Classification of Acute Stress using Linear and Non-Linear Heart Rate Variability Analysis Derived from Sternal ECG

George Tanev[1], Dorthe B. Saadi[2], Karsten Hoppe[3] and Helge B.D. Sorensen[4]

Abstract— **Chronic stress detection is an important factor in predicting and reducing the risk of cardiovascular disease. This work is a pilot study with a focus on developing a method for detecting short-term psychophysiological changes through heart rate variability (HRV) features. The purpose of this pilot study is to establish and to gain insight on a set of features that could be used to detect psychophysiological changes that occur during chronic stress. This study elicited four different types of arousal by images, sounds, mental tasks and rest, and classified them using linear and non-linear HRV features from electrocardiograms (ECG) acquired by the wireless wearable ePatch® recorder. The highest recognition rates were acquired for the neutral stage (90%), the acute stress stage (80%) and the baseline stage (80%) by sample entropy, detrended fluctuation analysis and normalized high frequency features. Standardizing non-linear HRV features for each subject was found to be an important factor for the improvement of the classification results.**

I. INTRODUCTION

Chronic stress is a psychophysiological state that occurs due to excessive levels and duration of stress. The human body's built-in endocrine response to acute (short-term) stress is essential for adaptation to changing situations such as the well known "fight or flight response" in times of perceived danger, in cases of anxiety or in emotional tension. Chronic stress can be broadly described as a prolonged exposure to an environmental demand (stressor) that exceeds the body's ability to react to a situation [1]. This overload on the endocrine system can increase risk of hypertension and cardiovascular diseases which are commonly linked to chronic stress. Sympathetic stimulation during acute stress typically increases cardiovascular properties of chronotropy (heart rate), inotropy (contractility), lucitropy (relaxation) and dromotropy (conduction velocity) while parasympathetic stimulation tends to decrease these same properties in non-stress situations [2].HRV, which represents the change in duration between heart beats over time, has been commonly used to track changes in cardiovascular activity which reflect the balance between the sympathetic and parasympathetic nervous systems [3].

[1]George Tanev is a Master's student at the Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 349, DK-2800 Kgs. Lyngby, Denmark

[2]Dorthe B. Saadi is an Industrial PhD student at the Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 349, DK-2800 Kgs. Lyngby, Denmark, and DELTA, Body Sensors Department, Venlighedsvej 4, DK-2970 Hørsholm, Denmark

[3]Karsten Hoppe is a Research Manager at DELTA, Body Sensors Department, Venlighedsvej 4, DK-2970 Hørsholm, Denmark

[4]Helge B.D. Sørensen is an Associate Professor at Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 349, DK-2800 Kgs. Lyngby, Denmark

Research in applying HRV analysis to the field of chronic stress is limited due to the lack of a "Golden Standard" and the lengthy experimental set-ups. The current methodologies for chronic stress diagnosis still rely heavily on questionnaire assessments and sympathetic tone measures such as cortisol or norepinepherine spillover in the blood, which are not considered robust enough to be golden standards [4], [5]. A methodology for detecting psychophysiological deviations from the baseline over long periods, as in the case of chronic stress, based on HRV would be a novel approach. The long term motivation and vision for this project is the implementation of such a methodology to monitoring chronic stress and thus reducing the risk of disease.

A pilot study was performed with its goal being to develop a linear and non-linear HRV feature set for acute stress classification. The purpose of developing this methodology is for its future application in experiments involving chronic stress. It is not expected that the same HRV features will be relevant in acute and chronic stress detection, but the availability of a large feature set will allow the identification of the relevant features in each respective scenario. This paper will cover a brief review of the current research in acute stress classification, the methodology used for acute stress elicitation and HRV analysis and the classification and discussion of these results.

A. HRV Parameters Used in Research

The most common ECG features that are used in stress classification in literature are based on either linear HRV features (time and frequency domain) or non-linear HRV features. Significant decreases in Low Frequency (LF) and High Frequency (HF) HRV components and significant increases in the LF/HF HRV ratio during induced mental stress compared to a rest period were found in the the work of Hjortskov et al. [6]. Valenza et al. [7] compared linear feature sets with non-linear feature sets to classify four emotional levels. The non-linear features attained a significant increase in classification rate between different emotions. Rates over 90% were achieved with a quadratic discriminant Bayesian classifier and feature reduction by means of principal component analysis. Mellilo et al. [8] applied a non-linear HRV feature set to classify stress in students during an exam situation. Classification rates of 90%, 86%, and 95% for accuracy, sensitivity and specificity, respectively, were achieved when using the Poincaré plot and approximate entropy parameters.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

Fig. 1. ePatch® ECG recorder correctly placed on sternum.

II. ACUTE STRESS CLASSIFICATION METHODOLOGY

Videos composed of pictures and sounds, a mental stress task stage using a Stroop test, and a rest stage were used to elicit varying degrees of short-term stress/arousal in the subjects. The images and sounds used in this study were from the recognized International Affective Image System (IAPS) [9] and the International Affective Digital Sounds (IADS) [10] databases, respectively. The sounds and images were graded based on valence (unpleasant to pleasant) and arousal (weak to strong) on a scale of 1 to 9, respectively.

A. Test Subjects

The subjects used for this acute stress study were 10 volunteer students (8 male) from the Technical University of Denmark, aged 22 to 26 years, that were deemed to be in good physical health, mental health and did not smoke. The subjects prepared themselves for the experiment by not exercising or consuming caffeine or alcohol for 12 hours before the experiment.

B. Recording and Measurements

Portability, wearability and comfort are key factors in the potential monitoring of long term stress and therefore the device chosen for this study was the ePatch® wireless wearable two-channel ECG recorder from DELTA Danish Electronics, Light and Acoustics. An example of the ePatch® placement on the sternum can be found in Fig. 1. The ECG was sampled at 512 Hz on both channels. Only the ECG acquired on channel 1 was used for analysis.

C. Experimental Procedure

The experimental procedure was broken down in four main stages: acute stress video, mental stress challenge, neutral video and a post-experimental baseline period. The experiment was precluded by a 10 minute period where ECG was recorded with the ePatch®. The subjects were given very little information about what to expect in the study as to not compromise the effects. Due to this reason, the baseline period was chosen to be at the end of the experiment.

1) Acute Stress Video: The lights were turned off and the window blinds were closed. The subject was seated in front of a computer screen and was given in-ear headphones. A video composed of images and sounds that were graded with a low valence and high arousal was presented to the subject for 6 minutes. The images and sounds changed every 6 seconds.

TABLE I
HRV ANALYSIS

Type	Feature	Description
Linear	Time Domain	Mean, SDNN, SDSD, NN50, pNN50
	Frequency Domain	Total Power, VLF, LFnorm, HFnorm, LF/HF
Non Linear	Approximate Entropy	AppEn(0.2), AppEn(r_{chon})
	Sample Entropy	SampEn(0.2), SampEn(r_{chon})
	DFA	$\alpha 1$, $\alpha 2$
	Correlation Dimension	D2
	Poincaré Plot	SD1, SD2, SD1/SD2

2) Mental Stress Challenge: The subject was assigned a Stroop test [11] as a mental task for a period of 6 minutes.

3) Neutral Video: A video composed of neutrally rated images and sounds was presented to the subject for 6 minutes. The images and sounds changed every 6 seconds.

4) Baseline from Recovery: After the neutral video, the lights were turned on and the subject was told that the experiment was over and that there would be a 10 minute rest period.

D. Analysis

The recorded ECGs were segmented into the four experimental stages: acute stress video, mental stress task, neutral video and a baseline (first 6 minutes of recovery period). These added up to 40 ECG segments, all 6 minutes long. The QRS detection was performed with an in-house algorithm. The HRV was calculated from the R-R (interbeat) time differences and then due to its irregular time sampling, it was resampled (HRV$_r$) by interpolation at 8 Hz.

For each subject, p, the non-linear features were standardized according to Equation 1, where $x_{i,p}$ is the feature value of subject p in stage i, and $\bar{x}_p$ and $x_{std,p}$ is the mean and standard devation of x for subject p within all four stages.

$$x_{stand(i,p)} = \frac{x_{i,p} - \bar{x}_p}{x_{std,p}} \textit{ for i=1,2,3,4} \tag{1}$$

The individual linear and non-linear parameters are described in the following subsections. A summary of all parameters can be found in table I and a flow chart of the data collection, signal processing and classification stages can be found in Fig. 2. All analysis was done using Matlab 7.10.0.499 (R2010a) and reference Matlab code for the non-linear functions can be found in the GitHub repository [12].

1) Linear Features: The time domain features were chosen to be the mean, standard deviation (SDNN), standard deviation between consecutive interbeat differences (SDSD), number pairs of adjacent interbeat intervals differing by more than 50 ms (NN50) and the percentage of NN50 from all the interbeat intervals (pNN50) [3].

The frequency domain features computed from the HRV$_r$ power spectrum density (PSD) were the normalized LF

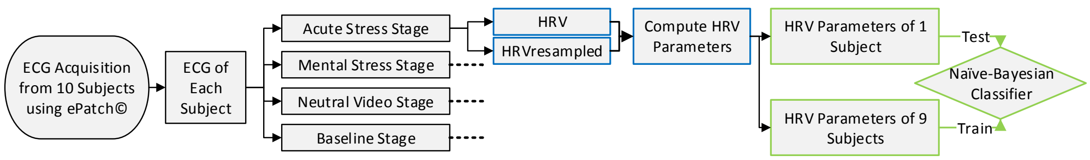

Fig. 2. Flow chart of data acquisition (black), signal processing (blue) and classification of experimental stages (green). Performance of the Naive Bayesian classifier was determined using the leave-one-out method (leaving out all four stages from one subject as a test and training with the rest).

(0.04-0.15 Hz), normalized HF (0.15-0.4 Hz), LF/HF, Very Low Frequency (VLF, $\leq$0.04 Hz) and total power as in Camm et al. [3]. These power values were calculated by integrating the power along the respective frequency locations of the PSD curve.

2) Non-Linear Features: The non-linear methods of analysis were chosen to be the approximate entropy (AppEn) [13] and sample entropy (SampEn) [14], Detrended Fluctuation Analysis (DFA) [15], Correlation Dimension (D2) [8] and Poincaré plot analysis [16].

The AppEn and SampEn are both measures that estimate the changing complexity of a system. For all entropy calculations, an embedding dimension $m = 2$ was used, while the use of two r (radius threshold) values was tested. The r values used were $r_{0.2} = 0.2SDNN$ (widely accepted as an appropriate r value) and the optimized r_{chon} value suggested by Chon et al. [17]. The HRV$_r$ data was used for the AppEn calculation and the original HRV data was used for the SampEn calculations.

DFA is a measure of long-range correlations in a time-series which in the case of HRV analysis, provides information about long-range correlations between interbeat differences by disregarding trends and nonstationarities in the data. Details on DFA computation can be found in [15]. The two slope parameters from the DFA log-log plot, $\alpha 1$ and $\alpha 2$, which represent short and long term fluctuations, respectively, were used.

The D2 is a measure of the dimensionality of the space a set of random points occupy in space. Details on D2 computation can be found in [18]. The original HRV series was used for the computation with an embedding dimension $m = 1$, a delay $\tau = 1$, for radii thresholds from 0.001 to 2 by steps of 0.0001, and taking the slope from the range -8 to -5.

A Poincaré plot is a common method of adjacent interbeat interval analysis. The plot is formed by a scatter plot of RR_{n+1} vs RR_n, which compares the relationship between adjacent RR interval times. The most common method of quantifying the resultant plot is ellipse fitting and finding the standard deviations, $SD1 = \sqrt{0.5 \cdot SDSD^2}$ and $SD2 = \sqrt{2 \cdot SDNN^2 - 0.5 \cdot SDNN^2}$ [16], which are parallel and perpendicular, respectively, to the line of identity. The SD1/SD2 ratio was also used.

E. Feature Selection and Classification

The classification was done using a Quadratic Discriminant Naive Bayesian classifier. Feature selection was done using a forward sequential search with 100 Monte Carlo repetitions. The performance parameter of the feature selection process was finding the minimum misclassification rate (MCR). The feature selection process was done with the parameters from 8 experimental segments held out (2 from each type). The feature selection process was performed 10 times and the features that occurred most often were included in a reduced feature set. The reduced feature set was then applied to a leave one subject out cross-validation, where all experimental stages from one subject were used as the test while the rest were used as training. An exhaustive approach with the reduced feature set was taken to find the best optimal feature combination which is presented in the results.

TABLE III

CONFUSION MATRIX OF NAIVE BAYESIAN CLASSIFIER USING THE LEAVE-ONE-OUT METHOD FOR 4-WAY DIFFERENTIATION OF STAGES IN EXPERIMENTAL SET-UP USING THREE FEATURES (HFNORM, SAMPEN(0.2) AND $\alpha 1$ - MCR = 20%)

Stage	Acute	Mental	Neutral	Baseline
Acute	**80%**	10%	10%	0%
Mental	20%	**70%**	10%	0%
Neutral	0%	10%	**90%**	0%
Baseline	0%	0%	20%	**80%**

III. RESULTS

Classification of the four experimental stage using the aforementioned HRV parameters was performed. The most commonly selected features in the feature selection process can be found in Table II with the mean and standard deviation values of all subjects in each of the four stages. The feature set that produced the lowest MCR (20%) was composed of HFnorm, SampEn(0.2) and $\alpha 1$. The results acquired by classification using those three features can be found in the confusion matrix in Table III and Box and Whisker plots of the feature values in Fig. 3. The highest recognition rate is 90% for the neutral stage and the lowest recognition rate is 70% for the mental stage. The overall classification rate is 80%.

IV. DISCUSSION

The classification shows high recognition of the neutral video (90%), baseline (80%) and acute video (80%) (Table III). The mental stress task is recognized 70% of the time and is misclassified either as the acute stress or neutral video stages. An interesting observation in the classification is that the mental task, acute stress video and neutral video are never misclassified as baseline. Baseline was misclassified 20% of the time as neutral, but that seems to be due to an outlier case

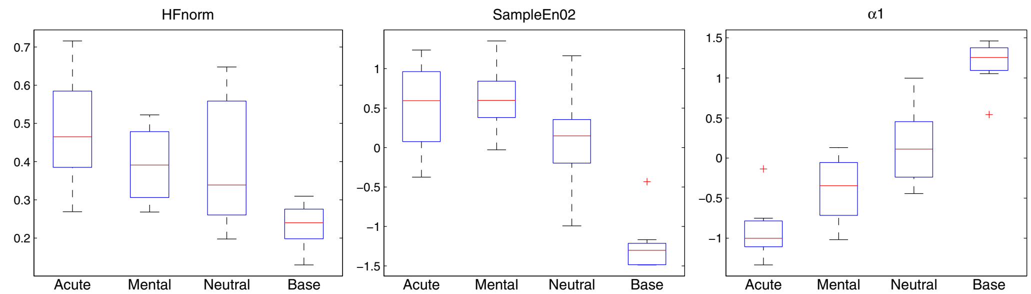

Fig. 3. Box and Whisker plot of the HFnorm, SampEn(0.2) and $\alpha1$ features used of classification for four experimental stages.

TABLE II
MEAN AND STANDARD DEVIATION VALUES OF SUBJECTS FOR EACH EXPERIMENTAL STAGE OF MOST COMMON FEATURES PRODUCED FROM THE FEATURE SELECTION PROCCESS.

Stage	LF/HF	HFnorm	AppEn(0.2)	SampleEn(0.2)	SampleEn(r_{chon})	$\alpha1$	D2
Acute	1.13 +/- 0.67	0.48 +/- 0.13	0.65 +/- 0.34	0.52 +/- 0.52	-0.33 +/- 0.70	-0.94 +/- 0.34	0.25 +/- 0.74
Mental	1.38 +/- 0.57	0.39 +/- 0.09	0.59 +/- 0.42	0.64 +/- 0.39	-0.61 +/- 0.44	-0.39 +/- 0.40	0.56 +/- 0.72
Neutral	1.84 +/- 1.21	0.39 +/- 0.16	0.08 +/- 0.53	0.10 +/- 0.57	-0.04 +/- 0.65	0.14 +/- 0.48	-0.05 +/- 0.84
Base	3.5 +/- 1.39	0.23 +/- 0.06	-1.31 +/- 0.10	-1.25 +/- 0.31	0.98 +/- 0.79	1.20 +/- 0.27	-0.76 +/- 0.72

that can be seen in the baseline stage of SampEn(0.2) and $\alpha1$ Box and Whisker plots in Fig. 3. Aside from this outlier in the baseline stage, it appears to be the most differentiable for all three features and also has a much smaller variability, which emphasises the importance of a good baseline.

Two important adjustments were made in the process of improving the classification. The first was the inclusion of the data from the baseline stage, which was at first not included, and the second was the standardization of the non-linear parameters. Before standardization, features among different subjects could all change proportionally between experimental stages, but have different absolute values or step sizes. Standardizing allowed for better inter-subject comparison and better overall classification results. Standardization was possible because all subjects underwent the same experimental procedure, with the same stages. The baseline ECG segment was also a key factor in the standardization procedure because once the baseline was added, a significant increase in classification rates was achieved.

Current research is focusing on adding more subjects to the study to increase statistical significance and also applying this methodology and feature set to determine relevant features in longer term chronic stress situations.

REFERENCES

[1] S. Cohen, D. Janicki-Deverts, and G. E. Miller, "Psychological stress and disease," *Jama*, vol. 298, no. 14, pp. 1685–1687, 2007.

[2] R. E. Klabunde, "Autonomic innervation of the heart and vasculature," March 2008. [Online]. Available: http://www.cvphysiology.com/Blood%20Pressure/BP008.htm

[3] A. J. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[4] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, "Effects of stress on heart rate complexityA comparison between short-term and chronic stress," *Biological Psychology*, vol. 80, no. 3, pp. 325–332, Mar. 2009.

[5] A. Lansdown and D. A. Rees, "The sympathetic nervous system in polycystic ovary syndrome: a novel therapeutic target?" *Clinical endocrinology*, vol. 77, no. 6, pp. 791–801, 2012.

[6] N. Hjortskov, D. Rissen, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *European Journal of Applied Physiology*, vol. 92, no. 1-2, pp. 84–89, 2004.

[7] G. Valenza, A. Lanata, and E. P. Scilingo, "The role of nonlinear dynamics in affective valence and arousal recognition," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 237–249, Apr. 2012.

[8] P. Melillo, M. Bracale, L. Pecchia *et al.*, "Nonlinear heart rate variability features for real-life stress detection. case study: students under stress due to university examination," *Biomed Eng Online*, vol. 10, no. 1, p. 96, 2011.

[9] P. J. Lang, M. M. Bradley, B. N. Cuthbert, "International affective picture system (iaps): Affective ratings of piture and instruction manual. technical report a-8," University of Florida, Tech. Rep., 2008.

[10] M. M. Bradley, P. J. Lang, "International affective digital sounds (2nd edition; iads-2): Affective ratings of sounds and instruction manual. technical report b-3," University of Florida, Tech. Rep., 2007.

[11] J. H. Krantz, "Stroop experiment," May 2013. [Online]. Available: http://psych.hanover.edu/JavaTest/CLE/Cognition/Cognition/stroop_instructions.html

[12] G. Tanev, "Non-Linear HRV Analysis," https://github.com/gtdevel/non-linear, 2013.

[13] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[14] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[15] T. Penzel, J. W. Kantelhardt, L. Grote, J.-H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *Biomedical Engineering, IEEE Transactions on*, vol. 50, no. 10, pp. 1143–1151, 2003.

[16] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability?" *Biomedical Engineering, IEEE Transactions on*, vol. 48, no. 11, pp. 1342–1347, 2001.

[17] K. Chon, C. Scully, and S. Lu, "Approximate entropy for all signals," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 28, no. 6, pp. 18–23, 2009.

[18] R. Carvajal, N. Wessel, M. Vallverdú, P. Caminal, and A. Voss, "Correlation dimension analysis of heart rate variability in patients with dilated cardiomyopathy," *Computer Methods and Programs in Biomedicine*, vol. 78, no. 2, pp. 133–140, 2005.