Prediction of Stress and Mental Workload during Police Academy Training Using Ultra-Short-Term Heart Rate Variability and Breathing Analysis

Abhishek Tiwari[1], Raymundo Cassani[1] Jean-Francois Gagnon[2],
Daniel Lafond[2], Sebastien Tremblay[3] and Tiago H. Falk[1]

Abstract*— Heart rate variability (HRV) has been studied in the context of human behavior analysis and many features have been extracted from the inter-beat interval (RR) time series and tested as correlates of constructs such as mental workload, stress and anxiety. Such constructs are crucial in assessing quality-of-life of individuals, as well as their overall performance when doing critical tasks. Most studies, however, have been conducted in controlled laboratory environments with artificially-induced psychological responses. While this assures that high quality data are collected, the amount of data is limited and the transferability of the findings to more ecologically-appropriate settings remains unknown. Additionally, it is desirable for such mental state monitoring systems to have high temporal resolution, thus allowing for quick feedback and adaptive decision making. In this article, we explore the use of features computed from time windows much shorter than typically reported in the literature. More specifically, we evaluate the potential of HRV and breathing features computed over so-called ultra-short-term segments (i.e., < 5 minutes) for stress and mental workload prediction. Experiments with 27 police academy trainees show that short time windows as low as 60 seconds can provide useful insights, in particular for mental workload assessment. Moreover, the fusion of HRV and breathing features showed to be an important aspect for reliable behavioural assessment in highly ecological settings.

I. INTRODUCTION

Police officers and first responders are regularly exposed to high stress and mental workload situations on a day-to-day basis. These situations can have both short (decrease in job performance and decision making abilities [1]) and long term consequences (stress induced health disorders [2]) on them. The increasing popularity of commercially available wearable devices for bio-signal monitoring has opened the possibility of monitoring these quality-of-life metrics non-intrusively using physiological correlates. One of the most popular modality has been the heart rate variability (HRV) which has correlates with various cognitive constructs, such as mental workload [3], psycho-social workload [4] (i.e., job stressors), stress [5], [6] , and anxiety [7], to name a few.

HRV is an indicator of the changes in the autonomic nervous system and has traditionally been quantified using time- and/or frequency-domain features. These features quantify the balance between the sympathetic and parasympathetic nervous systems [8]. The computation of these features requires the extraction of the inter-beat interval (RR) time series, commonly extracted from the peaks of the QRS complex of an electrocardiogram (ECG) signal, or from peaks of pulses measured in a photoplethysmogram (PPG). Conventional clinical assessment of HRV has typically relied on long duration time windows, around 24 h. Short-term HRV analysis, in turn, has explored time durations as little as 5 minutes and shown to achieve useful results [8].

Similarly, breathing parameters reflect the sympathetic and parasympathetic balance and have been noted to change with mental workload and stress [9]–[11], and other emotional states [12]. Recently, short-term breathing features have been used for stress prediction outside laboratory settings [11]. Typically, analysis window sizes similar to those used in short-term HRV analysis have been used [13], [14].

Though short-term analysis of cardiac and respiratory processes have shown useful for offline behavioural analyses, several applications exist in which faster time responses are needed [5], especially in life-saving situations that first responders face on a weekly basis. To this end, so-called ultra-short-term HRV analyses have been explored in which window durations smaller than 5 minutes are used. While some applications have been reported in the literature (e.g., [5], [6], [15]), these have relied in controlled laboratory environments, thus the transferability to highly ecological settings may be unwarranted.

Moreover, ultra-short-term analysis for alternate physiological modalities (e.g., breathing) has not been widely explored. Since short-term breathing and HRV have shown to be complementary modalities for behavioural assessment, here, we explore the use of HRV and breathing features computed from ultra-short-term time windows for the prediction of stress and mental workload. Experimental results with police academy trainees as they completed a 15-week lecture/hands-on course show that reliable results can be achieved with windows sizes as low as 1 minute, especially for mental workload assessment. Fusion of the two modalities was shown to outperform short-term analysis using the modalities individually. Overall, these findings suggest that ultra-short-term analysis of cardio-pulmonary signals can provide reliable correlates of stress and mental workload.

The remainder of this paper is organized as follows. Section II covers the materials and methods. Section III presents the experimental results and discussion. Lastly, Section IV presents the conclusions.

II. MATERIALS AND METHODS

A. Data Collection

Data was collected from 27 (6 females) police trainees taking a 15-week course at the Quebec National Police

Author affiliation: [1] Institut national de la recherche scientifique, INRS-EMT, Montreal, Canada, [2] Thales Research and Technology Canada, Québec, Québec, Canada, [3] Université Laval, Québec, Québec, Canada

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

TABLE I

SESSION DESCRIPTION OF THE SHOOTING RANGE EXERCISE

Session number	Description
1	Firearm handling & blank cartridge
2	Basic stance
3	While covered & kneeling stance
4	Displacement, rapid, flashlight, decision making
5	Recap

Academy (École Nationale de Police du Québec, ENPQ, Nicolet, Canada). The course included training and evaluation of various police-related skills (e.g., police car driving, hand-to-hand combat, arrest scenarios, criminal investigation, firearm usage). Participants were recruited at the beginning of the semester and the experimental protocol was approved by the Ethics Review Boards of the Institut National de la Recherche Scientifique (INRS), Université Laval, and ENPQ.

Data was collected in three waves. The first wave consisted of a longitudinal data collection in which participants were asked to wear a heart rate wrist monitor (Fitbit) at all possible times throughout the duration of the 15 weeks of the course. This was used to gauge a baseline HRV for each participant. The second wave, in turn, consisted of a shooting range exercise where participants were trained on the handling of firearms and related skills. In total five sessions (3 hour each) were held, each detailing a separate aspect of the task at hand. More details about the five sessions can be seen in Table I. Participants in the second wave were also asked to use a wearable device (BioHarness 3, Zephyr) that collected ECG (sampled at 250 Hz) and breathing curves (sampled at 18 Hz), which were all time aligned and streamed using the MuLES data acquisition software [16]. The exercise proceeded as it normally would, with the exception of an additional questionnaire at the end of each session in which the trainees reported their stress and mental workload ratings using a French version of the NASA-TLX [17] and fatigue using the Borg scale [18].

The final data collection wave was performed during the intervention simulator exercise. This exercise extends the shooting range one by focusing on aspects of decision making and intervention when a firearm is involved. These exercises follow standard classroom format along with hands-on trainee participation in a simulation environment involving teams of 1-4 people. The simulation was carried out using the setup depicted by Fig. 1. The simulation scenario was displayed via a video projector (shown in blue in the figure) and the trainees were free to interact with characters on the screen (e.g., suspects, victims, witnesses) and to move around in the simulation area as they wish, as well as to take cover behind different barricades (shown in red). To make the exercise more realistic, low velocity automated guns with non-harmful bullets were used during the simulation. Sensors embedded into the simulation area were used to record the timing and location of the shots taken by each trainee. During the simulation, the instructor could manipulate suspect behavior ranging from cooperative to hostile, thus directly modulating and/or responding to trainee actions and decisions. Each simulation lasted for several minutes. Similar to the shooting range exercise, trainees wore the Bioharness device, thus ECG and breathing curves were collected.

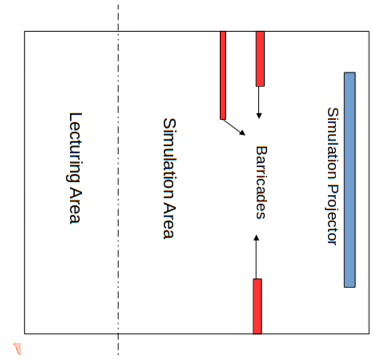

Fig. 1. Setup for the intervention simulation area at ENPQ

TABLE II

DIFFERENT GROUPS OF HRV AND BREATHING FEATURES TESTED

HRV -Time domain
mean · standard deviation · rmssd · pNN50 · coefficient of variation · mean of 1^{st} diff. · std. dev. of abs. of 1^{st} diff. · normalized mean of abs. 1^{st} diff.
HRV -Frequency domain
High freq. power (HF) · normalized HF · low freq. power (LF) · normalized LF · very low freq. power (VLF) · HF/LF
Breathing -Time domain
mean · standard deviation · range · skewness · kurtosis · mean of 1^{st} diff.
Breathing -Frequency domain
Band power in 5 equal bands from 0 to 1 Hz · Band power ratio low freq (0.05-0.25 Hz) and high freq (0.25-0.5 Hz) · spectral centroid · breathing rate (spectral peak)

B. Signal processing pipeline and feature extraction

ECG data was visually inspected and subjects whose clear QRS was missing were removed along with the corresponding breathing data. Following this, a simple pre-processing step using a bandpass filter (5-25 Hz) was performed on the ECG signal. The RR series was then extracted using an energy-based QRS detection algorithm, which is an adaption of the popular Pan-Tompkins algorithm [19]. The RR series was further filtered to remove outliers using range-based detection ($\geq 280ms$ and $\leq 1500ms$), moving average outlier detection, and a filter based on percentage change in consecutive RR values ($\leq 20\%$). The ECG processing was done using the toolbox introduced in [20].

From the enhanced RR series, standard time- and frequency-domain HRV features were extracted. These features, when computed from short-term durations less than 5 minutes, have been shown in the literature to correlate with mental workload [21] and stress [5]. Complete details about these measures can be found in [8].

In turn, the breathing raw signal was first downsampled to 6 Hz, then low-pass filtered to remove noise using a Chebychev 8^{th} order filter with a 2 Hz cutoff frequency. Finally, several time- and frequency-domain features were extracted from the enhanced breathing curve. A complete list of breathing and HRV features is presented in Table II. Overall, 15 HRV and 14 breathing features were computed for different window sizes (60, 90, 120, 180, 240, and 300 s), without overlap between consecutive windows.

C. Stress and mental workload classification

For evaluation, a 5-fold cross-validation setup was used. Here, stress and mental workload assessment was performed as a subject-wise binary classification task, where a classifier was trained to classify signal segments as high or low stress/mental workload. Binarization of the high/low labels was performed per subject and was based on the subject's reported NASA-TLX ratings. A support vector machine (SVM) classifier with a radial basis function (RBF) kernel was used. To explore the generalization performance of the classifier, the 5-fold test was repeated 50 times with different random seeds. To account for dataset class imbalance, we use balanced accuracy (BAC) as the performance figures-of-merit. Moreover, to assess feature importance recursive feature elimination was performed using the extra trees classifier (ETC) [22]. The sci-kit learn implementation of the SVM classifier and the ETC feature selection algorithm was used [22]. Overall, classifiers were trained with the top-15 features from the HRV, from the breathing, as well as from the fused feature sets.

A recent focus has been to assess the stability of the ultra-short term HRV estimates with respect to the 5 minute estimate [5], [15]. In order to explore which features stood out for each ultra-short-term duration and classification task, we analyze the list of most frequently-selected features (features appearing at-least 80% of the time across the 250 trials). Moreover, to assess stability across window duration, we also analyze the features which showed to be important across all of the window durations.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Classification results for stress and mental workload are shown in Tables III and IV, respectively. For stress, we can observe an increase in performance as segment duration increases, going from BAC=0.557 for 60-second segments to BAC=0.594 to 5-minute segments, thus indicating a loss of roughly 7% in accuracy from an ultra-short-term analysis. For breathing, on the other hand, the changes are smaller and go from BAC=0.541 (60 s) to 0.560 (300 s), thus suggesting a loss of 3.5%. When both modalities are fused, substantial improvement is seen in terms of BAC, thus further corroborating the complementarity of the two modalities [13], even at ultra-short-term analyses. Overall, for when used the fused set, a BAC=0.602 is achieved for 60 s ultra-short-term segments, relative to BAC=0.633 for 5-minute segments (i.e., a 5.1% drop). Note that the results achieved with the fused set under ultra-short-term analysis slightly outperform the HRV and breathing results attained with short-term analysis, thus showing the importance of the multimodal approach for applications that rely on very fast decision making.

TABLE III
PERFORMANCE COMPARISON FOR STRESS PREDICTION FOR VARYING WINDOW DURATIONS.

	Window duration (s)					
Modality	60	90	120	180	240	300
HRV	0.557	0.558	0.567	0.570	0.576	0.594
Breathing	0.541	0.545	0.548	0.547	0.536	0.560
Fused	0.602	0.608	0.616	0.617	0.619	0.633

TABLE IV
PERFORMANCE COMPARISON FOR MENTAL WORKLOAD PREDICTION FOR VARYING WINDOW DURATIONS.

	Window duration (s)					
Modality	60	90	120	180	240	300
HRV	0.561	0.557	0.569	0.569	0.566	0.579
Breathing	0.536	0.545	0.538	0.542	0.555	0.546
Fused	0.597	0.599	0.599	0.605	0.607	0.610

For mental workload assessment, the impact of ultra-short-term analysis seems to be even less pronounced than for stress. This may be due to the fact that stress is closely related to a sympathetic response [21] of the nervous system, which typically manifests itself in the lower frequency component of the heart rate [8], [21], which is harder to estimate using ultra-short-term HRV segments [5]. As such, a BAC=0.561 is achieved for HRV segments of 60 s, whereas BAC=0.579 for 300 s (3.2% drop). Similar findings are seen for the breathing features, where a drop of 1.8% is seen between 1- and 5-minute segments. Again, as was the case with stress, fusion of cardiac and pulmonary information showed useful and substantial improvements were seen with the fused feature set. Overall, a BAC=0.597 could be achieved with a 1-minute segment, thus comparing favourably to BAC=0.610 achieved with 5 minutes (i.e., a drop of 2.2%). It is important to note that this ultra-short-term analysis resulted in accuracy that outperformed HRV and breathing classifiers using 5-minute durations, thus further corroborating the usefulness of the proposed method in adaptive operational settings.

Table V lists the top consistent features for stress and mental workload. As can be seen the 'meanRR' and 'rmssd' features showed up in both cases. These features have been reported in the stress measurement literature in controlled environments [5], [6]. Stress is associated with a parasympathetic withdrawal along with a sympathetic activation, which can observed by changes in the high frequency component of HRV [21], which is correlated to rmssd [8]. In turn, the top breathing features have been shown previously to correlate to different emotions [12]. Breathing power spectrum shifts towards larger frequencies during stress along with changing breathing patterns have been reported in [14], [23].

For mental workload, recent reports have shown a lack of sympathetic activation with HRV for long term continuous mental workload task while observing a parasympathetic

TABLE V

CONSISTENT FEATURES FOR STRESS AND MENTAL WORKLOAD

Stress	Mental Workload
meanRR	meanRR
rmssd	rmssd
mean of 1^{st} diff	hf
normalized mean of abs 1^{st} diff	normalized mean of abs 1^{st} diff
breathing rate	br skewness
br skewness	br power (0-0.2Hz)
br power ratio	br power (0.2-0.4Hz)
br power (0.2-0.4Hz)	br power (0.4-0.6Hz)

withdrawal with rmsdd feature appearing as relevant over the full length of task [3]. The parasympathetic withdrawal also explains the occurrence of the high frequency components as a consistent feature. Various breathing band powers have shown to be consistent over all segment durations. This can explained by the fact that overall respiratory variability is reported to decrease along with increase in both respiration rate and sigh rate during mental tasks [10], [24]. These factors could shift the spectrum towards higher frequencies, along with a decrease in the spread of spectral peak captured by various power bands. Finally, breathing skewness appears to be a consistent feature for both constructs. A possible explanation could be related to increased sigh rates reported for both stress and mental workload [9], [10]. A sigh is characterized by a deep inhalation followed by slow exhalation, which could skew the amplitude distributions.

IV. CONCLUSIONS

In this work, we explored ultra-short-term analysis of heart rate variability and breathing signals for fast prediction of mental workload and stress in highly ecological settings. Experimental results involving police academy trainees during their hands-on exercises showed that ultra-short-term analysis on individual modalities results in performance losses lower than 7% when compared to short-term analyses. The losses were lower for mental workload than for stress prediction. Notwithstanding, when HRV and breathing features were fused, ultra-short-term analysis outperformed short-term analysis of individual modalities and remained inline with the performance achieved when both modalities were combined.

REFERENCES

[1] S. K. Nisar and M. I. Rasheed, "Stress and performance: Investigating relationship between occupational stress, career satisfaction, and job performance of police employees," *Journal of Public Affairs*, p. e1986, 2019.

[2] C. Regehr and V. R. LeBlanc, "Ptsd, acute stress, performance and decision-making in emergency service workers," *J Am Acad Psychiatry Law*, vol. 45, no. 2, pp. 184–192, 2017.

[3] G. Chaumet, A. Delaforge, and S. Delliaux, "Mental workload alters heart rate variability lowering non-linear dynamics," *Frontiers in physiology*, vol. 10, p. 565, 2019.

[4] F. Togo and M. Takahashi, "Heart rate variability in occupational health-a systematic review," *Industrial health*, vol. 47, no. 6, pp. 589–602, 2009.

[5] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term hrv features as surrogates of short term hrv: a case study on mental stress detection in real life," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–13, 2019.

[6] M. Zubair and C. Yoon, "Multilevel mental stress detection using ultra-short pulse rate variability series," *Biomedical Signal Processing and Control*, vol. 57, p. 101736, 2020.

[7] W.-H. Wen *et al.*, "Toward constructing a real-time social anxiety evaluation system: Exploring effective heart rate features," *IEEE Transactions on Affective Computing*, 2018.

[8] A. Camm *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.

[9] E. Vlemincx, I. Van Diest, and O. Van den Bergh, "A sigh of relief or a sigh to relieve: The psychological and physiological relief effect of deep breaths," *Physiology & behavior*, vol. 165, pp. 127–135, 2016.

[10] E. Vlemincx, J. Taelman, S. De Peuter, I. Van Diest, and O. Van Den Bergh, "Sigh rate and respiratory variability during mental load and sustained attention," *Psychophysiology*, vol. 48, no. 1, pp. 117–120, 2011.

[11] A. Tiwari, S. Narayanan, and T. H. Falk, "Breathing rate complexity features for "in-the-wild" stress and anxiety measurement," in *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 2019, pp. 1–5.

[12] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[13] A. Tiwari, J. L. Villatte, S. Narayanan, and T. H. Falk, "Prediction of psychological flexibility with multi-scale heart rate variability and breathing features in an "in-the-wild" setting," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2019, pp. 297–303.

[14] M. Parent, A. Tiwari, I. Albuquerque, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "A multimodal approach to improve the robustness of physiological stress prediction during physical activity," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 4131–4136.

[15] L. Pecchia, R. Castaldo, L. Montesinos, and P. Melillo, "Are ultra-short heart rate variability features good surrogates of short-term ones? state-of-the-art review and recommendations," *Healthcare technology letters*, vol. 5, no. 3, pp. 94–100, 2018.

[16] R. Cassani, H. Banville, and T. H. Falk, "Mules: An open source eeg acquisition and streaming server for quick and simple prototyping and recording," in *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, 2015, pp. 9–12.

[17] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[18] G. Borg, *Borg's perceived exertion and pain scales.* Human kinetics, 1998.

[19] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[20] J. A. Behar, A. A. Rosenberg, I. Weiser-Bitoun, O. Shemla, A. Alexandrovich, E. Konyukhov, and Y. Yaniv, "Physiozoo: A novel open access platform for heart rate variability analysis of mammalian electrocardiographic data," *Frontiers in physiology*, vol. 9, p. 1390, 2018.

[21] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis," *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 2015.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[23] S. Gandhi, M. S. Baghini, and S. Mukherji, "Mental stress assessment-a comparison between hrv based and respiration based techniques," in *2015 Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 1029–1032.

[24] M. Grassmann, E. Vlemincx, A. von Leupoldt, and O. Van den Bergh, "The role of respiratory measures to assess mental load in pilot selection," *Ergonomics*, vol. 59, no. 6, pp. 745–753, 2016.