Biomedical Signal Processing and Control 35 (2017) 19–29

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Predicting physiological parameters in fatiguing bicycling exercises using muscle activation timing

Petras Ražanskas[a], Antanas Verikas[a,b,*], Per-Arne Viberg[d], M. Charlotte Olsson[c]

[a] Department of Electric Power Systems, Kaunas University of Technology, Studentų g. 50, LT-51368 Kaunas, Lithuania
[b] Intelligent Systems Laboratory, Halmstad University, Box 823, SE-30118 Halmstad, Sweden
[c] School of Business, Engineering and Science, Halmstad University, Box 823, SE-30118 Halmstad, Sweden
[d] Swedish Adrenaline, Pilefeltsgatan 73, SE-30250 Halmstad, Sweden

ARTICLE INFO

Article history:
Received 29 October 2016
Received in revised form 15 January 2017
Accepted 22 February 2017

Keywords:
Random forests
Surface electromyography
Muscle activation patterns
Fatigue detection
Bicycling

ABSTRACT

This article is concerned with a novel technique for prediction of blood lactate concentration level and oxygen uptake rate from multi-channel surface electromyography (sEMG) signals. The approach is built on predictive models exploiting a set of novel time-domain variables computed from sEMG signals. Signals from three muscles of each leg, namely, vastus lateralis, rectus femoris, and semitendinosus were used in this study. The feature set includes parameters reflecting asymmetry between legs, phase shifts between activation of different muscles, active time percentages, and sEMG amplitude. Prediction ability of both linear and non-linear (random forests-based) models was explored.

The random forests models showed very good prediction accuracy and attained the coefficient of determination R^2 = 0.962 for lactate concentration level and R^2 = 0.980 for oxygen uptake rate. The linear models showed lower prediction accuracy. Comparable results were obtained also when sEMG amplitude data were removed from the training sets. A feature elimination algorithm allowed to build accurate random forests (R^2 > 0.9) using just six (lactate concentration level) or four (oxygen uptake rate) time-domain variables. Models created to predict blood lactate concentration rate relied on variables reflecting interaction between front and back leg muscles, while parameters computed from front muscles and interactions between two legs were the most important variables for models created to predict oxygen uptake rate.

1. Introduction

Over the years, cycling has become an important part of many people's lives. It provides environmentally friendly and cheap means of traveling short distances, and it can also serve as a beneficial exercise for cardiovascular health and general fitness [1,2]. Cycling has also become a popular professional sport, with multiple competitions held year-round across the globe.

During training, fatiguing cycling sessions are important, as they increase muscle power, raise lactate threshold, and improve oxygen consumption at high loads [3]. On the other hand, if the athletes are subjected to excessive fatigue and are not given enough time for proper recovery, this can lead to overtraining and loss of performance, and in the long term increases susceptibility to injuries [4,5]. For these reasons, it is highly desirable to establish methods for non-invasive fatigue estimation, which could help making appropriate adjustments of the training regimen. However, this task is complicated, because fatigue is also affected by psychological factors, which make subjective methods of fatigue assessment (e.g. Borg's Rating of Perceived Exertion [6]) unreliable. Indeed, highly motivated athletes are willing to continue the exercise much longer, and report lower levels of perceived fatigue than unmotivated athletes in the same circumstances. Therefore, establishing an objective frame of reference is important.

During an exercise, muscles tend to produce most of their power through the means of aerobic respiration. However, as the exercise intensity rises, the respiratory and cardiovascular systems cannot deliver sufficient oxygen to the muscles, and much less effective anaerobic glycolysis becomes the primary means of power production. One of the byproducts of this process is lactate, which is released into blood stream by the muscle cells to be carried to liver and metabolized back to glucose through a process called Cori cycle [7]. The hydrogen ions lower the pH of the interstitial fluid,

* Corresponding author at: Intelligent Systems Laboratory, Halmstad University, Box 823, SE-30118 Halmstad, Sweden.
E-mail addresses: petras.razanskas@hh.se (P. Ražanskas), antanas.verikas@hh.se (A. Verikas), pelle@swedishadrenaline.se (P.-A. Viberg), charlotte.olsson@hh.se (M.C. Olsson).

http://dx.doi.org/10.1016/j.bspc.2017.02.011
1746-8094/

which impedes the propagation of action potential along muscle fibers, leading to the loss of muscle power and consequently fatigue [8]. For these reasons, monitoring blood lactate concentration and oxygen uptake are popular ways to estimate the level of exercise intensity and general fatigue. It must be mentioned, however, that the majority of the theory linked to the cell–cell and intracellular lactate shuttles hypothesis, confirmed over the years, helped to understand that working muscles both produce lactate and use it as a fuel, see [9,10].

Several techniques can be used for non-invasive monitoring of muscle activity [11]. Of these, surface electromyography (sEMG) is the most popular, as the obtained readings are relatively precise, and the sensors are lightweight and cheap. Indeed, skin contact electrodes can be easily embedded into textile, such as cycling shorts, making data acquisition very easy [12,13]. Combined with a portable computational device, such as a smart-phone, this allows for real-time data analysis and feedback to the user with minimal intrusiveness as the exercise progresses.

The most common method of analyzing sEMG signals is spectral analysis. In the course of sustained muscle contraction at static medium or high load, the frequency components of the sEMG signal get compressed and shifted towards the lower frequencies [14]. Usually this effect can be tracked using a measure called median power frequency (MPF), which is defined as the frequency that divides the power spectrum of a signal into two parts of equal energy. However, this approach is generally not feasible when a dynamic exercise is involved, as the frequency content can be affected by changes in the muscle force and muscle length, or by muscle movement under the skin where electrodes are applied. As such, more complex time-frequency domain methods may be required [15–18]. However, Bonato et al. showed [19] that a lot of complications can be avoided in case of cyclic exercises by assuming that the sEMG signals resulting from an exercise are cyclostationary.

There have been a number of studies relating sEMG signals to fatigue, however, the results tend to be inconclusive. Jammes et al. showed that the root-mean-square amplitude (RMS) of the sEMG signal is strongly correlated to the measured load of the muscle [20]. On the other hand, the study by Pringle and Jones, among other findings, reported that there were no significant changes to the integrated sEMG signal as the subjects fatigued [21]. A study by Jensen et al. attempted to relate MPF of an sEMG signal to blood lactate concentration, and found that, contrary to the expected outcome, MPF increased as the majority of subjects fatigued [22]. Tenan et al. also investigated the relationship between MPF and blood lactate concentration and came to a conclusion that these two values were not strongly correlated [23]. By contrast, the study by Gerdle et al. used iso-kinetic knee extension exercise to validate the relevance of MPF and RMS as fatigue indicators, and reported very strong correlation between MPF and fatigue, while no such correlation was found between RMS and fatigue [24]. Arjunan et al. used six features, including MPF, to study the relationship between force of muscle contraction and muscle fatigue in isometric contraction experiments and come to conclusion that "*increase in synchronization index*" had the highest correlation with muscle fatigue [25].

Predictive models built using sEMG data have been used in many other fields. Prostheses control through classification of sEMG data into several classes of forearm motions [26,27], hand gesture recognition [28–30] and analysis of gait [31,17] are popular application fields. Anticipation of head movement for virtual-environment applications [32,33], a computer interface based on sEMG signals activated by wrist movements [34], prediction of forces affecting the lumbar spine [35], diagnostics of neuromuscular disorders [36], classification of different tremor types [37], motion pattern recognition [38], and prediction of knee abnormalities [39] are several other recent examples involving sEMG data-based predictive modelling. Apart from frequency and time-frequency domains-based analysis, nonnegative matrix factorization-based features [29,40], probabilistic features obtained from the generative models [39], decomposition of sEMG signals into a set of noise-free intrinsic mode functions followed by independent component analysis (ICA) [36], and the nonlinear Maximal Lyapunov Exponent features [30] have been utilized to build predictive models.

The study presented in this article extends our previous work [41], where extended sets of frequency-domain variables were used to build muscle-specific linear and non-linear, random forest-based models to predict blood lactate concentration level and oxygen uptake rate. While the results of the study were very positive – random forest-based models designed for prediction of blood lactate concentration level reached a coefficient of determination around $R^2 = 0.8$, and similar models designed to predict oxygen uptake reached around $R^2 = 0.9$ – it brought up a question, whether similar results could be achieved without relying on frequency domain features, but instead using timing data from several muscles in parallel. Therefore, in this paper we propose and investigate a set of features based on relationships between timing of activation and deactivation of selected leg muscles, and build linear and random forest-based models for prediction of blood lactate concentration level and oxygen uptake rate. Features of such type would significantly simplify the feature extraction process and obviate the need of sEMG normalization.

Our study was, to some extent, inspired Blake and Wakeling, who studied muscle coordination patterns during outdoor cycling and observed that muscle coordination fluctuated during the course of outdoor trial and coordination patterns significantly covaried with power output [42]. Our hypothesis is that muscle fatigue should reveal itself by varying parameters characterizing engagement of the muscles.

It must be noted that using muscle activation patterns as data for fatigue identification is not a completely new idea. Knaflitz and Molinari [43], among other factors, also investigated changes in muscle employment. However, the authors do not report any significant consistent changes in activation patterns of biceps femoris, rectus femoris, or gastrocnemius, suggesting that the subjects of their study, even though they subjectively reported muscle fatigue, did not adapt their muscle employment strategies to counteract effects of fatigue. However, it must be kept in mind that the study was carried out at relatively low loads.

A very important concern when working with timing events of an sEMG signal is a robust method of detecting muscle activation and deactivation. A number of procedures for detecting such change points in time series have been suggested over the years, both generally and applied specifically to sEMG signals. Single threshold methods [44] trigger when the signal exceeds a certain amplitude, and thus offer the simplest solution. However, such techniques are not robust and produce a lot of false positives. These techniques have been further improved upon by employing double threshold methods [45], which enforce a secondary threshold on the duration of the detected event. More complex, statistics based approaches have also been considered, such as maximum likelihood methods [46] and Bayesian analysis [47]. Finally, a cumulative sum of squares method [48] has been proposed as a viable alternative to the latter two procedures; it produces comparable results but requires significantly less computational effort.

2. Methodology

2.1. The testing protocol and data collection

Data from 9 test subjects – 5 men and 4 women, with the mean age of 35 ± 12 years, mean weight of 71 ± 11 kg, and mean height