Classification of Perceived Running Fatigue in Digital Sports

Bjoern Eskofier and Florian Hoenig
University of Erlangen-Nuremberg
Chair of Pattern Recognition
Martensstrasse 3, Erlangen, Germany
eskofier, hoenig@cs.fau.de

Pascal Kuehner
University of Koblenz-Landau
Department of Psychology
Fortstrasse 7, Landau, Germany
kuehner@uni-landau.de

Abstract

This paper presents methods for collecting and analyzing physiological and biomechanical data during recreational runs in order to classify an athlete's perceived fatigue state. Heart rate and its variability, running speed and stride frequency, GPS position and shoe heel compression were recorded continuously while runners moved freely outdoors. During their activity the sportsmen answered questions about their fatigue state in five-minute-intervals. Data from 84 one-hour-runs was collected for analysis. The data was analyzed using features computed for each step of the athlete to distinguish three levels of the runner's fatigue state with an accuracy of 75.3% across multiple study participants and 91.8% in the intraindividual case. The results show that for most participating runners, a heart rate variability periodogram feature and a step duration feature are best suited for classification of the perceived fatigue level. This information can be used to support sportsmen, for example by adapting their equipment to the specific needs of a fatigued athlete.

1. Introduction

Smart sensors embedded in clothes and equipment for sports afford novel opportunities to support and guide athletes. A prominent example is the "adidas_1" running shoe, which is the first shoe that features an embedded system (see Figure 1). This shoe is built to adapt to various running conditions like the prevailing surface situation, the runner's speed and fatigue state by changing the cushioning of the sole. A precise classification of the mentioned conditions is mandatory for the adaptation, which is conducted by a motor driven cable system inside the shoe. To facilitate the classification, the heel compression signal of the runner is continually measured and processed by the embedded microcontroller. A description of the "adidas_1", its functionality and embedded system hardware can be found below and in more detail in [2].

In this paper, we consider the very important task of classifying the fatigue state of a runner. Athletes can benefit a lot from that information. In the particular case of running with the "adidas_1", the shoe can be adapted accordingly, stabilizing exhausted muscles by providing more stiffness. However, the "adidas_1" shoe is just one example of smart sensors embedded in clothes. Similar actions could be taken in other endurance sports, where it is equally important to actively support an athlete by adapting the equipment to the amount of fatigue.

Many features of physiological and other signals have been proposed as indicators of fatigue by sports researchers and psychophysiologists, but these are normally individually evaluated and their use for classification purposes is rarely tested [4]. In this paper, we show how we apply pattern recognition techniques to identify a feature set to appraise the perceived fatigue level of sportsmen during a free one-hour-outdoor run, which to our knowledge has never been done before. Our features are derived from the physiological heart rate (HR) signal and the step signal as measured by the "adidas_1". The HR signal was chosen because previous studies have shown that especially heart rate variability (HRV) is very important for assessing psychological stress [4] and physiological fatigue [8].

We also obtained data about the perceived fatigue state of the runners by inquiring them periodically during their physical activity via a specially programmed mobile phone. The mobile phone automatically asked the athletes about their level of fatigue at predefined time points and recorded the answers that were on a scale from 0 to 6. Self-rating the fatigue state is a widely used method to get information about fatigue [1]. The information in the heel compression signal of the runners was represented by 19 features computed at each step of the athletes. Additionally, 10 physiological features were used to characterize HR and HRV in windows of 3 different lengths. These windows were placed in a causal context relating to the steps of the runners, i.e. derived from a section of the respective signals directly preceding each step. The steps in 5 phases of the run were labeled as belonging to one of 3 fatigue classes derived from the self-ratings of the runners: no or very little, medium and high perceived fatigue.

2   Data Collection

A total of 84 runners (30 female, 54 male) participated in the one-hour outdoor running study. The age of the subjects was 32.9 ± 7.9 years (average, standard deviation). The subjects were not specifically chosen according to running experience; instead, the group contained both low and high activity runners. The measurement system consisted of 3 separate devices. Firstly, we used a "Polar RS800 Running Computer" [10], which included an "S3 stride sensor" and chest strap. This system is capable of measuring running speed, stride frequency, barometric height, HR and the time between
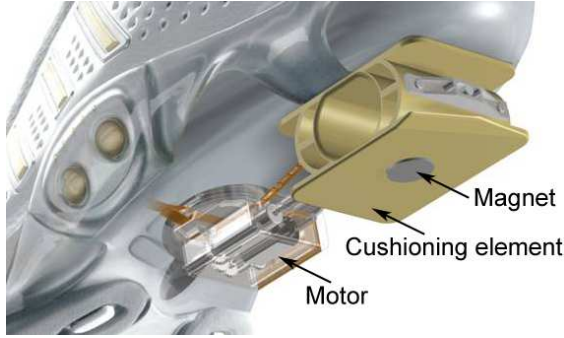
Figure 1. The "adidas_1" shoe, its cushioning element, magnet and motor unit.

Table 1. Grades for the athlete self-rating.

Spoken answer	Meaning
0	not at all
1	very little
2	little
3	somewhat
4	rather
5	very
6	extremely

Figure 2. Visualization of an example run.

two consecutive heart beats (RR-interval). We set the sampling interval for the first four signals to $5\,\mathrm{s}$, RR-intervals were measured with a resolution of $1\,\mathrm{ms}$.

We also continuously measured the heel compression signal of the runners using the "adidas_1" shoe. Figure 1 shows the measurement principle. A hall sensor mounted at the top of the cushioning element detects the magnetic field strength induced by a small magnet. The sensor was sampled with a rate f_s of $342\,\mathrm{Hz}$. The sensor-magnet distance d_m can then be computed from the field strength with an accuracy of $\pm 0.1\,\mathrm{mm}$.

Lastly, we used a Nokia 6110 Navigator cell phone with a custom-built Java software. It played sound files, recorded answers related to perceived fatigue questions and gathered GPS information. To facilitate the latter, we stored the data from an inbuilt GPS receiver. The phone was placed in a belt that was attached to the upper arm of the participants. The runners also wore a Bluetooth headset that was connected to the phone to capture their answers.

At the beginning of the experiments, the participants were standing and a short recorded instruction message was presented to them. Once the presentation was completed, they were asked about their fatigue state for the first time and then they were directed to start running. After that, we questioned them about their perceived fatigue state every 5 minutes. The athletes were instructed to answer each question about their subjective fatigue level with a self-rated grade as given in Table 1. From these ratings we derived 3 fatigue classes ω_k, where $k = 1, 2, 3$, with which we labeled each single step of the runners. Class ω_1 corresponded to no or very little perceived fatigue (self-ratings 0, 1, 2 - 34.6% of the labels), ω_2 denoted a medium level of fatigue (self-ratings 3, 4 - 41.1% of the labels) and ω_3 was used for considerable perceived fatigue (self-ratings 5, 6 - 24.3% of the labels). The steps were labeled according to the nearest recorded fatigue state time point.

After completion of the run, each participant was asked to fill in a questionnaire. Only 2 out of the 84 runners perceived a notable impediment by the equipment while running. An example run is visualized in Figure 2 using the Google Earth software. In this illustration, running speed is displayed as the height of the band along the running track. The fatigue classes are color coded: Green (light gray in b/w) means little or no perceived fatigue; orange (medium gray) is the medium; red (dark gray) represents the considerable perceived fatigue class.
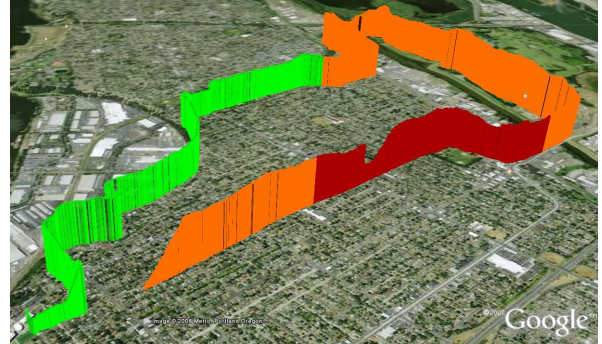
3 Feature Extraction

Out of the 84 study participants, 30 had to be excluded from further processing for various reasons. More specifically, 2 runners had incomplete audio data due to malfunctioning of the Bluetooth headset, and 5 other participants had incomplete data from the Polar RS800 system. The remaining 23 runners had to be excluded because of unusable data from the "adidas_1" shoe. In 8 of these cases, data collection was not possible because the "adidas_1" was not present in all shoe sizes at the beginning of the study, and therefore the runners had to use other shoe models. In the remaining 15 cases, the runners were mid- or forefoot strikers. The measurement system of the "adidas_1" is located at the heel of the shoe and can therefore only capture significant data for rearfoot strikers which account for more than 80% of the population [5]. To circumvent further problems with data collection, we made sure that the Bluetooth devices were working properly before the run, and applied a salt water solution to the chest straps of the RS800 systems to improve connectivity. We did not want to ask runners specifically whether they are fore- or midfoot strikers before the run to prevent a change in running style. Following this procedure, we had to cope with data loss for these runners. However, no additional bias was introduced thereby.

In total, 126,880 single step events were automatically detected in the 5 running phases that are defined in Table 2. We selected this number of phases because it gave us the maximum number of steps while the phases still did not overlap. The very beginning of the runs was not considered to ensure that the runners were warmed up and accustomed to data collection. The gaps between the phases were used to guarantee that the respective analysis windows for the intraindividual analysis (see section 4) did not overlap. Each step in the 5 running phases was labeled according to one of the 3 fa-
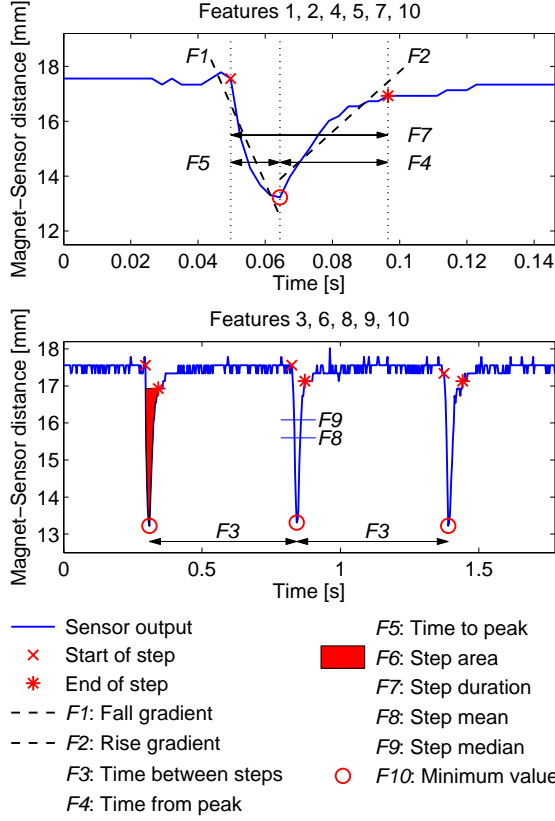
Figure 3. Depiction of the step signal features 1 to 10.

Table 2. The five run phases for analysis.

Running Phase	Interval [min]	# of steps
1	8.0 - 13.0	25513
2	18.5 - 23.5	25708
3	29.0 - 34.0	25208
4	39.5 - 44.5	25149
5	50.0 - 55.0	25302

Table 3. Overview of the step signal features 11 to 19.

Feature	Feature Description
$F11$	SD - values contained in one step
$F12$	SD - step minima ($F10$)
$F13$	SD - step means ($F8$)
$F14$	SD - step standard deviation ($F11$)
$F15$	SD - step duration ($F7$)
$F16$	SD - step area ($F7$)
$F17$	SD - time between steps ($F3$)
$F18$	SD - time to peak ($F5$)
$F19$	SD - time from peak ($F4$)

tigue categories defined in section 2. For each step, we extracted 19 features from the heel compression signal denoted by $F1 \ldots F19$. Figure 3 shows $F1 \ldots F10$. Features $F11 \ldots F19$ are all measuring the standard deviation (SD) of different attributes (see Table 3), SD is computed as follows:

$$\mathrm{SD}(\boldsymbol{c}) = \sigma_N = \left(\frac{1}{N-1} \sum_{k=1}^{N} (c_k - \bar{c})^2 \right)^{\frac{1}{2}} \quad (1)$$

where $\boldsymbol{c}$ is a vector containing the attribute data and varies depending on the attribute under investigation and N is the number of samples, i.e. the number of elements in $\boldsymbol{c}$. $\bar{c}$ is the average value of all the elements of $\boldsymbol{c}$. More specifically, for $F11$, $\boldsymbol{c}$ contains all the heel compression distance measurements d_m collected during a single step event (Fig. 3 top). For $F12 \ldots F19$, $\boldsymbol{c}$ contains the most recent 16 values of the corresponding feature. For example, for $F12$, each element of $\boldsymbol{c}$ is the minimum d_m that occurred within a single step, i.e. $F10$ and $\boldsymbol{c}$ has a total of 16 elements, one min. d_m value for each of the 16 most recent step events (see Fig. 3 bottom). $F13 \ldots F19$ have similar $\boldsymbol{c}$ values. For more details see Table 3. This set of step signal features has proven to be useful in our earlier experiments that aimed at surface or inclination classification with shoe data alone and we expected them to significantly contribute to perceived fatigue level classification as well.

For the physiological data we computed features in causal windows located before each step. The chosen interval times for the windows were $1\,\mathrm{min}$, $3\,\mathrm{min}$ and $5\,\mathrm{min}$, respectively. In theses windows, we calculated 3 features derived from the HR signal and 7 features derived from the more important RR interval data that gives an indication about HRV. For the HR data we computed mean, variance and offset HR value. The latter was calculated as the difference from the mean HR that the subjects had while standing and listening to the initial set of instructions and the mean HR in the analysis window. These 3 features are represented by μ_{HR}, σ_{HR} and $\bar{d}_{HR}$. From the RR interval times we calculated the variance, denoted by σ_{RR}, in each of the described windows. Additionally, we computed features using two different methods, the Poincaré plot (PP) [9] and Lomb-Scargle periodogram (LSP) [6] analysis. Both approaches have been proven useful in HRV evaluation [7, 11]. In the PP method, a vector of RR-interval times $\boldsymbol{t}_{RR} = (t_1, t_2, \ldots, t_L)$ of length L is first rewritten in two auxiliary vectors

$$\begin{aligned} \boldsymbol{t}_{RR}^{+} &= (t_1, t_2, \ldots, t_{L-1}) \\ \boldsymbol{t}_{RR}^{-} &= (t_2, t_3, \ldots, t_L). \end{aligned} \quad (2)$$

The PP then consists of the ordered pairs

$$\left(\boldsymbol{t}_{RR,i}^{+}, \boldsymbol{t}_{RR,i}^{-} \right), \quad i = 1 \ldots L - 1. \quad (3)$$

These ordered pairs generate a comet-like shape in a scatter plot for a healthy person [7]. From this representation, 3 features are derived. $SD1_{RR}$ and $SD2_{RR}$ are two standard PP descriptors defined as

$$\begin{aligned} SD1_{RR} &= \mathrm{SD} \left(\frac{\boldsymbol{t}_{RR}^{+} - \boldsymbol{t}_{RR}^{-}}{\sqrt{2}} \right) \\ SD2_{RR} &= \mathrm{SD} \left(\frac{\boldsymbol{t}_{RR}^{+} + \boldsymbol{t}_{RR}^{-}}{\sqrt{2}} \right) \end{aligned} \quad (4)$$

and correspond to the SD of the respective projections of the PP on the line of identity $y = x$ and the line perpendicular to it ($y = -x$). It is widely accepted that SD1 reflects the short-term RR-interval variability, and SD2 reflects both short-term and long-term variability [9]. The third feature derived from the PP is

the total variability $S_{RR} = SD1_{RR} * SD2_{RR}$. The PP procedure is susceptible to outliers, thus we implemented an outlier removal filter designed for PP analysis as described in [9]. In the second RR analysis method, the LSP evaluation, we first calculated a spectral estimate of the activity in the $0\,\mathrm{Hz} \ldots 1\,\mathrm{Hz}$ band using Lomb's method [6]. The LSP is a means of obtaining power spectral density estimates directly from irregularly sampled time series, avoiding the requirements of resampling at uniform intervals and replacement of unusable samples. Healey and Picard [4] have already successfully used 3 features derived from the LSP: the total energy in the low frequency (LF_{RR}) band ($0\,\mathrm{Hz} \ldots 0.08\,\mathrm{Hz}$) and in the high frequency (HF_{RR}) band ($0.15\,\mathrm{Hz} \ldots 0.5\,\mathrm{Hz}$) were calculated as the sum of spectral powers. The ratio $R_{RR} = LF_{RR}/HF_{RR}$ was used as a third feature. These features have been used to represent sympathetic tone. The parasympathetic nervous system modulates HR at all frequencies between $0\,\mathrm{Hz}$ and $0.5\,\mathrm{Hz}$, whereas the sympathetic system modulates only below $0.1\,\mathrm{Hz}$. We thus derive features that represent sympathetic and parasympathetic influence on the heart. The hypothesis is that increased fatigue leads to increased sympathetic activity.

4 Analysis

For the classification experiments, we used a linear discriminant classifier [3]. With this classifier, each vector of observed features x is assigned to the class ω_k in a way that the discriminant function g_k

$$g_k(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \frac{1}{2}\boldsymbol{\mu}_k^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \ln P(\omega_k) \quad (5)$$

is maximized. In this equation, $\boldsymbol{\mu}_k$ denotes the class mean, $\boldsymbol{\Sigma}$ the pooled covariance matrix and $P(\omega_k)$ the a priori probability of each class. For the first experiment, leave-one-runner-out cross-validation was performed, where the classifier was trained using all but the feature vectors from one specific runner, then classifying that runner's feature vectors according to maximum g_k and calculating the mean classification accuracy. We first tested each of the individual features in order to evaluate its performance for this interindividual classification. The 6 best results are given in Table 4. It can be seen that the LF_{RR} and R_{RR} features derived from the LSP perform best when analyzing a 5 min window. The $F7$ feature derived from the heel compression signal can also contribute to accurate classification. But, as stated in the introduction, single features were not expected to perform very well when analyzed individually. When using our complete feature set, we achieve a result of 75.3% correct classification. This result shows that recognition rates can significantly

be improved when using multiple features derived from different sensors.

In a second analysis, we evaluated the performance of our feature set in the case of intraindividual classification. We split each runner's feature dataset in 5 subsets corresponding to the 5 running phases and crossvalidated by training the classifier using all but the features from one phase, then classifying the steps in the remaining subset according to maximum g_k. In average, the recognition rate for this experiment was 91.8%, indicating that the feature set is very selective in the intraindividual case.

5 Conclusions

This research demonstrates the application of pattern recognition methods to detecting perceived running fatigue using features from heel compression and heart rate signal. We showed that in the interindividual case, we can significantly improve the recognition rate by using multiple features from 61.2% in the best one-feature case to 75.3%. Perfect performance was not expected due to inconsistencies in the athlete fatigue self-rating. However, when circumventing this problem by intraindividual analysis, we achieve very high recognition rates of 91.8% for 3 perceived fatigue levels. This suggests that an individually trained automatic system can very precisely support the athlete, for example by providing more shoe stiffness by the "adidas_1" running shoe when the sportsman gets fatigued.

Further work will be done to incorporate the additionally collected speed, stride frequency and altitude information. Moreover, an automatic selection process will be implemented to determine the ideal feature combination out of our partly overdetermined feature set to further improve the interindividual classification rate.

References

[1] G. Borg. *Borg's Perceived Exertion and Pain Scales.* Human Kinetics, Champaign, IL, USA, 1998.

[2] C. DiBenedetto, M. A. Oleson, and C. Roth et al. Intelligent footwear systems, 2004. United States Patent US 20040177531, European Patent EP 1582108.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Ed.).* Wiley-Interscience, November 2000.

[4] J. Healey and R. Picard. Smartcar: Detecting driver stress. *15th International Conference on Pattern Recognition, Proceedings*, 4:218–221, September 2000.

[5] B. A. Kerr, L. Beauchamp, and V. Fisher et al. Footstrike patterns in distance running. In B. M. Nigg and B. A. Kerr, editors, *Biomechanical aspects of sports shoes and playing surfaces*, pages 135–142. University Printing Calgary, 1983.

[6] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, February 1976.

[7] F. Marciano, M. Migaux, and D. Acanfora et al. Quantification of poincaré maps for the evaluation of heart rate variability. *Computers in Cardiology*, pages 577–580, 25-28 September 1994.

[8] V. Pichot, F. Roche, and J.-M. Gaspoz et al. Relation between heart rate variability and training load in middle-distance runners. *Medicine & Science in Sports & Exercise*, 32(10):1729–1736, October 2000.

[9] J. Piskoski and P. Guzik. Filtering poincaré plots. *Computational Methods in Science and Technology*, 11(1):39–48, June 2005.

[10] Polar Electro Oy, Finland. RS800 Running Computer. http://www.polar.fi/en/products/running/RS800, 2008.

[11] J. P. Spiers, B. Silke, and U. McDermott et al. Time and frequency domain assessment of heart rate variability: A theoretical and clinical appreciation. *Clinical Autonomic Research*, 3(2):145–158, April 1993.

Table 4. Ranking of the 6 best performing individual features

Feature	Window	Rank	Accuracy
LF_{RR}	5 min	1	61.2%
R_{RR}	5 min	2	60.8%
LF_{RR}	3 min	3	58.2%
R_{RR}	3 min	4	57.4%
$F7$	N/A	5	48.9%
$\bar{d}_{HR}$	5 min	6	44.1%