2018 ACCES23-CMBEC41

Joint Conference
Charlottetown PEI
May 8-11, 2018

# AUTOMATED DETECTION OF ANAEROBIC AND VENTILATORY THRESHOLDS FROM FREE-FORM BIOMETRIC DATA

Matthew Howe-Patterson, Eric Reiher, Matthew R Patterson, Bahareh Pourbabaee, Frederic Benard
*OMsignal Inc.*

## INTRODUCTION

As an athlete, knowing one's ventilatory and anaerobic thresholds (VT and AT respectively) can enable a more effective training regimen [2, 4]. The athlete can optimize his endurance by training at his VT, he can optimize his power and speed by training over his AT. Normally, determining these thresholds requires a subject to go through an incremental exercise protocol while their volume of $O_2$ inhaled (VO2 in L/min) and their total ventilation (VE in L/min) are measured. Changes of slope in the curve of VE vs. VO2, called inflection points are their VT and AT.

In Figure 1, the inflection at the intersection of regression-1 and regression-2 is VT and the inflection at the intersection of regression-2 and regression-3 is AT. The subject’s heart rate is measured throughout the protocol, and the heart rate corresponding to their VT and AT can then be used to inform a training regimen. [3] Figure 1 illustrates the inflection points during an incremental exercise protocol.

OMsignal apparel and algorithms can measure heart rate, breathing rate, and a unit-less correlate of breathing volume, providing the potential to find a subject’s VT and AT without the use of a metabolic cart. In addition, the OMrun platform collects data from runners who are not following any pre-specified protocol, and it is useful to infer these users VT and AT automatically from their free-form data.

Early work towards this goal at OMsignal demonstrated that a human was able to visually identify a users’ VT and AT in free-form data, from a suitably filtered plot of heart rate (HR) vs. ventilation. Since users were not following an incremental exercise protocol, these curves proved difficult to analyze automatically. The primary difficulty is the extremely uneven distribution of points that occurs in each regime. For example, a user might ramp up from walking to jogging speed in less than a minute, spend 50 minutes at that speed, and then may end with a single sprint lasting a minute. In an incremental protocol, all three of these regimes would last the same amount of time. Simple methods to correct for this imbalance, such as weighting each point in a regression by the inverse of its local density tend to emphasize noisy measurements.

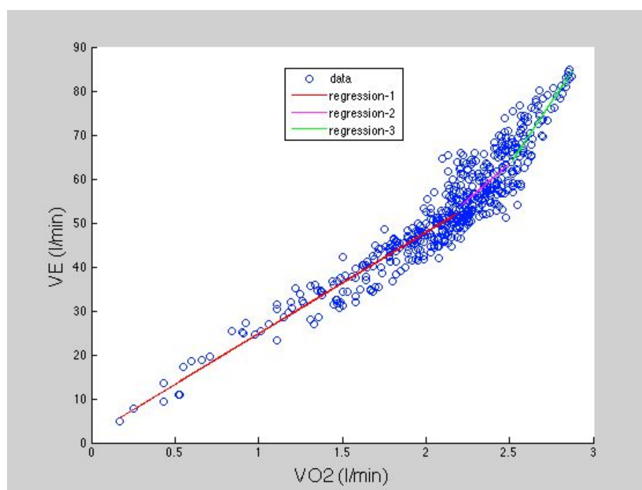


Figure 1: An example of the gold standard method for obtaining VT and AT by plotting ventilation to oxygen consumption for an incremental exercise test.

In light of the fact that humans can visually solve this problem, it can be considered a computer vision problem that can be solved using machine learning techniques. It was found that a system based on 5 free-form runs with minimal instruction to the users can determine VT and AT with an acceptable accuracy.

## DATA

All available accounts in OMsignal's OMrun database, with human annotated VT and AT values were assessed for suitability. These accounts include a range of individuals from amateur runners to competitive long distance runners. The suitability assessment involved looking at the ventilation-HR curve of the first five runs and the manually annotated AT and VT values to determine whether there is a suitable distribution of heart rates and ventilation values, and there are no obvious annotation errors. Some accounts were also rejected due to less than 5 minutes worth of data remaining after the biometric filter, indicating that these users' data suffer from a lot of noise. Finally, 193 accounts were used for algorithm development and validation. These 193 accounts were randomly divided into:

- A training set of 120 accounts (~62%). This data is used to train the machine learning model.
- A validation set of 33 accounts (~17%). This set is used to tune the hyperparameters of the machine learning model.
- A testing set of 40 accounts (~21%). To prevent overfitting, performance on this data is not assessed until the algorithm development is finalized.

In addition, 9 users had their gold standard VT and AT values determined at an exercise physiology lab. 6 of these users are not in any of the above sets. 3 of them are in the training set, although the human annotated values from the free-form data are used as the targets, rather than the gold standard values.

## PREPROCESSING

RR intervals, inhale amplitudes, and inhale to inhale intervals found by OMsignal's biometric algorithms are processed to give one tuple (finite, ordered list of elements) of breathing rate (BR, in breaths per minute), breathing depth (BD, which has no unit), and heart rate (HR, in BPM) for each second in each of the five runs. These are filtered to remove heart rates that are below 110 BPM or above 200 BPM. The unreliable measurements detected by non steady state behaviour (SD of BR > 15 or SD of HR > 10 in a 50 second window) or an implausibly low R peak detection rate (RR coverage value below 80% in a 50s window), are removed. The RR coverage metric is the sum of detected RR durations divided by the total elapsed time in the same window.

## AUGMENTATION

The number of available accounts for training is too small to prevent overfitting. It is also a very biased dataset in which physiologically plausible VT and AT values are not present in the data. Both issues are addressed through data augmentation. The augmentation process, applied to each user's data is as follows:

- Copies of the valid (ventilation,HR) tuples from the first five runs are created by repeatedly drawing bootstrap re-samples of the set of tuples. This introduces plausible variability in the density of the ventilation-HR curve while preserving the general shape and any inflection points.
- Each re-sampled group of five runs are created by shifting the HR values based on a target VT between 120 BPM and 175 BPM. This simulates the counterfactual scenario in which the same curve shape is associated with a different VT and AT value. Only augmented data with a shifted AT <= 195 BPM are generated.

## HEAT MAP EXTRACTION

All five runs are concatenated together. The distribution of these points are approximated with a 2D histogram, to which further processing is applied. The resulting representation is called the heat map. Examples are given in Figures 2 and 3.

1. All tuples with ventilation values in the 1st and 99th percentiles are removed, since these points tend to be outliers which can significantly degrade the quality of the heat map.

2. Ventilation is normalized to lie between 0 and 1, since it is a unit-less quantity in the OMsignal system and its absolute value is not meaningful for AT and VT assessment.
3. The 2D histogram has 45 bins for HR spanning [110,200], and 33 bins for ventilation.
4. The histogram is filtered with a 3x3 moving maximum filter followed by a 3x3 moving average filter.
5. All bins with a density below the maximum entropy density are set to zero.
6. All bins above the 95th percentile of the histogram bin densities are set to that value.
7. The histogram is re-normalized to sum to 1.

Steps 4 to 6 are applied to partially fix the uneven sampling problem described in the introduction.

## MODEL STRUCTURE AND TRAINING PROCEDURE

From the heat maps extracted from the training data, a random forest [1] is used to simultaneously predict an AT and a VT value. The random forest is trained with the following hyperparameters, chosen by trial and error based on the validation data performance:

- Number of Trees: 3000
- Loss Function: mean squared error
- Minimum Samples per Leaf: 100
- Number of Features per Tree: 39

Of these, the Number of Features per Tree was found to have the greatest effect on generalization error.

Once the hyperparameters were finalized, the final model was produced by training on the training and validation data sets, and tested against the testing set and the gold standard set.

## RESULTS

The results are presented in three ways. The 95th percentile absolute error for VT and AT individually are presented in Table 1 for each dataset. The percentage of the data where the predicted VT is within *x* BPM of the visually determined VT, and the predicted AT is within *x* BPM of the visually determined AT, is presented in Table 2 for each dataset.

The mean and worst case errors compared with the gold standard for the 6 unseen cases and 3 training cases where this was available, are presented in Table 3.

Table 1: Individual performance of VT and AT prediction with respect to the visually identified VT and AT values.

| **Data Set** | **95th Percentile of Absolute Error (BPM)** | | | |
|---|---|---|---|---|
| | ***Augmented Data*** | | ***Original Data*** | |
| | ***VT*** | ***AT*** | ***VT*** | ***AT*** |
| Training + Validation | 4.02 | 3.44 | 3.14 | 2.77 |
| Testing | 9.03 | 7.61 | 8.61 | 7.19 |

Table 2: Performance as measured by the percentage of data where both the VT and AT prediction meet a 7.5/10 BPM accuracy target with respect to the visually identified VT and AT values.

| **Data Set** | **% Where VT and AT Jointly Within *x* BPM** | | | |
|---|---|---|---|---|
| | ***Augmented Data*** | | ***Original Data*** | |
| | ***7.5*** | ***10*** | ***7.5*** | ***10*** |
| Training + Validation | 99.43 | 99.95 | 98.69 | 99.35 |
| Testing | 84.63 | 96.37 | 85.00 | 95.00 |

Table 3: Mean and worst case absolute error with respect to the gold standard VT and AT prediction on the data where this is available. The unseen dataset is composed of users who were not included in either the training, validation or testing sets.

| **Gold Standard Comparison** | **Absolute Error Vs The Gold Standard (BPM)** | | | |
|---|---|---|---|---|
| | ***Mean VT Error*** | ***Worst VT Error*** | ***Mean AT Error*** | ***Worst AT Error*** |
| Training | 3.67 | 5 | 3.5 | 4 |
| Unseen | 3.5 | 6 | 5.17 | 9 |

## DISCUSSION

An investigation into the errors on the testing set suggests that improvement is not possible by modifying the machine learning model. Figure 2 shows a heat map representation with a well formed ventilation-HR curve from the testing set. On this example, the manual and algorithmic AT and VT values are very close. Figure 3 shows the heat map for an example with one of the largest VT errors in the testing set. Given the ambiguous nature of the curve, it is unlikely that an algorithm could be reliably more accurate in this kind of situation. A human annotator would have difficulty in this situation as well due to the lack of obvious inflection points.

A limit of this system is that it cannot infer an accurate VT or AT if the user does not surpass this level in any of the 5 runs. A significant percentage of OMrun users only ever used the system while walking or lightly jogging, precluding an assessment of their VT and AT. These users were excluded from this analysis in the quality assessment described in the data section.

## CONCLUSION

An acceptable level of error for the VT and AT was chosen to be 7.5 BPM. 85% of the testing data falls within this acceptance threshold. A manual investigation of the testing data comparing the manual and algorithmic VT and AT values revealed that the remaining differences are likely due to the inherent uncertainty stemming from the use of free-form data. It is worth noting that two human annotators may differ by more than 7.5 BPM a small percentage of the time. Given the high correspondence between the gold standard VT and AT values and the algorithmically assessed values, this is an acceptable level of error for the average user.

Incremental exercise protocols with OMsignal apparel typically result in far less ambiguous ventilation-HR curves. Use cases that require a greater accuracy or precision than presented here should consider using such protocols in order to minimize the uncertainty introduced by free-form activity.

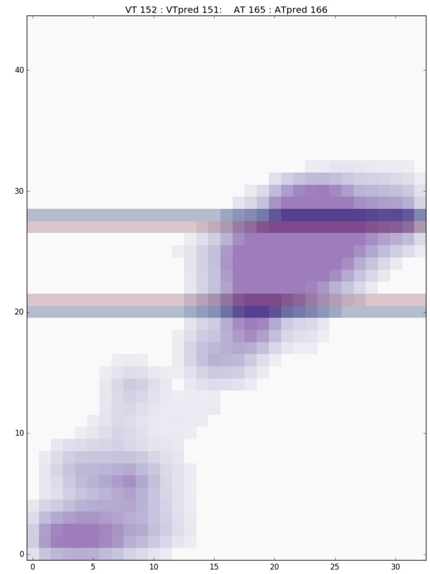


Figure 2: Example heat map. Ventilation along x-axis, heart rate along y-axis. Demonstrates good inflection points obtained from free-form run data. Manual (red bars) and algorithmic (blue bars) VT and AT values are very close.

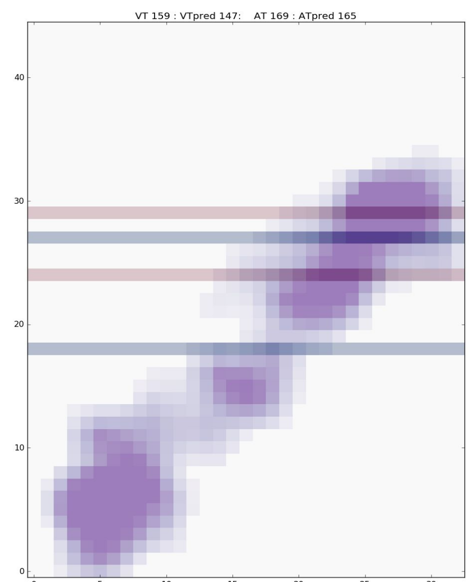


Figure 3: Example heat map. Ventilation along x-axis, heart rate along y-axis (in bin number). Demonstrates one form of ambiguity that can occur: there is no clear VT inflection point. As a result the manual VT (lower red bar) and algorithmic VT (lower blue bar) are far apart.

## REFERENCES

[1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.

[2] AR. Hoogeveen, G. Scheep and J. Hoogsteen, "The ventilatory threshold, heart rate and endurance performance: relationships in elite cyclists." *International Journal of Sports Medicine,* vol. 20 (02), pp 114-117, 1999.

[3] WD. McArdle, FI. Katch and VL. Katch, "Essentials of exercise physiology", *Lippincott & Wilkins*, pp. 291-294, 2006.

[4] K. Svedahl and BR. MacIntosh, "Anaerobic threshold: the concept and methods of measurement." *Canadian Journal of Applied Physiology,* vol. 28 (2), pp. 299-323, 2003.