicznym wyrażającym się w zmianach przedziałów czasowych kolejnych uderzeń serca. Zmienność ta nazywana jest zmiennością RR, gdzie R jest jednym z załamków zespołu QRS w zapisie EKG. HRV w analizie sygnałów jest zbiorem metod które umożliwiają badanie w/w zmienności. Ponadto HRV jest jednym z najbardziej obiecujących znaczników równowagi pomiędzy układem współczulnym i przywspółczulnym u człowieka. Pomimo że zbiór metod HRV jest duży, istnieje tylko kilka które pozwalają na zachowanie sekwencyjności sygnału w czasie. Co więcej, są one skomplikowane matematycznie i dają wyniki jakościowe a nie ilościowe dla oceny stresu.
Metody:
Zaproponowano nowy algorytm bazujący na krzywych Gaussa prezentujących prawdopodobieństwo wystąpień poszczególnych częstotliwości bicia serca. Ponadto, zaproponowano model matematyczny zmienności rytmu serca z zastosowaniem rozkładu wielogaussowskiego (parametry **Rgauss** oraz **Wgauss**). Ponadto, ponad 20 innych algorytmów, takich jak sieci neuronowe, analiza skupień, regresja PLS i inne zostało przebadanych pod kątem skuteczności oceny stresu przewlekłego. Ostatecznie algorytm bazujący na krzywych Gaussa zostaje wybrany jako ten, który uznano za najbardziej obiecujący.
Wyniki:

W wyniku zastosowania w/w algorytmu, otrzymaliśmy klasyfikację pacjentów w 4 głównych grupach. Parametry definiujące ten podział różnią się od siebie w sposób znaczący. Co więcej, nie znaleziono korelacji pomiędzy algorytmem bazującym na krzywych Gaussa i innymi badanymi algorytmami.
Obecnie, algorytm ten jest konsultowany z jedną z największych klinik kardiologicznych we Francji.
Celem przetwarzania dużej liczby plików z zapisami EKG, dwa toolboxy w środowisku MATLAB® zostały zaprojektowane. Są one udostępnione na stronie internetowej autora dla użytku osobistego, bez żadnych opłat. Jest to swoiste zaproszenie dla lekarzy, kardiologów, badaczy i innych do sprawdzenia tej metody w praktyce i zaproponowania konkretnych wartości do oceny stresu przewlekłego.

Table of contents

List of important abbreviations and symbols

HRV	Heart Rate Variability
HR	Heart Rate
$NNinterval$	time interval between two adjacent QRS complexes
$F_c(t)$	instantaneous cardiac frequency (as a function of time)
FM	Modal Frequency
AUC	Area Under Curve
$SDNN$	Standard Deviation of the NN interval
$SDANN$	Standard Deviation of the all 5-minute NN intervals
$RMSSD$	square root of the mean squared differences of successive NN intervals
$NN50$	number of successive NN intervals that differ more than 50ms
$pNN50$	ratio of NN50 and total number of NN intervals in the recording
$TINN$	Triangular Interpolation of NN intervals
LF/HF	ratio of Low and High cardiac Frequencies
AR	Auto Regressive (model)

1. Introduction

1.1. The research problem

Chronic mental stress is perceived as a risk factor of cardiovascular diseases [1]. Thus it can be considered as an element determining one's health and self-esteem. ECG signal is one of the possible non-invasive observations of a subject that contain information about mental stress. There are several more factors that could be observed seeking for stress prediction. Good examples are: breath monitoring, eye motion, sweat monitoring. Nevertheless ECG signal was chosen as a one that is the simplest to measure and can be evaluated quantitatively with great ease. ECG signal is treated as one of the most valuable markers of heart failure or disease. Moreover, measuring ECG signal is non-invasive and in consequence, there are no moral restrictions about sampling the data. No matter how many parameters of ECG signal we measure, up to today, there does not exist any clinically proved method of mental stress assessment based on ECG signal analysis. This poses The Research Problem.

1.2. Aim of the thesis

The aim of this thesis is to develop a novel method of ECG signal (precisely RR intervals) analysis, as stated in section 1.1. This method is supposed to be a mental stress predictor. Up to today there are many methods of the ECG signal analysis, but they do not give an insight into temporal evolution of the ECG signal. In most cases it implies loss of the information about the sequentiality of the signal. Thus it is impossible to correlate one's ECG and physiological response in terms of more broad analysis than in case of classical HRV, i.e. general sympathovagal balance. The reported work was conducted by the author during his internship in Laboratoire d'Analyse et d'Architecture des Systemès in Toulouse, France in 2010 and was supervised by Jean-Yves Fourniols.

2. ECG and HRV analysis

2.1. Introduction

Heart Rate Variability (HRV) is a physiological phenomenon of variations of heart beat-to-beat time intervals. This is also known as RR variability, where R is a peak of QRS complex (Fig. 2.2) of the ECG signal (Fig. 2.1). HRV in signal analysis is a set of methods measuring one's heart variability, standardized by American Task Force in 1996 [2]. It was introduced in '60s of the last century. Since this time the set of tools in HRV analysis has grown, as well as the clinical significance. Nowadays HRV is one of the most promising markers of sympathovagal balance [3]. HRV can be also used to detect many disorders and illnesses, which will be discussed further.
Let us consider a sample ECG signal as depicted in Fig. 2.1.

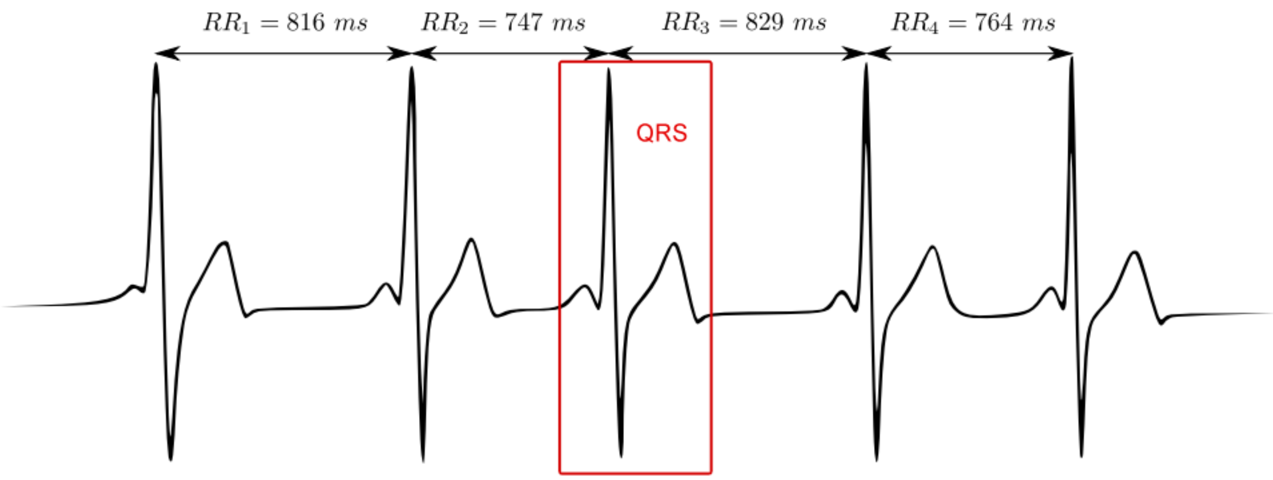


Fig. 2.1. Sample ECG signal [4]

In this signal we may distinguish so-called QRS complex, which in details is depicted in Fig. 2.2.

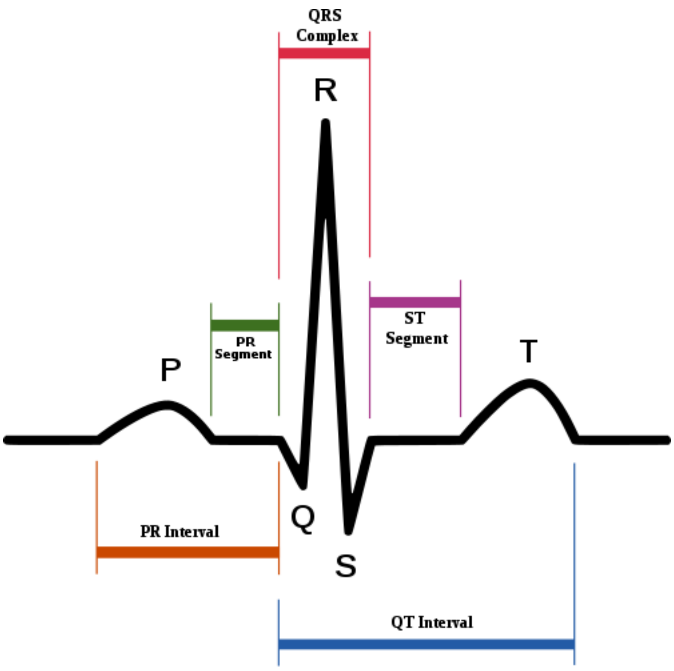


Fig. 2.2. Details of QRS complex

QRS complex corresponds to the depolarization of the right and left ventricles. Since it is usually the most visible part of the ECG trace, in this paper we consider only one input filetype in format RR interval expressed in ms. Where RR interval means time interval between two consecutive QRS complexes (strictly: two R-waves). In order to obtain such series of intervals from an ECG signal it is indispensable to perform several steps. The consecutive order of transformations from raw ECG signal to RR interval[1] is listed below:

- ECG signal acquisition
- Digitization by ADC
- Artefact detection
- QRS complex detection
- Calculation of intervals between consecutive R-waves

Once the measurements are given (usually in ASCII format) we can perform several types of analysis. In this paper, the standard HRV parameters were calculated using

[1] RR is also called normal-to-normal interval (NN interval); in practice it concerns time interval between two adjacent QRS complexes

Software for advanced HRV analysis developed in University of Kuopio, Finland by Juha-Pekka Niskanen, Mika P. Tarvainen, Perttu O. Ranta-aho, and Pasi A. Karjalainen [5].

2.2. Time domain methods

According to the standards [2] time domain parameters of HRV are typically calculated from 24h long ECG recording, and rather should not be applied to shorter time intervals.

2.2.1. Statistical methods

Two mutually dependent parameters directly obtained from NN intervals are mean Heart Rate (HR) expressed in pulsations/min. and mean NN expressed in ms.
SDNN - standard deviation of the NN interval which is simply square root of variance. The standard deviation is calculated as a mean of all 5-minute standard deviations(eq. 2.1) of NN interval during 24-h period.

$$s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{2.1}$$

It reflects all cyclic components responsible for variability in the entire recording. If SDNN is calculated over a 24h period it reflects short-term high frequency variations as well as low frequencies. It should be stressed that statistically the longer the recording is, the higher variance is obtained, and in consequence SDNN increases.
Also **SDANN** can be calculated. It is a standard deviation of all 5-minute NN interval means. It measures the variability over the intervals shorter than 5 minutes.
RMSSD is the square root of the mean squared differences of successive NN intervals.
NN50 is the number of successive NN intervals that differ more than 50 ms.
pNN50 is the ratio of NN50 and total number of NN intervals in the recording (eq. 2.2) expressed in [%].

$$pNN50 = \frac{NN50}{No\ samples} * 100[\%] \tag{2.2}$$

2.2.2. Geometrical methods

The explanation of these parameters is shown in fig. 2.3, where D is sample density distribution of NN intervals. The most frequent NN interval length is denoted by X. The maximum of this histogram is denoted by $Y = D(X)$. **HRV triangular index** is the value obtained by dividing the area integral of D by the maximum Y.

$$HRV\ index = (total\ number\ of\ all\ NN\ intervals)/Y \tag{2.3}$$

TINN - triangular interpolation of NN - is the baseline width of the NN interval distribution measured as a base of a triangle which approximate this distribution; namely: multilinear function q constructed such that $q(t) = 0$ for $t \leq N$ and $t \geq M$ and $q(X) = Y$. Moreover the integral $\int_0^\infty (D(t) - q(t))^2 dt$ is the minimum of all occurrences of N and M values. Final result is expressed in ms and is expressed by the formula 2.4.

$$TINN = M - N \tag{2.4}$$

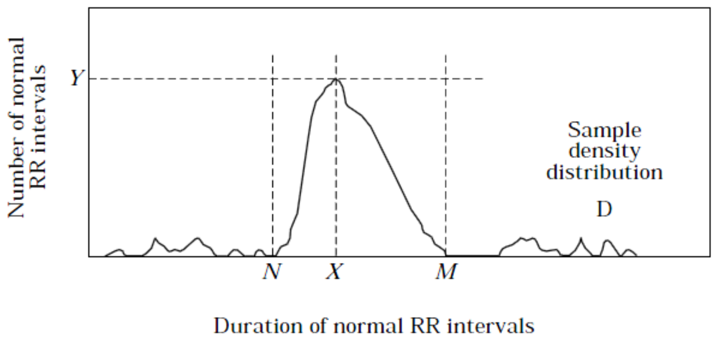


Fig. 2.3. Geometrical interpretation of HRV triangular index and TINN [2]

2.3. Frequency domain methods

In frequency domain methods we can distinguish two major approaches to HRV modelling: *non-parametric* and *parametric* spectral methods.

2.3.1. Non-parametric spectral analysis

In non-parametric spectral analysis, the most frequently used method is the Fast Fourier Transform (FFT) which enables us to calculate Power Spectral Density (PSD) of the HRV signal. On the contrary to time domain methods, frequency domain calculations are rather performed on short-term recordings, i.e. 2 to 5 minutes. One can consider spectral analysis of 24h recording, but it has to be stressed that using Fast Fourier Transform we assume stationarity of the signal. It cannot be assumed that ECG signal is stationary in such a long time period. The solution to this problem seems to be wavelet, packet wavelet transform and other [6–8]. Nevertheless, this method is quite complex, and similarly to previous methods does not give clear values for mental stress assessment. Thus it will be not considered in this paper.

In spectral analysis, several frequency bands can be distinguished: Ultra Low Frequencies (ULF) corresponding to 10^{-4} up to $3 \cdot 10^{-3}$ Hz, Very Low Frequencies (VLF) corresponding to $3 \cdot 10^{-3}$ up to 0.04 Hz, Low Frequencies (LF) corresponding to 0.04 to 0.15 Hz and High Frequencies (HF) corresponding to 0.15 to 0.4 Hz. In spectral analysis, the PSD in each spectra is calculated. PSD is expressed in $[ms^2]$, or in normalized units (n.u.) with normalization factor equal to $100 \cdot (TotalPower - VLF - ULF)^{-1}$. It should be stressed that slope of the 24h spectrum can be linearly fitted in log-log scale. From the technical point of view, in order to properly calculate all of the components, appropriate length of recording and sampling frequency[2] should be chosen. Moreover ectopic beats, arrythmia, missing data and noise effects may interfere with the final result of spectral analysis, and these phenomena should be taken into account.

2.3.2. Parametric spectral analysis

As mentioned in the earlier section, the problem of stationarity is an obstacle in long-term signal analysis. One of the solutions is usage of an autoregressive model, obtaining stationary signal. Reasonable order of AR model is an order of 16 which is the value used in software used for HRV analysis [5].

[2] Nyquist-Shannon sampling theorem

2.4. Non-linear methods

Non-linear methods are the most complex methods for ECG signal analysis, both mathematically, and from the point of view of physiological interpretation. The most frequently used method is so-called Lorenz Plots [9, 10], also known as Poincarè Plots which is supposed to be better in terms of reliability. They include both: short and long-term analysis. Sample Poincarè Plot is depicted in Fig. 2.4 [11].
A typical Poincarè plot (Fig. 2.4) constitutes from RR interval of the current normal beat on the abscissa and the RR interval of the succeeding normal beat on the ordinate. An ellipse is fitted into the resulting data point. Then two parameters are calculated: standard deviation of beat-to-beat RR variability (SD1) and the standard deviation of long-term RR variability (SD2), as well as the SD1/SD2 ratio.

It has to be stressed that Wavelet Transform and Wavelet Packet Transform seem to be the methods that solve the problem of stationarity of the signal, mentioned in section 2.3.1. Nevertheless, this is quite novel method and is not so broadly used as classical frequency analysis.

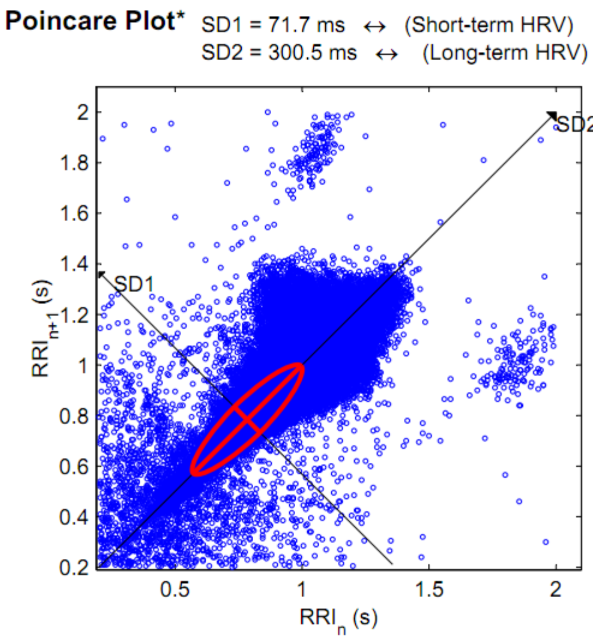


Fig. 2.4. Sample Poincaré plot

2.5. Physiological interpretation

The correlation of an ECG signal analysis and physiological response is still complicated and in some cases vague. The very crucial point about physiological response is the meaning of frequency domain analysis, namely **LF, HF** and **LF/HF**. In 1993 Malik et al [12] reported that both, low and high-frequency components of HRV represent modulated vagal and sympathetic activity of the autonomic nervous system, which suggest response to regulatory mechanism. Moreover, the Task Force [2] concluded that the ratio of **LF/HF** reflect sympathovagal balance. An exhausting explanation of the correlation between HRV and physiological response was given in the Committee Report [3].

We find two significant factors of HRV analysis, described in section 2.2.1. They concern 24 hours recordings:

- $pNN50 > 30\%$ - as a measure of probable arrhythmia [13].
- $NN50 \in [500; 2000]$ - as a measure of diabetic neuropathy [14].

3. A novel algorithm for HRV analysis

3.1. The research problem

There are numerous papers reporting changes in HRV due to mental stress [15–17]. Mainly LF/HF, SDNN, RMSSD are significantly affected. Unfortunately, these investigations provide only tendencies, rather than threshold values for stress assessment. So far, there is no algorithm capable of measuring mental stress level directly from ECG recorded in **non-laboratory conditions**. Our goal is to create a method which allows to perform broad time-domain analysis, which on the contrary to standard HRV methods, will give an insight into evolution of signal. This method would permit to establish symptoms of stress, and finally create a SCORE stress index. Furthermore, we introduce another two statistical parameters which satisfies our criteria.

3.2. General assumptions

According to standards [2] we will use 24h Holter recordings, as those which contain the entire day-night cycle. Our investigations are based on records consisting of RR intervals expressed in seconds.

3.3. Modal Frequency (FM)

FM is the heart rate at which one's heart passes most of the time. In other words, FM is the Heart Rate (HR) associated with maximal value of the histogram obtained from the entire recording, with width of the bin equal one. Sample frequency histogram is shown in Fig. 3.1. We think that modal frequency is better way to measure heart rate than mean of HR, which gives averaged-value of the entire HR. In consequence, modal frequency is supposed to be a measure of heart rate in rest. We postulate that FM can be a good predictor of tachycardia and bradycardia arrhythmias and general outline of one's activities during the day (i.e. general lifestyle). Nevertheless, this hypothesis needs clinical investigation.

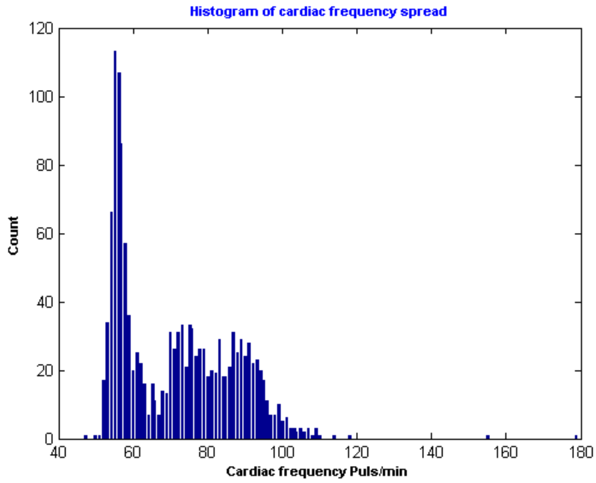


Fig. 3.1. Spread of all frequencies with one modal frequency equal to 55

3.4. Area Under Curve (AUC)

Let us define instantaneous cardiac frequency $F_c(t)$ 3.1.

$$F_c(t) = \frac{60}{RRinterval}[puls./min] \tag{3.1}$$

While calculating $F_c(t)$ from raw RR interval recording, we proceed as follows:

1. If in one second time period, more than one heart beat occurs, we calculate the mean of all intervals at this second.

2. If in one second time period, no heart beat occurs, we associate mean of two surrounding intervals to this second.

The AUC is inspired by I,MR chart used for quality control in industry. It is calculated for 2 s, up to 12 s windows[3].

$$AUC = \sum_{it=2}^{12} \frac{1}{nbM \cdot d_2(it)} \sum_{i=it}^{nbM} |min(F_{c_{i-it}}) - max(F_{c_{i-it}})| \tag{3.2}$$

Where:

- nbM - number of samples

[3]This method was developed by a statistician on demand of a company and is not published (internal report)

- $F_c(t)$ - instantaneous cardiac frequency.
- $d_2(it)$ parameter is used for evaluation of statistical significance of the results ($p < 0.0001$).

Low heart variability is characteristic for patients with chronic heart failure [18]. It is proven, that physical activity enhances variability of heart [19]. Accordingly to these facts, we introduce a well-known method, mostly applied in industry for quality control, as we believe that it will give better description of heart's variability than standard HRV parameters (e.g. SDNN, RMSSD, etc.).

3.5. Statistical analysis by Gaussian curves

Let us remind the formula for normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp \frac{-(x-\mu)^2}{2\sigma^2} \tag{3.3}$$

Where σ^2 is the variance and μ is the mean of observed variable x. Using equation 3.3 we choose random time-window (e.g. 10 minutes) of an ECG signal recording and trace normal distribution of HRs for each window. Sample plot is depicted in Fig. 3.2
It has to be stressed that HR in this case is the actual number of heart beats per minute and has nothing in common with instantaneous cardiac frequency ($F_c(t)$) defined in eq. 3.1. Moreover HR is always an integer.

Having Gaussian curves for each time window, we simply superpose them on one plot, obtaining sample distribution as in Fig. 3.3

3.5.1. Gaussian contour

Taking normal distribution traced for each window within an entire recording as depicted in Fig. 3.3, we take a contour of all Gaussian curves (Fig. 3.4); If one do not feel comfortable with an intuitive interpretation of contour, we introduce a mathematical formula presented in equation 3.4.

$$\underset{p \in \langle 0;250 \rangle}{\forall} \; \underset{i \in \langle 0; i_{max} \rangle}{\forall} C(p) = max(f_i(p)) \tag{3.4}$$

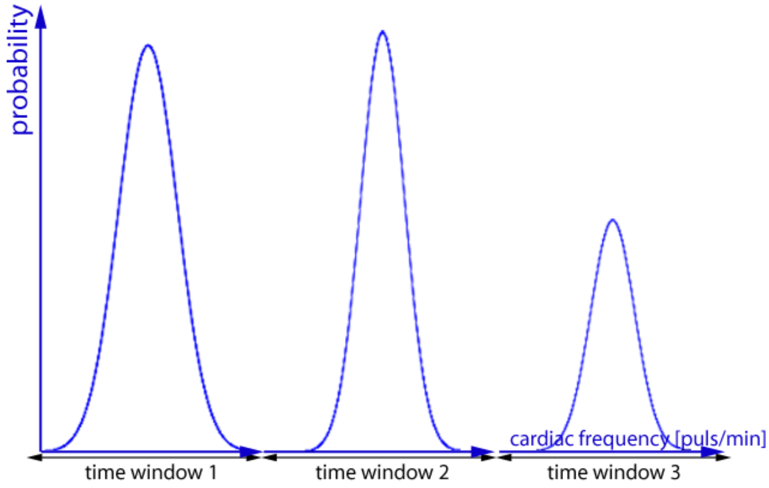


Fig. 3.2. Base curves which will be used to produce superposition of all curves

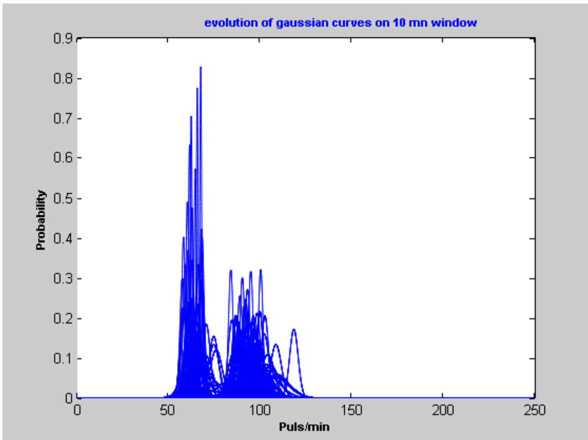


Fig. 3.3. Double Idle Mode

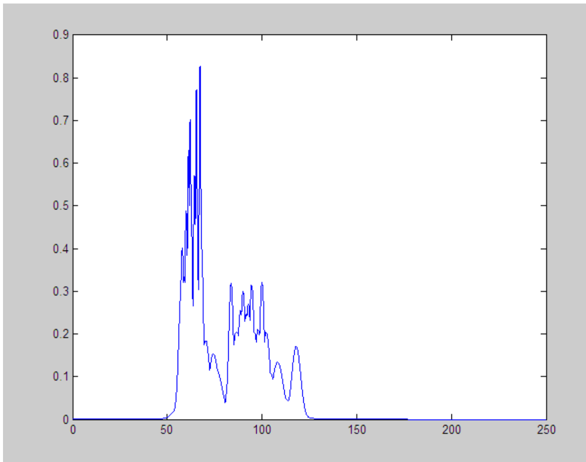


Fig. 3.4. Double Idle Mode - contour

Where p is HR (puls./min.), $f_i(p)$ signifies normal distribution function within one window.

In order to measure quantitatively Gaussian representation of ECG signal we introduce the two following parameters: $Rgauss(p)$ and $Wgauss(x)$

3.5.2. Rgauss

Denoting $C_1(p)$ and $C_2(p)$ as two contours obtained for normal distribution functions with different windows, integrating them and calculating their ratio we obtain:

$$Rgauss(p) = \frac{\int_0^{250} C_1(p)dp}{\int_0^{250} C_2(p)dp} \tag{3.5}$$

Thus $Rgauss(p)$ can be interpreted as a ratio of the areas under two contours, each computed for different window length.

3.5.3. Wgauss

Let us consider one more time a contour of normal distributions $C(p)$. Denoting $max(C(p))$ by $pmax$, we introduce function $Wgauss(x)$ which will measure HR spread with a certain level of probability; strictly speaking:
Find interval u within the contour $C(p)$ that consists of all heart pulsations p that satisfies condition:

$$u: \ C(p) \geq x \cdot p_{max} \tag{3.6}$$

$$Wgauss(x) = max(u) - min(u) \tag{3.7}$$

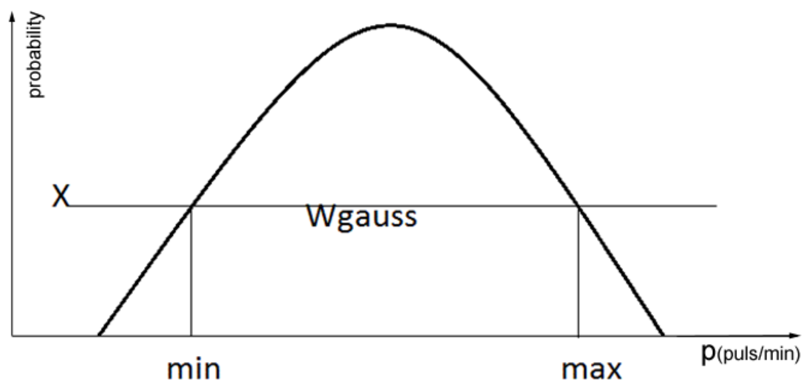


Fig. 3.5. Calculation of Wgauss

4. Results

The recordings for our tests constitute a database of 43 recordings from random chosen subjects, where 9 subjects responded to stress survey consisting of following 14 questions [20]:

1. In the last month, how often have you been upset because of something that happened unexpectedly?
2. In the last month, how often have you felt that you were unable to control the important things in your life?
3. In the last month, how often have you felt nervous and “stressed”?
4. In the last month, how often have you dealt successfully with day to day problems and annoyances?
5. In the last month, how often have you felt that you were effectively coping with important changes that were occurring in your life?
6. In the last month, how often have you felt confident about your ability to handle your personal problems?
7. In the last month, how often have you felt that things were going your way?
8. In the last month, how often have you found that you could not cope with all the things that you had to do?
9. In the last month, how often have you been able to control irritations in your life?
10. In the last month, how often have you been angered because of things that happened that were outside of your control?
11. In the last month, how often have you found yourself thinking about things that you have to accomplish?
12. In the last month, how often have you been able to control the way you spend your time?

13. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

Scale:

0 - never

1 - almost never

2 - sometimes

3 - fairly often

4 - very often

NOTE: items 4,5,6,7,9,10,13 are measured in the reverse directions.

Each subject had performed 24h Holter test. The output files were text files with RR interval given in milliseconds. Then the files were preprocessed, i.e. misinterpreted intervals were deleted as well as redundant data. For each subject we had calculated normal distribution for 10 minutes window. Subjects were subdivided into 4 categories: single idle mode (Fig. 4.1), double idle mode (Fig. 3.3), more than 2 idle modes (Fig. 4.2) and uncategorized (Fig. 4.3).

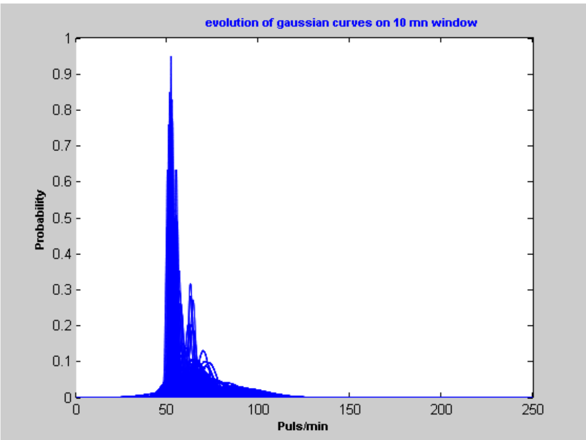


Fig. 4.1. Single Idle Mode

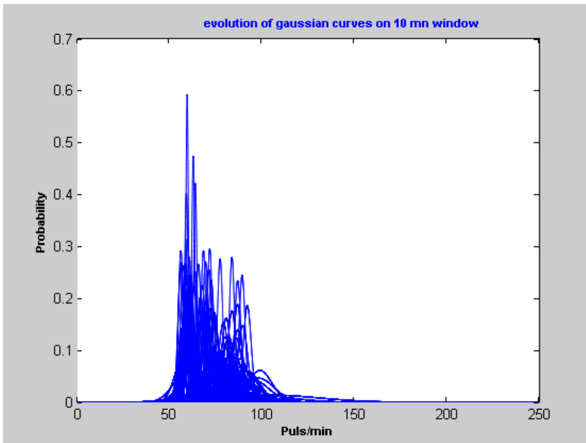


Fig. 4.2. More than two Idle Modes

For each subject's normal distribution, we had measured Rgauss according to equation 3.5 and Wgauss according to equation 3.7. Rgauss was calculated as a ratio of contour obtained for 5 min window, and 30 min window. 5 minute window is supposed to be a measure of short-term changes, and 30 min window is supposed to be a measure of long-term changes in ECG signal. Nevertheless in clinical use different windows may

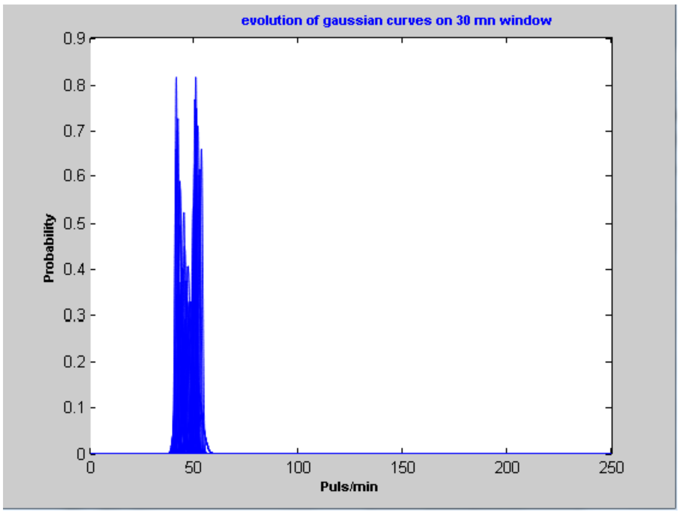


Fig. 4.3. Unclassified

show a predictive value. Moreover, Wgauss was measured at three levels: $0.2pmax$, $0.6pmax$, and $0.8pmax$ where $pmax$ means maximal probability of the Gaussian contour.

After calculations, we compared each component with HRV analysis methods mentioned in section 3. They didn't correlate with each other - maximal correlation using Multiple Linear Regression was less than 0.01.

Among the subjects that responded to stress survey, there was no tendency observed or correlation between ECG signal-based analysis and stress SCORE index determined by the survey. Obviously, it does not mean that there was no correlation between those two factors, since in practice it requires a very large number of samples. However we had found subjects that have common HRV parameters, but different Gaussians parameters (tab. 4.1).

Let us define a relative difference is defined by equation 4.1.

$$d_r = \frac{x - y}{max(|x|, |y|)} \tag{4.1}$$

Tab. 4.1. Comparison of HRV parameters obtained for two subjects

	Patient 1 (Fig. 4.1)	**Patient 2 (Fig. 4.2)**	**Relative difference** [%]
MeanRR	0,831[s]	0,965[s]	13,89
STD	0,101[s]	0,115[s]	12,17
RMSSD	110,9	111,1	0,18
pNN50	12,4	13,3	6,77
RR triangular index	0,085	0,084	1,18
TINN	1305	1390	6,12
FM	60	52	13,33
AUC	3,5937	0,8648	75,94
Rgauss	2,9626	2,5034	15,50
Wgauss$(0,2 \cdot p_{max})$	3,77	1,92	49,07
Wgauss$(0,6 \cdot p_{max})$	5,84	5,58	4,45
Wgauss$(0,8 \cdot p_{max})$	39,97	16,58	58,52

Moreover, the histogram of all patients is depicted in Fig. 4.4 (p-value < 0.001). The bin is equal to one.

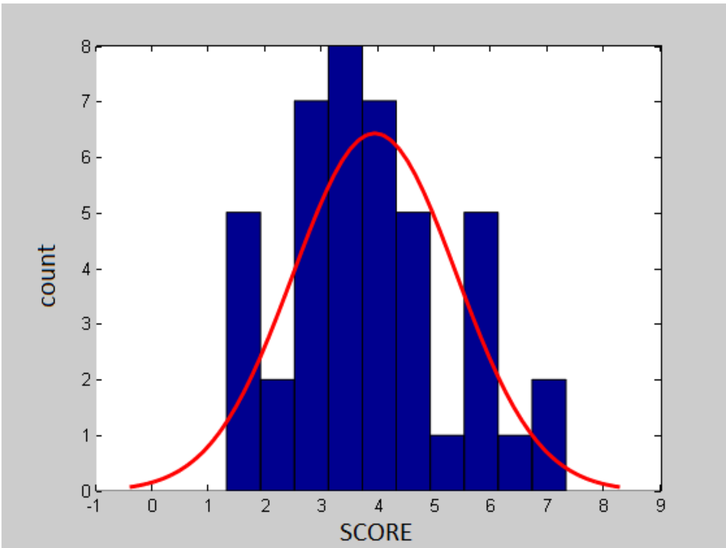


Fig. 4.4. Histogram of 43 SCOREs

Finally we present three algorithms:

1. **SCORE Brack**

   An algorithm developed on internal demand of the company related to health care. It bases only on two parameters: FM and AUC (eq. 4.2). The original method was established on very small group of subjects, i.e. about 10. Since we doubt in pertinence of this algorithm, it will not be discussed in details.

$$SCORE(FM, AUC) = const + a.FM + b.AUC + d(FM + e)^2 + f(AUC + g)^2 \quad (4.2)$$

where $const, a, b, d, e, f, g$ are constants.

2. **SCORE Brack Survey**

   An algorithm based on SCORE Brack algorithm. The constants were established basing on stress survey acquired from 9 subjects.

3. **SCORE Gauss**

   An algorithm based on several parameters according to eq. 4.3.

$$\begin{aligned} SCORE \; = \; & \\ & 0.15FM + 0.1AUC + 0.3Rgauss \\ & + 0.1pNN50 + 0.1Wgauss(0.2p_{max}) \\ & + 0.1Wgauss(0.6p_{max}) + 0.15Wgauss(0.8p_{max}) \end{aligned} \quad (4.3)$$

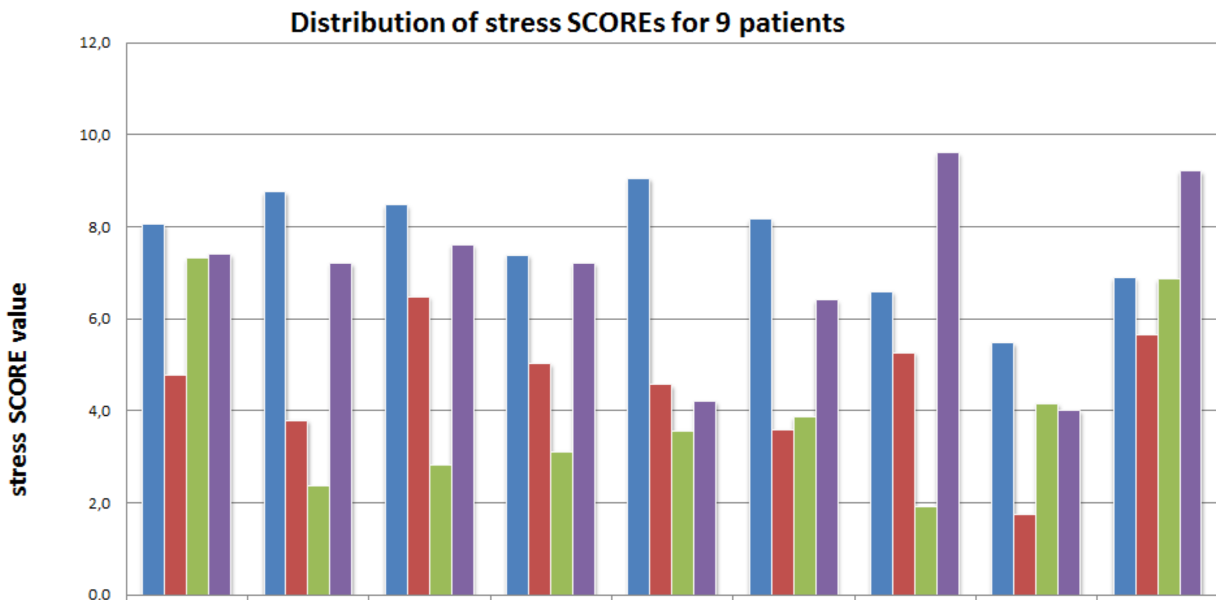
The weight coefficients were chosen according to the estimated significance. Each variable can take 4 color codes with corresponding numerical values:

- Green = 1
- Yellow = 4
- Orange = 7
- Red = 10

The classification of color codes was established, basing on median value of each parameter. The Rgauss parameter is a ratio of 5 and 30 minutes windows which correspond to short and long term Gaussian evolution.

5. Discussion

In our opinion, the single idle mode has the greatest potential to be the symptom of low mental stress, since no matter what stressor was applied, it always returns to single idle mode. On the contrary when more than one idle mode is recognized, there could be more explications in terms of physiological response; the subject is under mental stress, and thus basic cardiac frequency is shifted toward higher frequencies [15] and finally we can observe Gaussian distributions that are shifted to the right. Second explanation could be physical activity of the subject where increased HR is observed [19, 21]. There can exist other factors that explain this phenomenon. The data shown in table 4.1 includes the data computed for two subjects with Gaussians distributions depicted in figure 4.1 and figure 4.2 for Patient 1 and Patient 2 respectively. This proves that despite the HRV parameters are relatively similar, the parameters based on Gaussian analysis are very different (i.e. Wgauss$(0,2 \cdot p_{max})$ and Wgauss$(0,8 \cdot p_{max})$). Even by visual inspection we can distinguish different idle modes. It means that our method introduces new information to ECG signal analysis, and may reveal a predictive value.



	1	2	3	4	5	6	7	8	9
SCOREBrack	8,0	8,7	8,5	7,4	9,0	8,2	6,6	5,5	6,9
SCOREBrackSurvey	4,8	3,8	6,5	5,0	4,6	3,6	5,3	1,7	5,6
SCOREGauss	7,3	2,4	2,8	3,1	3,6	3,9	1,9	4,2	6,9
StressPerceived	7,4	7,2	7,6	7,2	4,2	6,4	9,6	4	9,2

Fig. 4.5. Distribution of 9 patients' SCOREs

While taking into consideration the distribution of different SCOREs as depicted in Fig. 6.2 we can see that the results between different algorithms and stress perceived diverge. Nevertheless, we do not discriminate the predictive value of these methods. This is due to lack of subjects to investigate.

6. Development of Matlab® toolbox for Gaussian analysis of HRV

In order to easily perform the calculations, I created two graphical toolboxes in Matlab®. First which evaluates stress SCORE basing on FM and AUC, second that bases on FM, AUC, Rgauss, Wgauss and pNN50.
We dispose a database of more than 40 patients with measurements performed by SORIN equipment. It means that both toolboxes use the following file format as shown in Listing 1.

```
Surname First Name Hook up date : 01/01/2000 additional data
RR (ms) export from 12:31:33 to 11:37:35
12:31:33   990 A
12:31:34   1000  C
12:31:35   1000  C
...
```

Listing 1. File format

First two lines are header lines, the rest are measurements with first colon containing time in format HH:MM:SS and second colon RR interval expressed in ms.
Moreover, both toolboxes provide stress SCOREs evaluation with scale of stress from 1 up to 10 , where 1 is no stress and 10 is hyperstress.

6.1. Toolbox based on FM and AUC

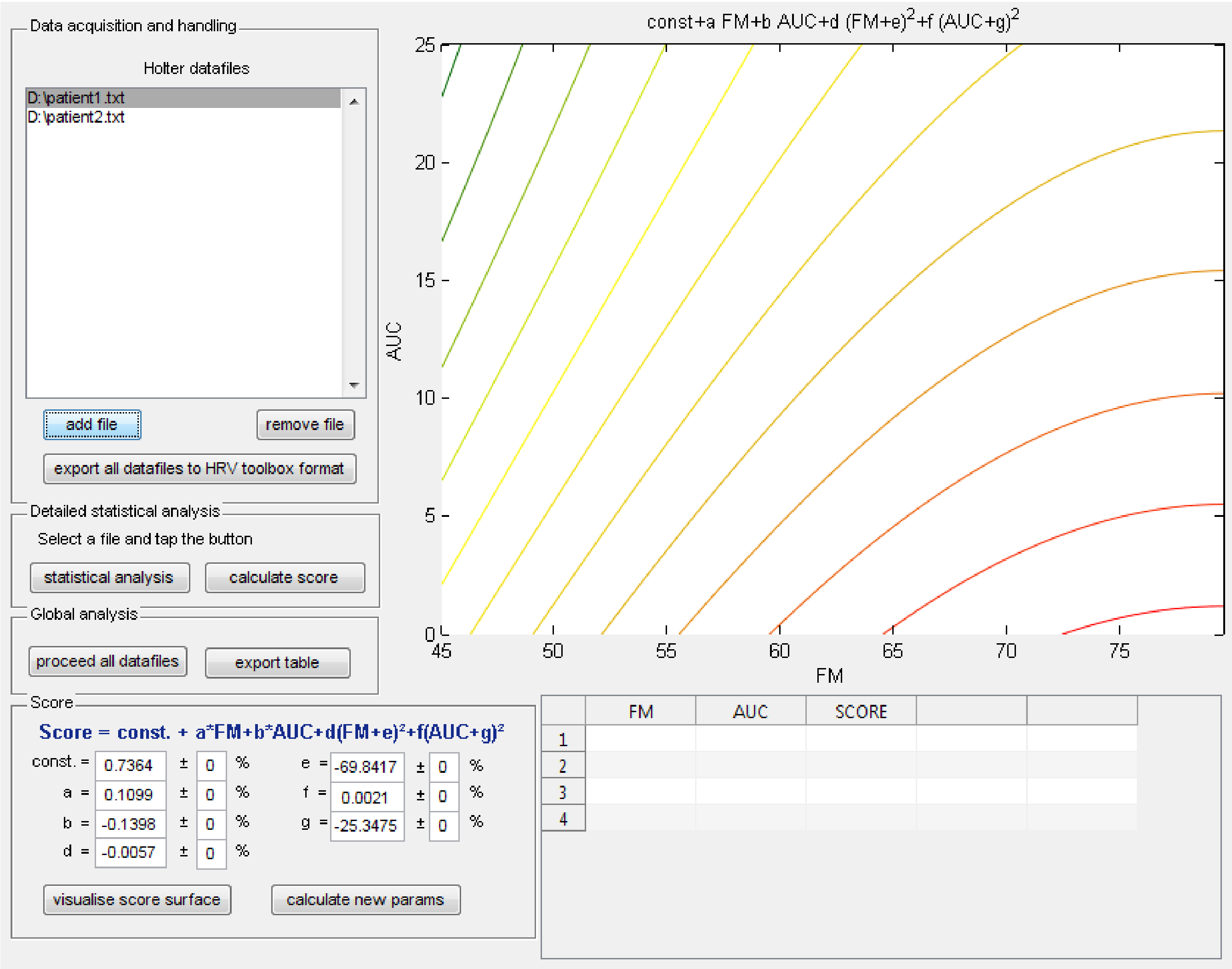


Fig. 6.1. Toolbox for stress SCORE evaluation based on FM and AUC

6.1.1. Buttons

Add file - add files to the 'Holter datafiles' list. Only first two colons are taken into consideration.

Remove file -removes selected file from the list

Export all datafiles to HRV toolbox format - exports all data files in the table to HRV toolbox [5] format, that is only one colon with RR interval expressed in ms. Default folder is ./HRVdata

Statistical analysis shows results of statistical analysis performed on currently se-

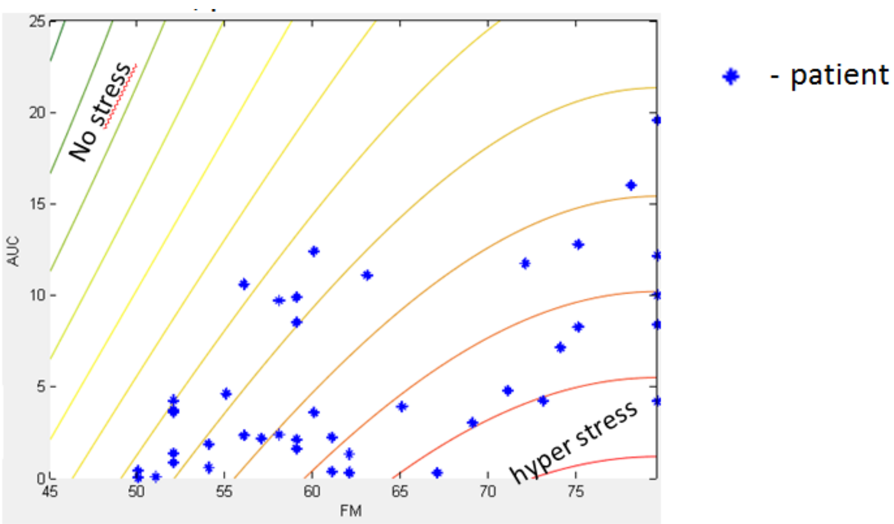


Fig. 6.2. Distribution of 43 patients SCOREs for the Brack's algorithm

lected file. Statistical analysis means instantaneous cardiac frequency vs. time and Gaussian curves evolution.

Calculate score - calculates stress SCORE of currently selected patient in the list. The formula used for calculations is visible in the tab 'Score'

Proceed all datafiles - calculates SCOREs of all patients in the list, and displays result in the table in lower right corner. Note that treatment of one 24h recording may take several minutes depending on speed of CPU.

Export table - exports the entire table with data to .csv file

Visualize score surface - shows SCORE formula surface in 3D. Since SCORE is a function of two variables, the resulting visualization is a surface.

Calculate new params - permits to calculate new parameters of SCORE (const., a, b, d, e, f, g) by Multiple Linear Regression. Note that before calculation of new parameters, you have to calculate all scores (proceed all datafiles button), and then fill in prevision colon in the table located in lower right corner. In this case, the final score formula will be calculated in order to fit the prevision scores. They can be based on e.g. stress surveys, or medical inspection.

6.1.2. Plot of SCORE function

The plot in the central part of GUI is projection of SCORE surface on 2D. On X axis we observe FM, and Y axis AUC. When SCOREs are calculated, the patients are assigned to this surface and are noted by blue stars. While changing SCORE parameters, the surface projection changes. Note that FM and AUC remain constant. Thus this is only the projection that changes.

6.1.3. Error function boxes

Concerns: 'Score' tab.

It is possible to investigate error influence of each parameter on the final SCORE. The boxes at the right hand side of each constant allow us to calculate each SCORE with an error. This option is valid only with 'Calculate score' button. By default we obtain result with 0% error (SCORE), with custom error influence (SCOREu), and four other errors (SCOREx where x is 1,5,10,20 % respectively).

6.2. Toolbox based on FM, AUC and Gaussian analysis

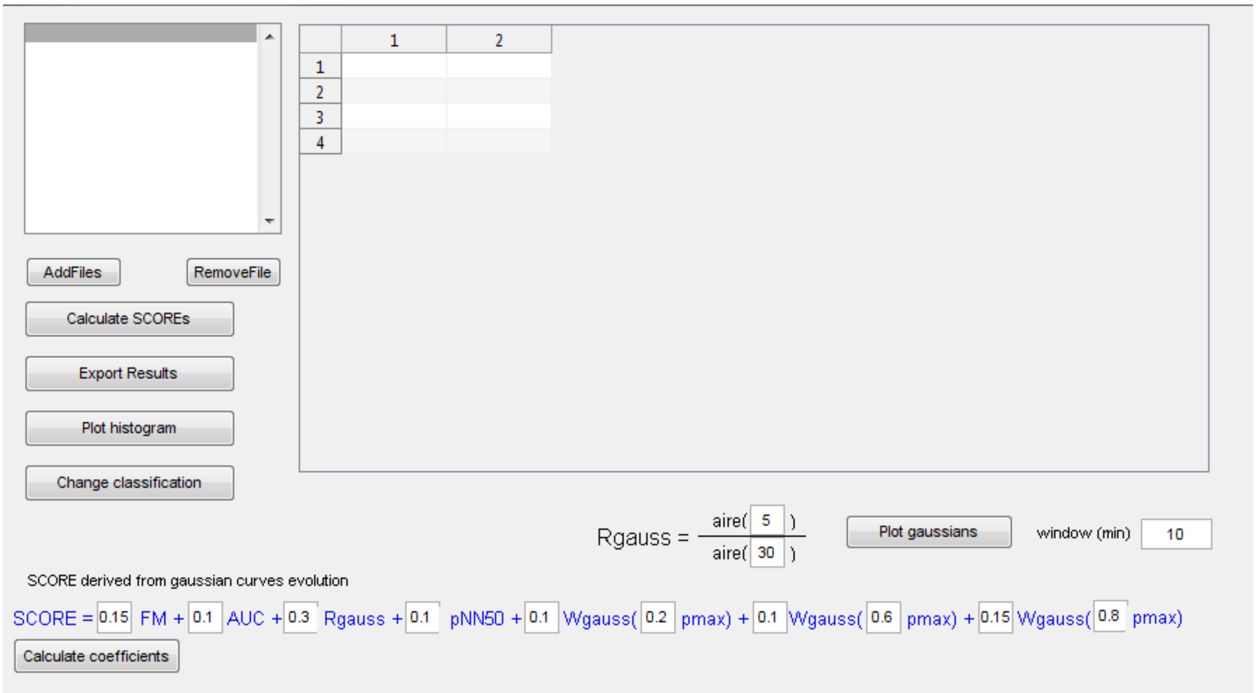


Fig. 6.3. Toolbox for stress SCORE evaluation based on FM, AUC and Gaussian analysis

6.2.1. Buttons

Add file - add files to the 'Holter datafiles' list. Only first two colons are taken into consideration.
Remove file -removes selected file from the list
Calculate SCOREs calculate scores of all patients in the list, and displays result in the table. Note that treatment of one 24h recording may take several minutes depending on speed of CPU.
Export results - exports the entire table with data to .csv file
Plot histogram - plots a histogram of calculated SCOREs and perform normal fitting. Note that you have to calculate SCOREs before plotting a histogram.
Change classification - uses a sub-GUI in order to change the classification of the patients.
Plot Gaussians - serves to plot normal distributions of windowed ECG signal. The window can be modified by EditBox on the right.
Calculate coefficients - calculate Multiple Linear Regression based on Prevision entered by user. Note that it is necessary to calculate SCOREs before using this button.
Change classification - While calculating coefficients, we use color codes instead of direct values. Color codes can be changed by 'change classification' button, and they have following values:

- Green = 1
- Yellow = 4
- Orange = 7
- Red = 10

6.2.2. EditBoxes

- **SCORE formula editboxes**
  We can change weights of each coefficient. Moreover it is possible to change level of measure of Wgauss parameter

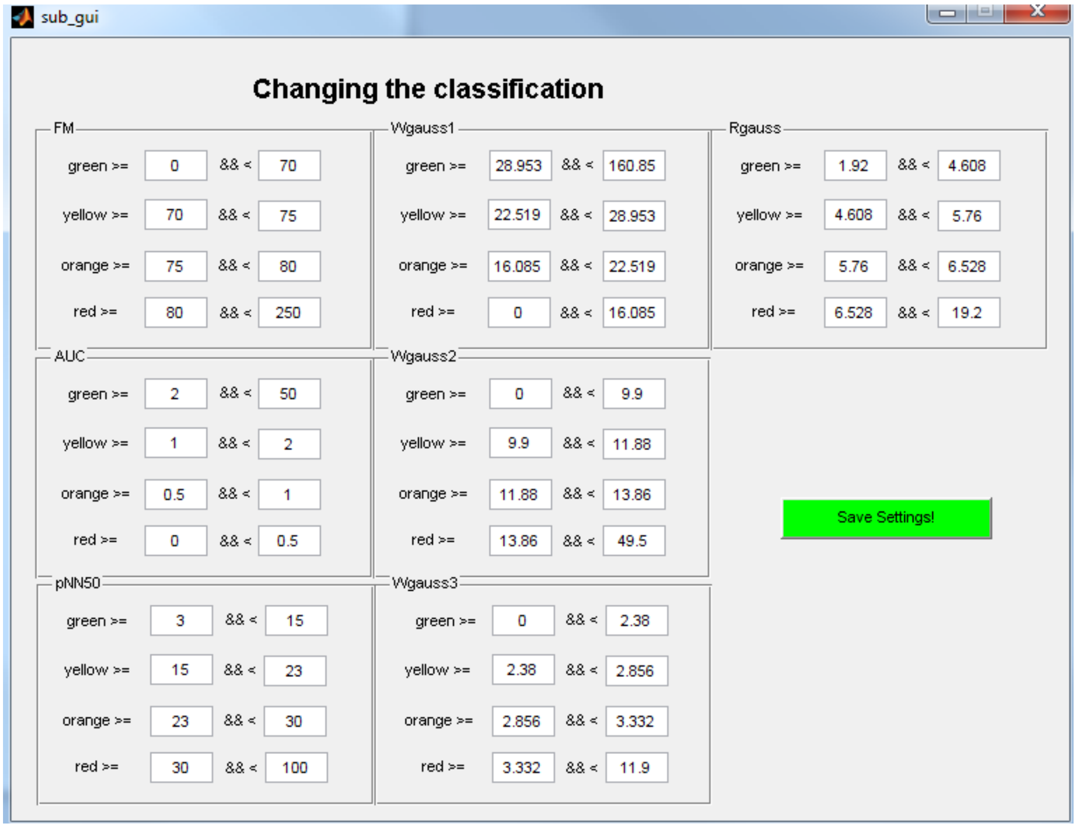


Fig. 6.4. GUI for changing the classification of the parameters

- **Rgauss editboxes**

  It is possible to window length of each contour of Gaussians.

6.3. Most vital algorithms in Matlab®

Since the algorithms like Rgauss (eq. 3.5) or Wgauss (eq. 3.7) are easily implementable once we have an array with Heart Rate $HR(t)$ as a function of time. Time is meant to be expressed in minutes, obtaining HR in pulsations per minute. Hence it is the crucial process to obtain function $HR(t)$ from the source file shown in Listing 2.

```
%Reading the sourcefile - ignoring two first lines
[HH, MM, SS, interval]= textread(filename, '%2d:%2d:%2d\t%d%*[^\n]', '
    headerlines', 2);

%we check at which our the recording started
heure_debut = HH(1);
```

```
%looking for midnight
for i=2:length(interval)
    if (HH(i-1,1)== 23) && (HH(i,1)==00)
        %midnight found !
    break ;
    end
end

%After midnight we add 24 hours (it means this is a new day)
for p=i:length(interval)
    HH(p,1) = HH(p,1) + 24;
end

%transforming the time into seconds
temps = HH.*3600 + MM.*60 + SS; %


%Since the recording did not start at midnight we have to normalize it
%in such way that we do not have day "longer" than 24 hours
temps11=temps(1,1);
for i=1:length(temps)
    temps(i,1)=temps(i,1) - temps11;
end
%now the time is within the limits of 0H 00M 00S up to 23H 59M 59S

%we are looking if there is an integer number of minutes in the recording
reste = mod(temps(length(temps)),60);

%temps_max is the integer number of minutes
temps_max = temps(length(temps),1)-reste;

%allocating a matrix for the cardiac frequency in function of time
fc=zeros((temps_max)/60+1,1);


%calculating cardiac frequency
%note that with this calculations, cardiac frequency is always integer
for i=1:temps_max
```

```
  %indice is the current minute of the recording
    indice=floor(temps(i,1)/60)+1;
  %for a given minute (indice) we increase fc by 1 at each occurrence
    fc(indice)=fc(indice)+1;
end



%we treat that fc lower than 35 puls/minute are errors of the recording
fc(fc<=35)=[];
```

Listing 2. HR(t)

Second algorithm, shown in Listing 3 calculates three functions: instantaneous cardiac frequency – $F_c(t)$, Modal Frequency – FM and Area Under Curve – AUC.

```
[temps, interval] = readfile(path);
fc=calculateFc(path);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Looking for two beats occuring at the same second%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
len =length(temps);
for i=(length(temps)):-1:1

   %looking for the same time occurences the recording
   T=find(temps==temps(i,1));
   if length(T)>1
    %if there are more than 1 occurences, we save the indexes
    for j=length(T):-1:2
        dokasowania(i*j)=T(j);

    end
  %we assign time interval(RR) as a mean of all RR intervals occured at
      that second
    interval(T(1))=mean(interval(T));

   end
end
```

```
%deleting doubled times and intervals
dokasowania(dokasowania==0)=[];
temps(dokasowania(:))=[];
interval(dokasowania(:))=[];

%assigning instantaneous cardiac frequency for each second
for i=1:length(interval)
    interval(i)=(60*1000)./interval(i);
end

%deleting frequencies lower than 35 treating them as errors in the
    recording.
indices=find(interval<35);
interval(indices)=[];
temps(indices)=[];


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Looking for every second without any heart beat %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
len=length(temps);
for i=1:(len-1)

  %looking for these seconds where no heart beat occured
    dif = temps(i+1)-temps(i);
    if dif>1
    %when such a second found, we assign a mean of neighboring intervals
        to that second
        avg=mean([interval(i),interval(i+1)]);
        tempstmp=temps([(i+1):length(temps)]);

        intervaltmp=interval([(i+1):length(interval)]);

        for j=1:(dif-1)
            temps (i+j)=temps(i)+j;
            interval(i+j)=avg;
        end
```

```
        for p=dif:length(tempstmp)
            temps(i+p)=tempstmp(p);
            interval(i+p)=intervaltmp(p);
        end

    end
end

%standard deviation table for normalization of moving range chart
dn(1)=1;
dn(2)=1.128;
dn(3)=1.693;
dn(4)=2.059;
dn(5)=2.326;
dn(6)=2.534;
dn(7)=2.704;
dn(8)=2.847;
dn(9)=2.97;
dn(10)=3.078;
dn(11)=3.173;
dn(12)=3.258;

%calculating moving range chart 2 - 12 seconds

R=zeros(length(interval),12);
for i=2:(length(interval))
    R(i,2)= abs(interval(i)-interval(i-1));
end

for j=2:11
    for i=(j+1):length(interval)
        R(i,j+1)=max(interval(i-j:i))-min(interval(i-j:i));
    end
end

srednia = mean(R);
```

```
%final result
AUC=sum(srednia./dn);

%calculating the histogram of cardiac frequencies
FMcount = zeros(1,250);
for i=1:length(fc)
    FMcount(fc(i))=FMcount(fc(i))+1;
end

%looking for the modal frequency
for i=1:length(FMcount)
    if FMcount(i) == max(FMcount) ;
        FM = i;
        break;
    end
end


FM=i;

%assumption of the algorithm
if FM > 79.5
    FM=79.5;
end
```

Listing 3. calculation of AUC

7. Conclusions

This thesis sums up three months of internship, i.e. about 500h of workload (developing algorithms, coding), where I have tested more than 20 different algorithms. The toolboxes take roughly 2000 lines of code, where about 500 lines are unique and written manually.

The intention of this thesis is not to give a direct algorithm of stress evaluation, but to give a tool for doctors and especially for cardiologists that will be validated by them in practice. The method presented in this thesis gives a new insight into temporal evolution of ECG signal, retaining its sequentiality. This could not be obtained by other or earlier reported method. This sequentiality holds information concerning one's activities during the day, and permits to classify normal activities as the states in equilibrium, and abnormal states as indicators of potential mental stress. It is obvious that this method needs broad clinical investigation, as well as verification against many other parameters than presented in this work (i.e. sex, age, medical treatment, illnesses, etc.). Concerning possible application of the proposed methods as stress indicator, question arises, about non-invasive measurement of stress level in order to validate the results. Although, the pertinence of stress survey was scientifically investigated [20], other stress surveys have to be taken into consideration.

The minimal satisfactory number of subjects to investigate is about 2000, which will permit to combine all algorithms with much more sophisticated methods, for example Neural Network prediction, or Neural Network Clustering.

Finally, although this algorithm is consulted with cardiologists from one of the biggest cardiology clinic in France, this thesis is an invitation for other cardiologists and researchers to check described above methods in practice. Toolboxes presented in the section 6 are available freely on the website of the author: `http://tajga.zapto.org` or on-demand by mailing the author: s.kosowski@gmail.com

References

[1] Bong Suk Cha Jong Ku Park Soon Koo Baik Sei Jin Chang Myung Gun Kang, Sang Baek Koh. "Job stress and cardiovascular risk factors in male workers". *Preventive medicine*, 40:583–588, 2005.

[2] Task Force of The European Society of Cardiology, The North American Society of Pacing, and Electrophysiology. "Heart rate variability, standards of measurement, physiological interpretation, and clinical use". *European Heart Journal*, 17: 354–381, 1996.

[3] Eckberg DL Grossman P Kaufmann PG Malik M Nagaraja HN Porges SW Saul JP Stone PH van der Molen MW Berntson GG, Bigger JT Jr. "Heart rate variability: origins, methods, and interpretive caveats". *Psychophysiology*, 34:623–648, 1997.

[4] Tomasz Pięciak. "Analiza zmienności rytmu serca". http://home.agh.edu.pl/~pieciak/SygnalyBiomedyczne/laboratorium_hrv.pdf.

[5] Perttu O. Ranta-aho Juha-Pekka Niskanen, Mika P. Tarvainen and Pasi A. Karjalainen. "Software for advanced HRV analysis". 2002.

[6] Alan R. Hargens. Kunihiko Tanaka. "Wavelet packet transform for RR interval variability". *Medical Engineering & Physics*, 26:313–319, 2004.

[7] Robert A.Hovis D.J.Hebert Vladimir Shusterman Soundararajan Ezekiel, Abdullah A. Alshehri. "Heart rate variability (HRV) signal processing by using wavelet based multifractal analysis". *Northern University Ada, Ohio, USA; University of Pittsburgh, PA, USA*, 2001.

[8] Boriana G. Piryova Nina Y. Belova, Stoyan V. Mihaylov. "Wavelet transform: a better approach for the evaluation of instantaneous changes in heart rate variability". *Autonomic Neuroscience: Basic and Clinical*, 131:107–122, 2007.

[9] Xavier Copie MD Anne Staunton BSc RN Katerina Hnatkova, PhD and MD Marek Malik, PhD. "Numeric processing of lorenz plots of RR intervals from long-term ECGs ". *Journal of Electrocardiology*, 28:74–81, 1995.

[10] M. Palaniswami M. Brennan and P. Kamen. “Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability”. *IEEE Trans Biomed Eng*, 48:1342–1347, 2001.

[11] Héctor Pérez-Grovas Claudia Lerma, Oscar Infante and Marco V.José. “Poincarè plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients”. *Clinical Physiology and Functional Imaging*, 23:72–80, 2003.

[12] MD Marek Malik, PhD and MD A. John Camm. “Components of heart rate variability – what they really mean and what we really measure”. *The American Journal of Cardiology*, 72:821–822, 1993.

[13] K. Suresh Kumar B. Anuradha and V. C. Veera Reddy. “Classification of cardiac signals using time domain methods”. *ARPN Journal of Engineering and Applied Sciences*, 2008.

[14] C M Shapiro J A Stewart W Reid D J Ewing, J M M Neilson. “Twenty four hour heart rate variability: effects of posture, sleep, and time of day in healthy controls and comparison with bedside tests of autonomic function in diabetic patients”. *British Heart Journal*, 65:239–244, 1991.

[15] Lilianne R. Mujica-Parodi, Gülce N. Dikecligil. “Ambulatory and challenge-associated heart rate variability measures predict cardiac responses to real-world acute emotional stress”. *Biol. Psychiatry*, 67:1185–1190, 2010.

[16] M. Nyssen J. Haueisen (Eds.). J. Vander Sloten, P. Verdonck. “Influence of mental stress on heart rate and heart rate variability”. *IFMBE Proceedings*, 22:1366–1369, 2008.

[17] Sook-hyun Kim Desok Kim, Yunhwan Seo and Suntae Jung. “Short term analysis of long term patterns of heart rate variability in subjects under mental stress”. *International Conference on BioMedical Engineering and Informatics.*, 2008.

[18] S Capewell T M MacDonald J M M Neilson D J Ewing. J Nolan, A D Flapan. “Decreased cardiac parasympathetic activity in chronic heart failure and its relation to left ventricular function”. *British Heart Journal*, 67:482–485, 1992.

[19] Keijo Peuhkurinen Raimo Kettunen Rainer Rauramaa Petri Tuomainen, T. "Regular physical exercise, heart rate variability and turbulence in a 6-year randomized controlled trial in middle-aged men: The dnasco study". *Life Sciences*, 77: 2723–2734, 2005.

[20] G.M. Williamson S. Cohen. "Perceived stress in a probability sample of the united states". *The Social Psychology of Health*, pages 30–67, 1988.

[21] Anne Favre-Juvin. Véronique-Aurélie Bricout, Simon DeChenaud. "Analyses of heart rate variability in young soccer players: The effects of sport activity". *Autonomic Neuroscience: Basic and Clinical*, 154:112–116, 2010.