Heart Rate Variability and its Association with Second Ventilatory Threshold Estimation in Maximal Exercise Test

Iman Alikhani, Kai Noponen, Mikko Tulppo, Juha Peltonen, Elias Lehtonen, Tapio Seppänen

***Abstract*— During incremental exercise, two ventilatory thresholds (VT1, VT2) can normally be identified from gas exchange and ventilatory measurements, such as oxygen uptake, carbon dioxide production and ventilation. In this paper, we attempt to estimate the VT2 using HRV indices derived from a wearable electrocardiogram during a maximal exercise test. The exercise test is conducted on a treadmill that raises its speed by 0.5 km/h every minute. We have 42 measured exercise tests from 24 healthy male volunteers. Three experts determined the VT2 in each exercise test independently and we used principal component subspace reconstruction of their determinations to compute a collective VT2 for our machine learning model. The results demonstrate that the VT2 can be estimated from HRV using the proposed method with a reasonable performance during a maximal exercise test. In 28 out of 42 exercise tests, the HRV-derived threshold (HRVT) is within a minute (one phase) of the collective expert's determination.**

I. Introduction

Second ventilatory threshold (VT2) is defined as the point during exercise at which the ventilation starts to rise at a faster rate than the oxygen uptake [2]. There are various methods in the literature to determine the VT2 from the gas exchange variables [3,4,5]. The methods are typically experimented on an incremental exercise protocol, for instance a treadmill or bicycle ergometer [3,5] with an intensity progression rate.

Heart rate variability (HRV) is constantly decreasing during incremental exercise tests [10] and estimation of the VT2 based on HRV has been studied in the literature. Some methods find an association between the variability of (an) HRV index (indices) with the VT2 [6, 7]. However, this can be conducted using machine learning models, to allow the model to train itself using descriptive features and markers of HRV. A recent study [8] used deep learning models to estimate the VT2 using single-lead electrocardiography signals.

In this paper, we aim to propose a method for estimation of the VT2 using HRV. Using the proposed machine learning approach, we classify one-minute episodes of the run (which we call a phase in this paper) independently to either before-VT2 or after-VT2 and then we find the HRV-derived threshold (HRVT) that is associated with the VT2. To validate our method, we compare the HRVT with the VT2 determined by the experts on a dataset with 42 runs.

II. Data

The database was originally designed to evaluate the effects of physical activity on exercise capacity [1]. However, it fits the purpose of this article to study the estimation of the VT2 using HRV indices.

A. Graded Maximal Exercise Test

The participants performed a graded treadmill maximal exercise test, starting at the speed of 8 km/h and increasing every minute by 0.5 km/h until the voluntary declaration of exhaustion. Ventilation and gas exchange were reported as a mean value per one-minute-phase measured using an M909 ergospirometry (Medikro, Kuopio, Finland). HRV was measured simultaneously using Polar RS800 (Polar Electro Oy, Kempele, Finland) during the exercise test.

B. Participants

The subjects underwent the measurement on two occasions, once before and once after a two-month training program. Entirely we have 42 runs of exercise tests collected from 24 healthy male volunteers [Age: 29.8 ± 2.8 years, body mass index (BMI): 25.0 ± 2.5 kg.m^{-2}, maximum oxygen uptake (VO_2 max): 55.9 ± 6.9 ml/kg/min].

The participants were not allowed to drink and eat three hours before the measurement, and they were instructed not to consume alcoholic beverages from the preceding test day. The participants underwent a resting ECG test using a 12-lead standard ECG and a physician confirmed their cardiac health.

The study was performed according to the Declaration of Helsinki, the local ethical committee of the Hospital District approved the protocol, and all the participants gave their written informed consent.

C. Determination of Second Ventilatory Threshold

VTs were determined independently by three experts using ventilatory indices that were measured by the spirometer. The expert's determinations of the VT2 phases are depicted in Figure 1. Figure 1 shows that the VT2 determination is not an unchallenging task even from the ventilatory signals, as there are cases with more than a couple of phases of difference in the experts' opinions. This variability though on the determination of VT2 is not an unknown matter within the sports physiology discipline [9]. The experts determined the

Iman Alikhani (corresponding author), Kai Noponen and Tapio Seppänen are with the Center for Machine Vision and Signal Analysis of the University of Oulu, P.O. BOX 8000, FI-90014 Finland (e-mail: iman.alikhani@oulu.fi, kai.noponen@oulu.fi, tapio.seppanen@oulu.fi).

Mikko Tulppo is with the Research Unit of Biomedicine, Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, P.O. Box 5000, FI-90014, Finland (e-mail: mikko.tulppo@oulu.fi)

Juha Peltonen and Elias Lehtonen are with HULA – Helsinki Sports and Exercise Medicine Clinic, Foundation for Sports and Exercise Medicine, Mäkelänkatu 47, 00550 Helsinki, Finland. Juha Peltonen is also with the Department of Sports and Exercise Medicine, Clinicum, University of Helsinki, Mäkelänkatu 47, 00550 Helsinki, Finland. (e-mail: juha.peltonen@helsinki.fi, elias.lehtonen@hula.fi)

first ventilatory threshold (VT1) by the V-slope method (i.e., identifying a consistent increase in the CO_2 output/O_2 uptake relationship), and VT2 was primarily determined by identifying a consistent increase in the ventilation/CO2 output relationship. In addition, an increase in the ventilation/O2 uptake relationship was used in both VTs.

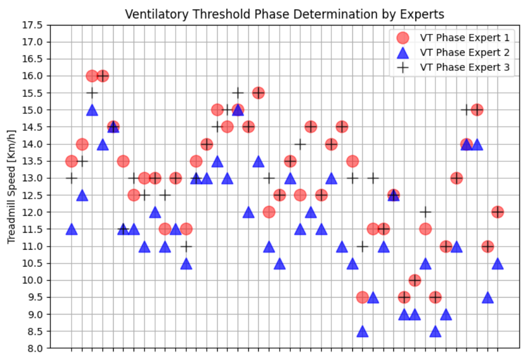

Figure 1. In this figure, the expert's determinations of the VT2 phase are depicted. The x-axis represents the exercise tests, and the y-axis represents treadmill speed.

III. Method

To estimate the VT2 from HRV, we first use a machine learning model to classify each one-minute phase to either before-VT2 or after-VT2 phase, and in a post-processing step, we define the specific phase, where estimations distinctly switch from before-VT2 to after-VT2 as HRVT (which is the estimated VT2 phase). Therefore, we formulate the problem to 1) a binary classifications of each minute of the exercise test and 2) defining the HRVT which is associated with VT2. We describe the process in the following subsections.

A. Heart Rate Variability Preprocessing

Beat-to-beat interval known as RR interval (RRI) is derived by the heart rate monitoring device and streamed out. RRI time series typically contain outliers and artifacts. Considering that our measurement is conducted in an exercise context, the HR is higher and HRV is lower than when resting, and in most cases the amount of motion artifacts is higher than in the resting state measurements. Hence, we remove RRI values over 1700 ms or under 250 ms (outliers) and the ones with more than 70 ms difference to the adjacent values (artifacts). On average, the criteria above accounted for 2.5% of the RRI values in our dataset, and we replaced those with linear interpolation using their adjacent RRI values. Given that the RRI time series is not naturally equidistantly sampled, we linearly interpolate it with an 8 Hz sampling rate for power spectral density (PSD) estimation.

B. Feature Extraction

We extracted a set of features for each one-minute phase of the exercise test. Each run of the exercise test then contains a feature matrix of size X_m^n, where m is the number of features, and n is the number of phases the subject ran in that particular exercise test. From each one-minute phase of the exercise test, we derive the following feature set:

- Time-domain features: Mean heart rate (HR_{mean}) and the root mean square of successive differences between normal heartbeats (rMSSD)
- Frequency-domain features: High (HF; 0.15-1 Hz) and low (LF; 0.04-0.15 Hz) frequency power. Normalized low-frequency power (LF_{nu}) is also added to the feature vector. Note that the HF frequency band is defined higher, due to the measurement context.
- non-linear features: α1 and α2 and their ratio in detrended fluctuation analysis (DFA), plus sample entropy [10].
- Treadmill speed

Figure 2 illustrates the measured data in an example exercise test. In this figure, the transitions between the phases are highlighted with green vertical lines, and the features are extracted from the HRV data depicted in the second last sub-figure. Ventilatory indices are in the top two sub-figures. The m in X_m^n in this sample case is equal to 15, indicating that from this run, we derive 15 samples, out of which seven are labeled with before-VT2 and eight with after-VT2. Respectively, we labeled each phase with a binary flag (bottom sub-figure), where zero corresponds to before-VT2 and one to after-VT2.

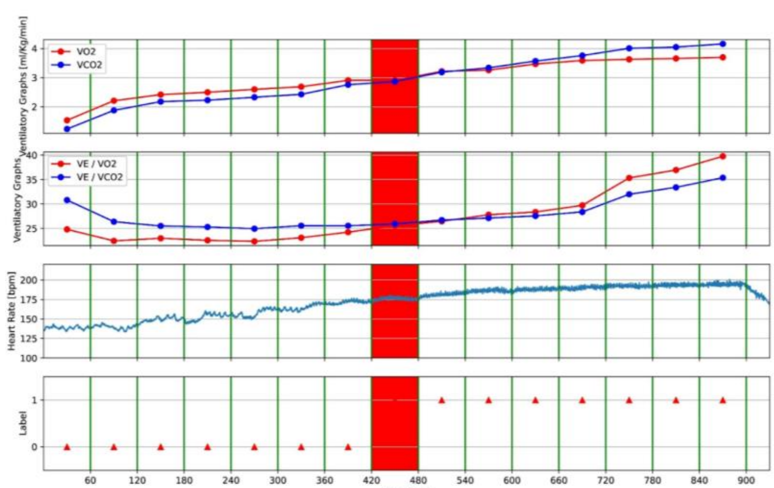

Figure 2 . Sample exercise test. The two top sub-figures depict the ventilatory graphs, using which the experts determined the reference VT2, highlighted as a red phase. Second bottom sub-figure shows the HRV of the subject and the bottom sub-figure shows the respective labels assigned to each phase.

C. Calculation of Collective VT2

We possess three sets of VT2 determinations for each exercise test (as shown in Figure 1), using which we calculate a collective VT2 to draw a single label per exercise test to facilitate our supervised machine learning problem. The collective VT2 phase is computed based on the first principal component (PC) subspace reconstruction of the experts' opinions. The three sets of VT2 phase numbers are normalized for PC analysis and the mean of the expert's average VT2 phase is added to the subspace reconstruction. Given that the subspace reconstruction leads to real numbers with fractional parts, the collective VT2 phase is then calculated by rounding that number to the nearest integer value.

With positive loadings of equal magnitude, the first PC represents the weighted average of the expert VT2 phase determinations that explains the most variance (92,0 %) of the VT2s across the subjects. The second and third PCs represent differences in the opinions between the experts as there are both positive and negative loadings, and they contribute 5.6 % and 2.5 % to the variance, respectively. This way, the collective VT2 can be viewed as the main intrinsic factor behind the expert opinions that best explains the subject-wise differences as a whole but is resilient to the individual characteristics of the experts' opinions.

D. Model Selection and Training

We derived the described features from each one-minute exercise phase and assigned them to a corresponding binary label from the calculated collective labels (either before-VT2 or after-VT2 phase). Each exercise test creates multiple samples for our machine learning training and testing, and to avoid having the data of the same subject in both training and testing sets, we trained the classifier in a four-fold grouped cross-validation setting, where the group is the subject identifiers.

We trained a random forest classifier with 20 estimators and a maximum tree depth of four. The classifier returns a binary output, which leads to a sequence of binary labels for each run. The actual decision on the VT2 location is made in a post-processing step described in the next sub-section.

E. HRVT Determination

As mentioned earlier, the model is formulated to have single-minute HRV indices alongside the treadmill speed as its input and outputs a binary value, corresponding to before-VT2 or after-VT2. But the purpose of this step is to make the final decision on the VT2 based on the output of the classifier for each stage. We call the estimated phase as HRVT phase which stands for HRV-derived threshold and is associated with the phase that the subject reaches to his/her VT2. HRVT is the phase that (at least) the next two consecutive phases are predicted as after-VT2.

F. Evaluation Metrics

To evaluate the results of our study, we use two sets of metrics, including the binary classification evaluation metrics, and the metrics for HRVT estimation. For the classification, we calculate precision, recall, f1-score, and the number of supported samples in each class. For the HRVT estimation, we count the number of subjects, that their HRVT phase is as close as one phase with the collective expert's VT2 phase.

IV. Results

VT2 determination is not a trivial task, even in a maximal exercise test. Disagreements between the three experts that determined the VT2 based on the ventilatory markers (Figure 1) in our study are evidence of the problem complexity. We must accept that there is no sharply defined ground truth for this problem in the described setting, and therefore we should acknowledge its natural consequence in HRV-based determination.

TABLE I. Model Classification Results on Each Fold

Fold	Set	Label	Precision	Recall	f1-score	Support
1	Train	Before VT2	0.95	0.95	0.95	279
		After VT2	0.92	0.93	0.93	195
	Test	Before VT2	0.91	1.00	0.95	92
		After VT2	1.00	0.87	0.93	67
2	Train	Before VT2	0.95	0.97	0.96	277
		After VT2	0.96	0.93	0.95	197
	Test	Before VT2	0.96	0.85	0.90	94
		After VT2	0.82	0.95	0.88	65
3	Train	Before VT2	0.96	0.96	0.96	277
		After VT2	0.95	0.94	0.94	198
	Test	Before VT2	0.91	0.93	0.92	94
		After VT2	0.89	0.86	0.87	64
4	Train	Before VT2	0.97	0.97	0.97	280
		After VT2	0.95	0.95	0.95	196
	Test	Before VT2	0.88	0.96	0.92	91
		After VT2	0.93	0.82	0.87	66

Our result shows a promising estimation of VT2 in most of our sample runs. In 28 out of 42 cases, none or a single wrongly classified phase is found, which means VT2 is estimated in quite good agreement with our collective expert's opinions. As expected, most misclassified phases are in the vicinity of the collective VT2 phase, and the beginning and ending phases are easier for the model to classify correctly.

Table 1 shows the performance of model training on each fold. The figures are separated as per the label and the set on each fold. The figures in the table summarize the binary classification performance of the random forest classifier, which takes a minute-long phase of the exercise test and classifies whether that phase is before-VT2 or after-VT2. There are 59 out of 633 misclassified phases.

We have also depicted the output of the model in Figure 3. There are three sub-figures and each row in this figure represents an exercise test. The left subfigure is the reference according to the collective expert's opinion. The length of each one-minute phase of the run is noted on the top left of the sub-figure in green. We aligned all the exercise tests based on their collective VT2 phase, which is illustrated by a vertical red line. The blue phases are before-VT2, and the yellow phases are after-VT2. The middle sub-figure shows the classification result of the random forest model, and the right sub-figure shows the difference between the reference and classification results. The yellow phases in the right sub-figure are the misclassified ones.

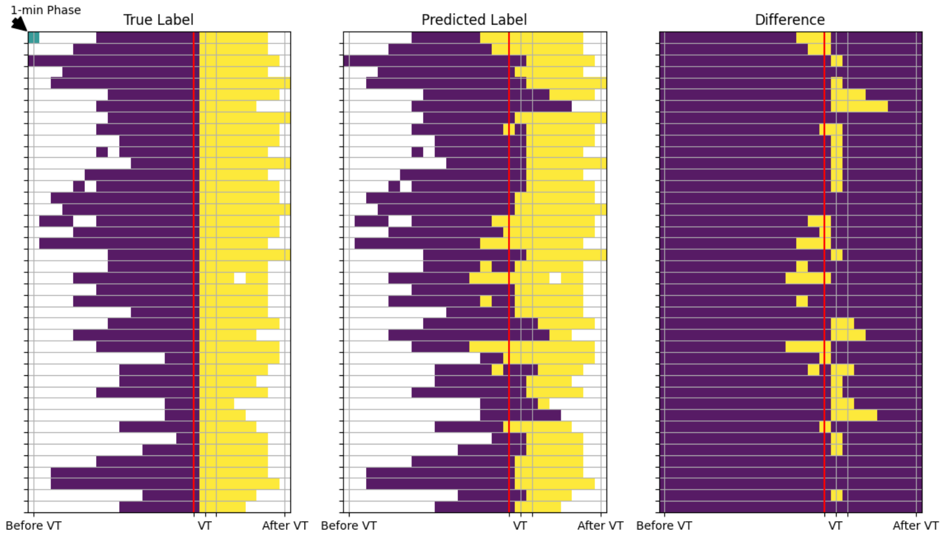

Figure 3 The left sub-figure shows the subject's exercise test aligned based on their collective VT2 threshold (noted by a red line). Blue cells are before VT2 and yellow cells are after VT2 phases. The middle subplot shows the predictions by the trained model in the cross-validated setting. The right subplot shows the difference between the prediction and the expert labels (misclassified phases are in yellow).

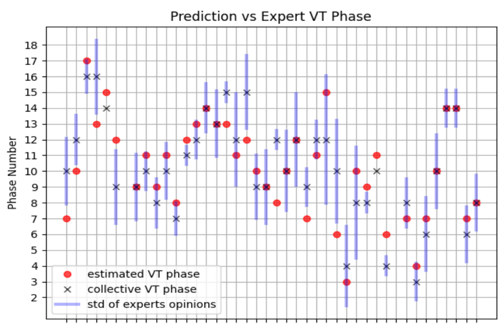

Figure 4 HRVT phase for each test versus collective VT2 phase from the experts. Blue bars are the standard deviation of experts' determinations.

But as described in the previous section (subsection E in Method) the final estimation in this paper is HRVT. From each row of the middle sub-figure of Figure 3, we define HRVT. Figure 4 illustrates how well the HRVT phase falls to the expert's VT2 determinations. In two of the exercise tests, all the phases are predicted as before-VT2 (can be seen in Figure 3 middle sub-figure), which led to unknown HRVT and therefore missing from Figure 4.

V. Conclusion

In this study, we proposed a model to estimate VT2 using HRV in a maximal exercise test. HRV is measured using a wearable device, while the subjects were running on a treadmill, on which the speed increased every minute. We first classify the one-minute phases into a before-VT2 or after-VT2 using a random forest classifier. Then we find the HRVT phase, which is associated with the VT2. The reference VT2s for each exercise test were determined independently by three experts and is aggregated to a collective VT2. The results demonstrated promising figures, both concerning the binary classification of the one-minute phases and the association between HRVT and VT2.

There are limitations in this study, including the fact that the determination of VT using ventilatory markers is still a subjective task. Therefore, the determination of experts might be different as illustrated in Figure 1. VT determinations can be complemented by e.g., blood samples for lactate measurements since lactate accumulation occurs when the runner reaches his/her VT. But taking blood samples during the exercise introduces a pause distortion into the equation which affects cardiac system performance and HRV indices.

References

[1] Hautala, A., Martinmaki, K., Kiviniemi, A., Kinnunen, H., Virtanen, P., Jaatinen, J., & Tulppo, M. (2012). Effects of habitual physical activity on response to endurance training. Journal of sports sciences, 30(6), 563-569.

[2] Reybrouck, T., Ghesquiere, J., Weymans, M., & Amery, A. (1986). Ventilatory threshold measurement to evaluate maximal endurance performance. International journal of sports medicine, 7(01), 26-29.

[3] Gaskill, S. E., Ruby, B. C., Walker, A. J., Sanchez, O. A., SERFASS, R. C., & LEON, A. S. (2001). Validity and reliability of combining three methods to determine ventilatory threshold. Medicine & Science in Sports & Exercise, 33(11), 1841-1848.

[4] Cheng, B., Kuipers, H., Snyder, A. C., Keizer, H. A., Jeukendrup, A., & Hesselink, M. (1992). A new approach for the determination of ventilatory and lactate thresholds. International journal of sports medicine, 13(07), 518-522.

[5] Gregg, J. S., Wyatt, F. B., & Kilgore, J. L. (2010). Determination of ventilatory threshold through quadratic regression analysis. The Journal of Strength & Conditioning Research, 24(9), 2512-2515.

[6] Karapetian, G. K., Engels, H. J., & Gretebeck, R. J. (2008). Use of heart rate variability to estimate LT and VT. International journal of sports medicine, 29(08), 652-657.

[7] Dourado, V. Z., & Guerra, R. L. F. (2013). Reliability and validity of heart rate variability threshold assessment during an incremental shuttle-walk test in middle-aged and older adults. Brazilian Journal of Medical and Biological Research, 46, 194-199.

[8] Miura, K., Goto, S., Katsumata, Y., Ikura, H., Shiraishi, Y., Sato, K., & Fukuda, K. (2020). Feasibility of the deep learning method for estimating the ventilatory threshold with electrocardiography data. NPJ digital medicine, 3(1), 1-7.

[9] Kaczmarek, S., Habedank, D., Obst, A., Dörr, M., Völzke, H., Gläser, S., & Ewert, R. (2019). Interobserver variability of ventilatory anaerobic threshold in asymptomatic volunteers. Multidisciplinary respiratory medicine, 14(1), 1-10.

[10] Tulppo, M. P., Makikallio, T. H., Takala, T. E., Seppanen, T. H. H. V., & Huikuri, H. V. (1996). Quantitative beat-to-beat analysis of heart rate dynamics during exercise. American journal of physiology-heart and circulatory physiology, 271(1), H244-H252.