Identification of Athletes During Walking and Jogging Based on Gait and Electrocardiographic Patterns

Peter Christ(✉) and Ulrich Rückert

Cognitronics and Sensor Systems Group, CITEC, Bielefeld University,
Universitätsstr. 21-23, 33615 Bielefeld, Germany
{pchrist,rueckert}@cit-ec.uni-bielefeld.de
http://www.ks.cit-ec.uni-bielefeld.de

Abstract. We propose a biometric method for identifying athletes based on information extracted from the gait style and the electrocardiographic (ECG) waveform. The required signals are recorded within a non-clinical acquisition setup using a wireless body sensor attached to a chest strap with integrated textile electrodes. Our method combines both sources of information to allow identification despite severe intra-subjects variations in the gait patterns (walking and jogging) and motion related artefacts in the ECG patterns. For identification we use features extracted in time and frequency domain and a standard classifier. Within a treadmill experiment with 22 subjects we obtained an accuracy of 98.1 % for velocities from 3 to 9 km/h. On a second data set consisting of 9 subjects and two sessions of recording, our method achieved 93.8 % despite variations in the patterns due to reapplying the body sensor and an increased velocity (up to 11 km/h).

Keywords: Human identification · Accelerometer · Electrocardiograph (ECG) · Wireless body sensor (WBS) · Pattern recognition

1 Introduction

The identification of humans is important for various applications such as surveillance systems, authorization checks at doors or electronic devices (e.g. computer, smartphone). A variety of biometric characteristics have been investigated such as information from fingerprint, iris and retina, human face, voice, gait or electrocardiograph.

Previous work has shown that discerning, reproducible information on the human is found in the ECG waveform, especially around the QRS complex [7,11]. Moreover, biomechanical differences between the gait style of humans have been investigated and used for identification within video and acceleration sensor based applications [12,18].

We propose a biometric measure combining both sources of information: characteristics in the electrocardiograph (ECG) waveform and the gait style.

M. Fernández-Chimeno et al. (Eds.): BIOSTEC 2013, CCIS 452, pp. 240–257, 2014.
DOI: 10.1007/978-3-662-44485-6_17

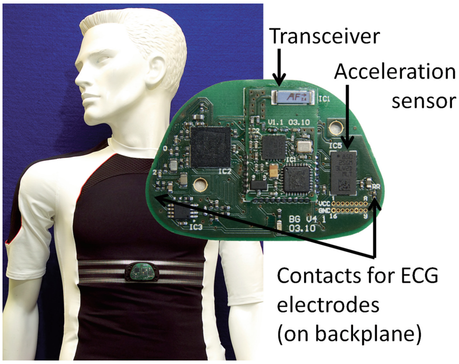


Fig. 1. Our self-made wireless body sensor (WBS) and its integration into a chest strap. The WBS can measure a person's electrocardiograph (ECG) and accelerations of the body along three orthogonal axes.

Unlike other applications, our approach focuses on the identification of athletes during physical exercise using a compact wireless body sensor (WBS) which is worn around the chest (see Fig. 1). The WBS is typically used to measure the heart-rate and the body accelerations of athletes. Our identification method additionally utilizes the sensor measurements to identify the athlete, enabling an automatic annotation of sensor data with the subject's identity. Our goal is to overcome the drawbacks of a manual annotation of measurements for applications in sports medicine and athlete training research. Furthermore, recognizing the subject allows to automatically load personal settings on the WBS or the sport equipment for a customized training. Our identification method is in particular interesting for a WBS which is used with several athletes of a mid-sized group.

Our identification method uses features in time and frequency domain to extract characteristics on the subject which are used as input to a classifier for identification. By combining information from gait and ECG we can successfully identify subjects despite of artefacts in the ECG caused by a slipping of the ECG electrodes and severe variations in the gait patterns between walking and jogging.

Previous work in this field focused on the identification of humans from either gait or ECG waveform characteristics. Mainly ECGs were used which were recorded at rest or with a clinical acquisition setup. The gait based identification was carried out for walking velocities.

Rong et al. proposed a method which uses measurements recorded during walking with an accelerometer located at the subject's waist [19]. The method utilises a segmentation into gait cycles to extract gait patterns. Dynamic time warping is applied to compensate natural changes in walking speed. The actual gait segment is then compared with a reference pattern of the subject and a 1-nearest neighbour classifier is used to recognize the subject. Ailisto et al. evaluated an accelerometer based identification based on similarities between gait segments to protect portable devices [2]. Mäntyjärvi et al. evaluated a gait based

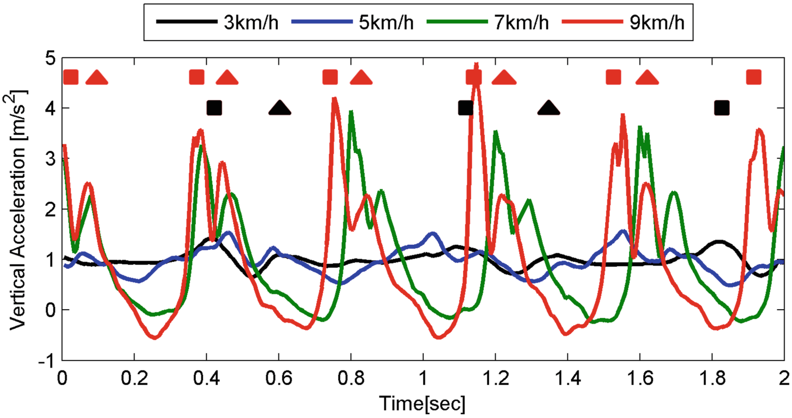


Fig. 2. Vertical acceleration data of a subject walking and jogging at velocities from 3 to 9 km/h. Each stride is represented by two consecutive peaks which correspond to the heel strike (square) and the toe strike (triangle). These peaks are marked for 9 km/h (red) and for 3 km/h (black). Velocity can be increased with either longer strides (increase in signal amplitude) or a higher step frequency (color figure online).

identification for different walking velocities using correlation coefficients derived from a template comparison, frequency coefficients and a histogram based comparison [16]. Gafurov et al. proposed two methods based on histogram similarity and gait cycle length to distinguish acceleration measurements recorded at the lower leg [12].

Several methods have been proposed to identify a human based on ECG measurements. Biel et al. used data from a standard 12-lead ECG recorded during rest to identify subjects using multivariate analysis [4]. Furthermore, the study showed that identification is possible with even one-lead ECGs. Shen et al. also utilises data from one-lead ECGs to distinguish subjects using a template matching and a decision-based neural network [21]. Chan et al. identifies subjects based on ECGs recorded within a non-clinical acquisition setup where the subjects were holding two electrodes on the pads of their thumbs [7]. For classification, three qualitative measures were used: percent residual difference, correlation coefficient, and a novel distance measure based on wavelet transform.

This paper is organized as follows: Sect. 2 describes the identification of a subject based on acceleration and ECG measurements. Information on preprocessing, feature extraction and used classifiers is given. Section 3 explains the conducted experiments for data collection. Section 4 presents the experimental results of our identification method. The results are summarised and discussed in Sect. 5 and a prospect on our future work is given.

2 Identification of a Subject

This section describes the identification of a subject based on gait style and ECG waveform characteristics. We describe the preprocessing of the signals, the feature extraction and the classifiers used for identification.

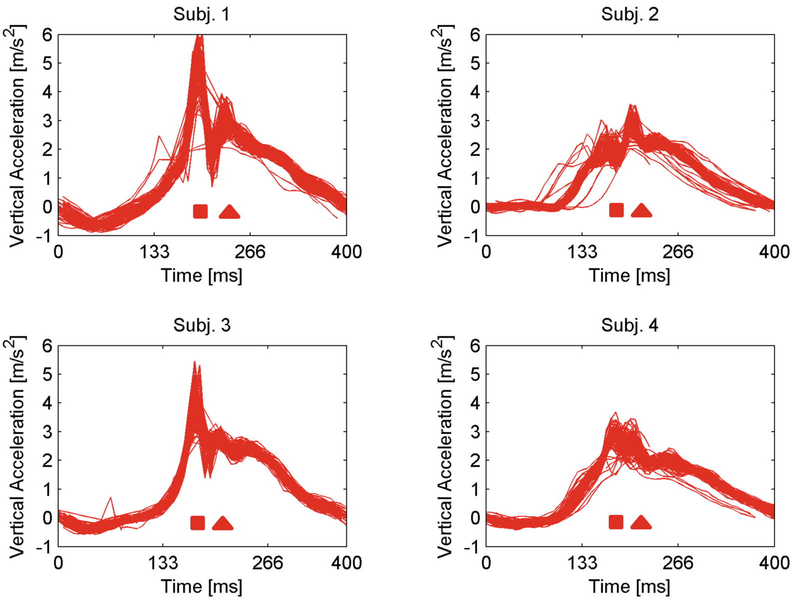


Fig. 3. Alignment of 100 consecutive strides of four subjects jogging at 9 km/h. The vertical acceleration signals were automatically segmented into strides and cross-correlation was used to align the strides. The peaks related to the heel strikes (square) and toe strikes (triangle) significantly differ in shape between the subjects.

2.1 Gait Analysis for Identification

Previous work has shown that gait differs between humans and that the gait style is fairly stable for a subject [3,18]. Bianchi et al. stated that the variability across humans depends on different kinematic strategies rather than on biomechanical characteristics [3]. Their study showed that subjects are different in the ability of minimising energy oscillations of their body segments for transferring mechanical energy.

In order to measure these inter-subject differences, severe intra-subject variations in the gait patterns between walking and jogging have to be taken into account. The intra-subject variations are a result of an adaptation of the gait to achieve different velocities. The velocity of a person is described by stride length and stride frequency. According to Weyand et al., longer strides are achieved by applying greater support forces to the ground which significantly increases the amplitude of the vertical acceleration signal, whereas the step frequency changes frequency components of the signal [24].

Samples of vertical acceleration data of one subject walking and jogging at different velocities on a treadmill are shown in Fig. 2. Strides are presented by two consecutive peaks corresponding to the heel and toe strikes. Significant changes in amplitude and an almost doubling of the step frequency can be observed between walking at 3 km/h and jogging at 9 km/h.

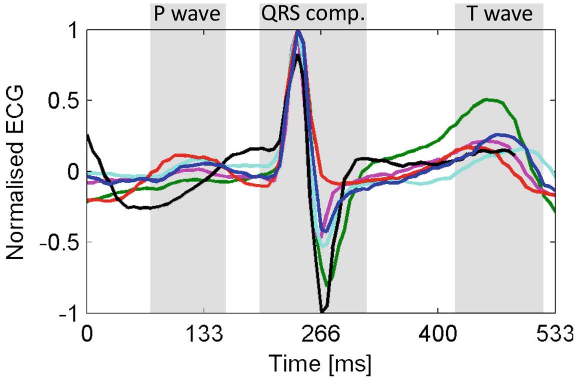


Fig. 4. Comparison of heartbeat segments of six subjects (different colours). The DC-offset was removed and the heartbeat segments were aligned using cross-correlation. We use inter-subject variations in the ECG waveform to identify subjects (color figure online).

Despite this intra-subject variability in the gait patterns, we observed inter-subject variations in acceleration signals recorded during walking and jogging [10]. In particular, heel and toe strikes differ in the vertical acceleration signal's shape between subjects (see Fig. 3). The peak acceleration of the heel strikes varies between the four subjects about 2 m/s^2.

2.2 ECG Analysis for Identification

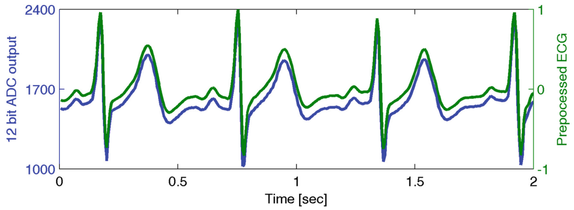
Inter-subject variability is also found in the ECG's waveform. The variations depend on position, size and anatomy of the heart, age, sex, relative body weight, chest configuration and various other factors [13,22]. Figure 4 shows sample heartbeat segments from six subjects recorded with our WBS. The ECG reflects the electrical activity of the heart and consists of the P wave followed by the QRS complex and the T wave [11, chap. 2]. Discerning information on the subjects is found in the QRS complex, the P and the T wave.

Chan et al. observed a high degree of reproducibility of information extracted from the QRS complex of a person through several sessions of recording [7]. Furthermore, a higher identification accuracy was determined for the P wave than the T wave.

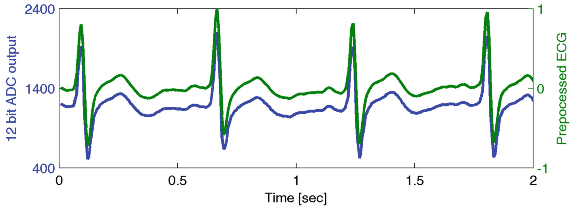
During physical exercise these characteristics can be superposed by motion related artefacts. These artefacts are caused by a slipping of the ECG electrodes and variations in the contact resistance during body movements [9]. Figure 7 shows disturbances in the ECGs of two subjects recorded during jogging on a treadmill.

2.3 Preprocessing of Acceleration and ECG Signals

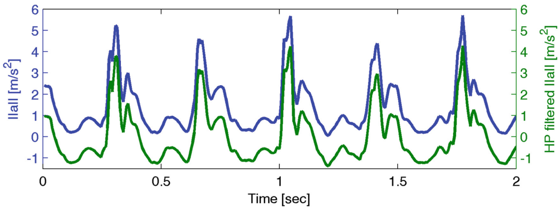
ECGs recorded with our WBSs showed hardware-related differences in the DC-offset making an ECG associable to a WBS. Furthermore, using textile ECG



(a) ADC output and preprocessed ECG of subject 1.



(b) ADC output and preprocessed ECG of subject 2.



(c) Magnitude of the acceleration vector $||\mathbf{a}||$ and the result of the offset reduction by the high-pass filter.

Fig. 5. The 12 bit analog-to-digital converter (ADC) output and the preprocessed ECG in comparison. An offset of 300 between the ADC output of the two different subjects was removed by the preprocessing. In Fig. 5c the offset due to the static acceleration of gravity and a sensor-related zero-g-level offset are reduced after preprocessing.

electrodes, the skin contact resistance decreases over time because of an increased transpiration which results in changes in the DC-offset. In order to avoid classification errors, we removed the DC-offset using a 4th-order high-pass butterworth filter with a cutoff frequency of $f_c = 0.67$ Hz. Additionally, we applied a low-pass filter with a cutoff frequency of $f_c = 40$ Hz to remove noise in the ECG signal.

With a decrease in skin contact resistance after a few minutes of exercise, we observed an increase in the ECG signal's amplitude which improved the signal-to-noise ratio. We normalised the signal's amplitude to assure that ECG segments are comparable. The results of the ECG preprocessing are shown in Figs. 5a and b.

For the frequency analysis of the acceleration measurements, we approximated the dynamic accelerations by applying a 4th-order butterworth high-pass filter with a cutoff frequency of $f_c = 0.1$ Hz to the magnitude of the acceleration vector $\mathbf{a} = (a_{AP}, a_{ML}, a_V)$; a_{AP} denotes anteroposterior accelerations, a_{ML}

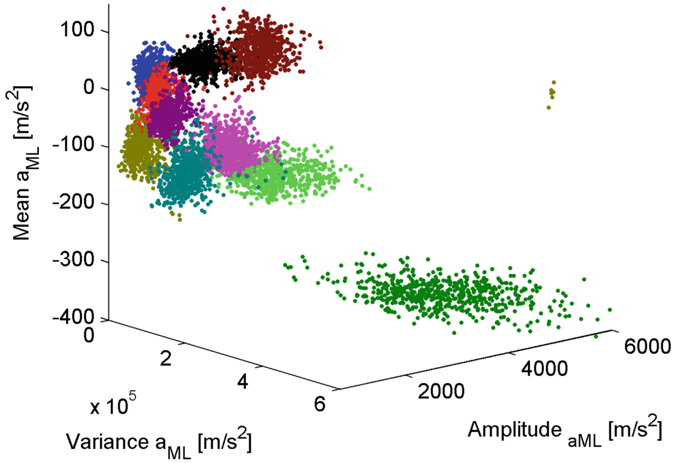


Fig. 6. Visualisation of time domain features extracted from mediolateral accelerations a_{ML} of ten subjects at 9 km/h. Clusters are observable for the different subjects. In our feature selection we obtained a good identification performance based on the mean, the variance, the amplitude and the root-mean-square (RMS) features (see Table 3).

mediolateral accelerations and a_V vertical (up-down) accelerations. The high-pass filter reduced the impact of the static acceleration due to gravity and a sensor-related offset (zero-g level offset). The results of this preprocessing step are shown in Fig. 5c.

2.4 Feature Extraction for Identification

In order to access characteristics of a subject in the acceleration and ECG measurements, we extracted features in the time and the frequency domain.

The features were calculated within a sliding window with no overlap and length N. Each window at time t consists of N measurements $x(t:t+N-1) = x(t), x(t+1), ..., x(t+N-1)$. We empirically determined an appropriate window length of two seconds ($N = 300$).

2.5 Time Domain Features

In the time-domain we calculated the variance, amplitude, mean and root mean square (RMS) along the three orthogonal axes a_{AP}, a_{ML} and a_V of the windowed acceleration signals. The variance, mean and amplitude of a_{ML} are visualised in Fig. 6. Discriminative clusters can be observed for the different subjects.

From the ECG signal we calculated a feature measuring the closeness of an unknown heartbeat segment to five reference patterns stored for each subject. This step requires a segmentation of the ECG signal into heartbeats. We used a QRS detection based on the algorithm of [1] in its implementation of Schloegl in the BioSig toolbox [23]. The five reference heartbeat segments were chosen

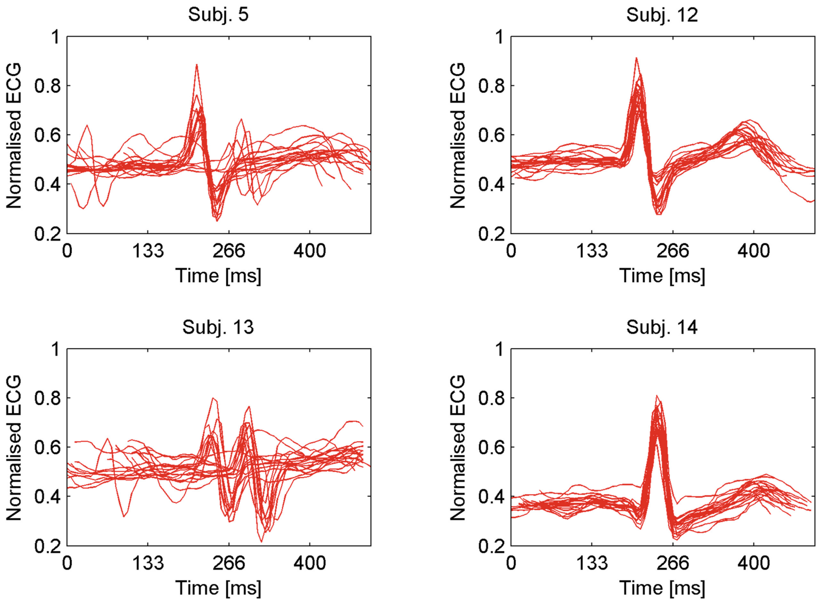


Fig. 7. Alignment of 20 heartbeat segments of four subjects recorded during jogging on a treadmill. A correct placement of the chest strap is important for an identification based on a similarity measure between heartbeat segments. Motion related artefacts and poor skin contact can disturb the ECG-signal (see subjects 5 and 13).

randomly from the ECG data of each subject. However, we assured that only heartbeat segments without severe disturbances were chosen. For identification, an unknown segment x was aligned to each reference segment y using cross-correlation:

$$R_{xy}(m) = \frac{1}{N} \sum_{j=0}^{N-m-1} y(j+m)x(j) \tag{1}$$

where N is the length of a segment and m the offset with $m = 0, 1, ..., 2N - 1$. We calculated the Pearson's correlation coefficient as a measure of similarity between the two segments. The Pearson's correlation coefficient is defined as the covariance (cov) of the two segments divided by the product of their standard deviation σ:

$$r(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y} \tag{2}$$

Figure 7 shows the alignment of 20 heartbeat segments of four subjects. The QRS-detection and the alignment are sensitive to motion-related artefacts (see subjects 5 and 13).

For heartbeat segments without major disturbances the alignment centred the segments around the QRS complex. The discerning information in this region of the ECG is fairly stable in relation to morphology changes in the ECG waveform during effort.

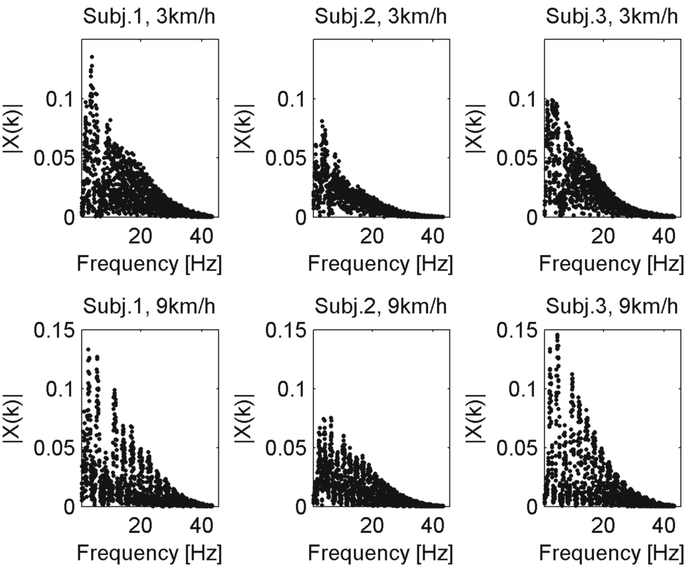


Fig. 8. The FFT amplitude spectra of the ECG signals of three subjects during walking (3 km/h) and jogging (9 km/h). The amplitude spectra show differences between the subjects but also vary with the velocity.

2.6 Frequency Domain Features

In the frequency domain we use the discrete Fourier transform (DFT) to extract frequency components of each window. The DFT is defined as:

$$X(k) = \sum_{j=t}^{t+K-1} x(j) e^{-i2\pi k \frac{j}{K}}, \quad k = 0, ..., K-1 \tag{3}$$

where K is the number of outputs $X(k)$. We used a 512-point fast Fourier transform (FFT) algorithm to compute the DFT efficiently for our windows of the length $N = 300$. Therefore, each window $\mathbf{x}(t\!:\!t+N-1)$ was padded with trailing zeros to the length of $K = 512$. Before calculating the FFT, a Hamming window function was applied to each window to reduce the spectral leakage.

Figure 8 shows the FFT amplitude spectra of ECGs of three subjects recorded during walking (3 km/h) and jogging (9 km/h). Despite velocity related variations in the amplitude spectra, differences can be observed between the three subjects.

We calculated additional frequency domain features from the amplitude spectrum (FFT features): the variance, the mean, the Fourier coefficient with the highest amplitude and the Shannon entropy SE:

$$SE = -\sum_{k=0}^{K-1} |X(k)| \log_2(|X(k)|) \tag{4}$$

where $X(k)$ is the output of the DFT of length K.

2.7 Methods for Classification

We used a standard classifier to identify the subject based on the extracted features. The identification performance was determined by evaluating three different classifiers: artificial neural network (ANN), support vector machine (SVM), and random forest (RF).

2.8 Artificial Neural Network (ANN)

We used a feed-forward ANN with 25 neurons with tangent sigmoid activation functions in one hidden layer to associate the extracted features with the subjects' identities. The ANN was trained using back-propagation which is a supervised learning method [14]. During training the prediction of the network is compared to the known target value (subject's identity) and the weights are modified to minimize the mean square error. These errors propagate backwards from the output layer to the hidden layer [14]. The network was trained using the scaled conjugate gradient algorithm described in [17]. The weights and bias values of the neurons were updated using a gradient descent with momentum.

2.9 Support Vector Machine (SVM)

We used a ν-SVM [20] with a sigmoid kernel in its implementation in the LIBSVM[1] [8]. SVMs are fundamentally a two-class classifier. Various methods have been proposed how to use SVMs for multi-class problems [5, chap. 7]. We used a one-against-one method which constructs $n(n-1)/2$ classifiers where n is the number of classes to distinguish. Each classifier is trained on tuples from two classes. A voting strategy is then applied to determine the winning class [15].

2.10 Random Forest (RF)

A random forest is a classifier consisting of a combination of tree predictors. The growth of each tree is governed by independently and identically distributed random vectors [6]. Each tree votes for one class and the class which occurs most frequently is the output of the classifier. RF classifiers are fast in the training phase and the training time is linear to the number of trees used. The testing of an unknown tuple is performed on each tree independently and is therefore parallelisable. We used a RF consisting of 100 trees, with each tree being constructed of ten randomly chosen features.

3 Subjects and Data Collection

For the evaluation of our identification method, we recorded data within two experiments of subjects who volunteered to participate in the study. The subjects

[1] LIBSVM: library for support vector machines.

Table 1. Characteristics of the subjects who participated in the two experiments.

(a) First experiment with 22 subjects (15 men, 7 women).

Characteristic	Mean ± SD	Range
Age (yr)	26.6 ± 4.0	18-33
Height (cm)	179.8 ± 9.6	160-198
Weight (kg)	76.7 ± 11.1	58-108

(b) Second experiment with 9 male subjects.

Characteristic	Mean ± SD	Range
Age (yr)	27.0 ± 3.7	21-33
Height (cm)	180.2 ± 9.6	160-198
Weight (kg)	77.3 ± 10.8	65-108

were informed verbally and in writing in advance and signed an informed consent document.

In the first experiment 22 healthy subjects participated (see Table 1(a)). The data was collected using the treadmills in the gymnasium of our university. Velocities between 3 to 9 km/h were chosen to cover slow, normal, and fast walking as well as jogging. The treadmill was set to no incline and the velocity was manually increased by 2 km/h every two minutes. This procedure was repeated twice for each subject (total: 16 min per subject).

In order to estimate the impact of variations in the gait and ECG patterns resulting from reapplying the body sensor (electrode placement and conductance, acceleration sensor orientation), we repeated the experiment with a smaller group of nine male subjects from which data was collected within two independent sessions of recording (see Table 1(b)). Both sessions, which were one week apart, covered data at velocities from 3 to 11 km/h (total: 20 min per subject).

The accelerations of the upper body and the ECG were recorded with a self-made WBS (see Fig. 1). The WBS measures accelerations within a range of $\pm 6\text{m/s}^2$ along three orthogonally oriented axes using a commercial off-the-shelf accelerometer (ST LIS3LV02DL). The ECG is digitized using the analog-to-digital converter of a TI MSP430 microcontroller. Body accelerations and ECG were measured with a 150 Hz sampling rate and a 12 bit resolution (range 0 to 4095). The measurements were sent wirelessly to a nearby receiver for recording.

The subjects were given an explanation on how to place the chest strap with the WBS tightly around the chest. However, we didn't verify the correct placement of the WBS to assure real world conditions. Furthermore, no instructions were given on how to perform the exercise.

4 Results

This section describes the evaluation of the athlete identification on the data collected during walking and jogging on the treadmills.

4.1 Evaluation Methods

All features were calculated on windows of acceleration and ECG measurements of two seconds. No overlap of the windows was chosen to ensure fully discriminative training and testing data. We concatenated features of two consecutive

Table 2. Accuracy (ACC) and overall specificity ($\overline{S}$) of the identification of the 22 subjects. The results were determined with three different classifiers on a feature space combining acceleration and ECG features (combination C8, see Table 3).

Classifier	ACC	$\overline{S}$
ANN	94.2 %	99.8 %
SVM	90.4 %	99.5 %
RF	98.1 %	99.9 %

windows to have samples of four seconds of data to identify the subject. For the data of the first experiment with 22 subjects, we determined the identification performance using a ten-fold cross-validation. The two sessions of recording from the second experiment with 9 subjects were used for training the classifier (first session) and for evaluation of the resulting model (second session).

For the evaluation, we used three statistical measures: sensitivity, specificity and accuracy. In order to calculate the statistics we obtained the number of true positive samples TP_i, true negative samples TN_i, false positive samples FP_i, and false negative samples FN_i from the classifier's output. For a class i the sensitivity R_i is defined as:

$$R_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} * 100 \tag{5}$$

The sensitivity (also referred to as recall) measures the percentage of correctly classified positive samples in relation to all positive samples. For negative samples the specificity S_i is defined as:

$$S_i = \frac{\mathrm{TN}_i}{\mathrm{TN}_i + \mathrm{FP}_i} * 100 \tag{6}$$

We calculated the overall sensitivity $\overline{R}$ and the overall specificity $\overline{S}$ as a class-based weighted average. For our multi-class problem we refer to the overall sensitivity as the accuracy of the classifier:

$$ACC = \overline{R} = \sum_{i=1}^{n} p_i R_i \tag{7}$$

where n denotes the number of classes and p_i the probability of the occurrence of the class in the test data. In our data from the two experiments the samples are equally distributed for the $n = 22$ and $n = 9$ subjects ($p_i = 1/n, \forall i$). The overall specificity $\overline{S}$ is calculated accordingly. The optimum of the statistical measures is 100 %.

4.2 Results of the Athlete Identification

The following results were obtained for the data from the first experiment with the 22 subjects. We determined the identification performance for three standard

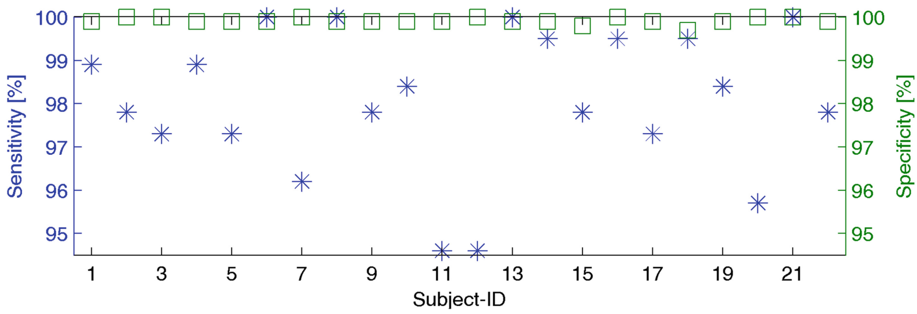


Fig. 9. Class-specific sensitivity (stars) and specificity (squares) results of the identification of the 22 subjects (RF classifier, feature combination $C8$). The sensitivity varied between 94.6 to 99.5 %. The specificity was over 99.7 % for all subjects.

classifiers: ANN, SVM and RF. The classifiers and their parametrization are described in Sect. 2.7. We achieved up to 98.1 % accuracy (see Table 2) with the RF classifier using a feature space combining acceleration and ECG based features. The lowest accuracy of 90.4 % was obtained with the SVM. For all three classifiers, we obtained an overall specificity $\overline{S}$ of more than 99 %.

The class-specific sensitivity (see Eq. 5) of the identification varied between 94.6 to 99.5 % for the different subjects (RF classifier, see Fig. 9). We observed only low deviations in the identification's specificity between the 22 subjects. A class-specific specificity (see Eq. 6) of more than 99.7 % was achieved for all subjects.

We performed a feature selection using the ANN classifier to determine the impact of the different features and to identify combinations C with a high classification performance (see Table 3). We obtained a similar identification accuracy based on acceleration (86.6 %, $C6$) and ECG (84.8 %, $C4$) features. In combination, the accuracy improved to 94.2 % ($C8$).

The ECG contained more information on the subject in the frequency domain than the acceleration measurements (12.3 % higher accuracy). Frequencies of up to 10 Hz contained the most discriminant information of the acceleration measurements. A reduction of the frequency band from 40 to 10 Hz reduced the identification accuracy by only 3.8 %. For the ECG measurements, a reduction from 40 Hz to 15 Hz resulted in a 8.8 % lower accuracy. Overall, we obtained an accuracy of 72.4 % ($C1$) for features extracted from the ECG in the frequency domain.

We found that correlation coefficients describing the similarity between heartbeat segments provide useful insights to identify subjects (80.3 % accuracy, $C2$). To reduce the dimensionality of the feature space, we averaged the correlation coefficients corresponding to the five reference segments per subject. This averaging resulted in a 7.7 % lower accuracy. However, in combination with other features this difference was negligible (0.4 % for $C8$).

Table 3. Identification accuracy for the 22 subjects using different feature combinations C. We obtained a similar accuracy with acceleration and ECG based features (see $C4$, $C6$). Combining both improved the accuracy (see $C8$). The feature selection was performed using the ANN classifier. * denotes the use of the average over the five correlation coefficients per subject.

C	Acceleration feat.			ECG feat.			ACC
	Time dom.	FFT coef.	FFT feat.	FFT coef.	FFT feat.	Corr. coef.	
$C1$	-	-	-	x	x	-	72.4 %
$C2$	-	-	-	-	-	x	80.3 %
$C3$	x	-	-	-	-	-	83.3 %
$C4$	-	-	-	x	x	x	84.8 %
$C5$	x	-	-	-	x	-	86.5 %
$C6$	x	x	x	-	-	-	86.6 %
$C7$	x	-	-	-	-	x	93.6 %
$C8$	x	x	x	x	x	x*	94.2 %

The time domain features calculated from the acceleration signals showed a good accuracy (83.3 %, $C3$). Additional information on the gait in the frequency domain improved the identification accuracy to 86.6 % ($C6$).

By combining the time domain features of the acceleration data with the correlation coefficients derived from the ECG, we achieved a high accuracy of 93.6 % ($C7$), which is only 0.6 % less than using the full feature set ($C8$).

For the time domain features extracted from the acceleration signals, we analysed the impact of the different acceleration axes on the subject's identification accuracy. The highest accuracy was obtained for the anteroposterior accelerations (a_{AP}). The mediolateral accelerations (a_{ML}) showed a 4.4 % and the vertical accelerations (a_V) a 16.2 % lower accuracy.

We additionally evaluated our approach using a hold-out validation for which the data set was split into 66 % training data and 34 % testing data. A hold-out validation avoids temporal proximity between training and testing data and allows therefore a more accurate estimation of the generalization performance. We noted only a slight decrease in accuracy by 0.9 % for the RF classifier.

To estimate the impact of the number of subjects in the data set on the identification performance, we randomly selected eleven out of the twenty-two subjects and repeated the evaluation. With the smaller group of athletes to distinguish, the overall accuracy improved by 1.2 % (RF classifier).

The above results were obtained for the data from the first experiment with one session of recording per subject. However, reapplying the body sensor can change the waveform of the signals due to a different position of the ECG electrodes in relation to the heart, differences in the conductance of the electrodes or alternations in the acceleration sensor's orientation. In order to estimate the

Table 4. Equal error ratio (ERR) and accuracy (ACC) of other gait and ECG based identification methods. For comparison, our results on the data sets from the first and second experiment are listed below. N denotes the number of subjects who participated in the experiments.

	Type	Velocities	N	ERR	ACC
Mäntyjärvi et al. (2005)	Gait	slow, normal and fast walking	36	7.0 %	-
Ailisto et al. (2005)	Gait	normal walking	36	6.4 %	-
Gafurov et al. (2006)	Gait	normal walking	21	5.0 %	-
Rong et al. (2007)	Gait	normal walking	21	5.6 %	-
Chan et al. (2008)	ECG	-	50	-	89.0 %
First experiment (1 session)	Gait & ECG	3, 5, 7 and 9 km/h	22	1.1 %	98.1 %
Second experiment (2 sessions)	Gait & ECG	3, 5, 7, 9 and 11 km/h	9	2.5 %	93.8 %

impact of this variability in the gait and ECG patterns, we evaluated the identification method on the two independent sessions of recording from the second experiment. The first session was used to train the RF classifier and the second session functioned as an independent test set. We obtained an accuracy of 93.8 % with a class-specific sensitivity between 83.6 to 89.7 % for three subjects and above 94.5 % for the other six subjects; the highest sensitivity was 99.4 %. The overall specificity was 99.2 % showing only minor variations between the nine subjects (range: 97.4 to 99.9 %). In comparison to the results from the first experiment, the overall accuracy decreased by 4.3 % (22 subjects) and 5.5 % (reduced subset of 11 subjects) mainly because of the outlying results of the three subjects. However, the results showed that identification is possible with a good accuracy despite reapplying the body sensors and an increased velocity of up to 11 km/h. Furthermore, the second evaluation points out that the identification method is stable for short-term physiological variations in the ECG and alternations in the gait patterns.

In order to compare our results with existing work, we additionally calculated the equal error rate (ERR) of the RF classifier on feature combination $C8$. The ERR is the rate at which both accept and reject errors are equal. For our data set from the first experiment (22 subjects), we obtained an ERR of 1.1 % and for the evaluation on the data from the second session of the second experiment (9 subjects) the ERR was 2.5 %. Compared to other approaches, which are based on only gait characteristics, our achieved ERR is lower (see Table 4). For a comparison of our approach with an ECG based identification we have chosen the method of Chan et al. because the results are also based on data from non-clinical ECGs [7]. With 98.1 % our accuracy is higher than Chan et al. results (89 %). However, with an identification on ECG characteristics only, we obtained a lower accuracy (84.8 %, $C4$). Overall, our high performance is achieved by combining ECG and gait characteristics. We believe that motion related artefacts in the ECG, and a high variability in the gait patterns between changing from slow walking to jogging, reduce the identification performance when we use only one source of information.

5 Discussion and Conclusions

This paper is concerned with the identification of humans during walking and jogging using a single wireless body sensor module attached to a chest strap. Our approach focuses on recognising a human using a biometric measure based on the characteristics in the gait style and the ECG of the human and is hence independent of the used hardware. Thus, our system overcomes the drawbacks of an identification based on the WBS's serial number or an radio-frequency based identification (RFID) which recognises the hardware but not the subject itself.

We have collected data from 22 subjects on a treadmill at velocities from 3 to 9 km/h using a WBS attached to a chest strap. To assure real world conditions, no advice was given on how to perform the exercise and the correct placement of the chest strap was not verified. Despite severe variations in the gait patterns and motion-related artefacts in the ECG, which occur due to real world conditions and physical exercise, our method achieves up to 98 % accuracy.

In order to estimate the impact of variations in the ECG and gait patterns resulting from reapplying the body sensor and short-term physiological alternations, we repeated the experiment with nine subjects with two sessions of recoding per subject which were one week apart. The first session was used for training the classifier and the second session served as test data. With 93.8 % the accuracy of the identification is still high considering also the extended range of velocity classes (3 to 11 km/h).

Our feature selection showed a good identification accuracy for time domain features extracted from the acceleration signals. By using simple and low-dimensional features on the acceleration signal our method can potentially be implemented on computationally constrained platforms, such as a microcontroller on a WBS.

Our identification method can presumably not be extended to an unlimited number of subjects. The individual characteristics in the subject's ECG and gait patterns are extremely difficult to capture and may change over time because of an adaptation to physical exercise. However, we believe our method is well suited to provide an automatic annotation of sensor measurements from several WBSs with the subject's identity for use in sports medicine and athletic training research. Moreover, our method helps to customize a training session by loading personal settings of the recognized athlete on the WBS or other sport equipment.

Our future work includes the evaluation of the identification method within team sports. In particular, we want to recognize handball players in order to support a real-time vision-based tracking of these players.

Acknowledgements. This research was supported by the DFG CoE 277: Cognitive Interaction Technology (CITEC).

References

1. Afonso, V.X., Tompkins, W.J., Nguyen, T.Q., Luo, S.: ECG beat detection using filter banks. Trans. Biomed. Eng. **46**(2), 192–202 (1999)
2. Ailisto, H.J., Lindholm, M., Mantyjarvi, J., Vildjiounaite, E., Makela, S.M.: Identifying people from gait pattern with accelerometers. In: Society of Photo-Optical Instrumentation Engineers, pp. 7–14 (2005)
3. Bianchi, L., Angelini, D., Lacquaniti, F.: Individual characteristics of human walking mechanics. Pflügers Arch. Eur. J. Physiol. **436**, 343–356 (1998)
4. Biel, L., Pettersson, O., Philipson, L., Wide, P.: ECG analysis: a new approach in human identification. IEEE Trans. Instrum. Meas. **50**(3), 808–812 (2001)
5. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York Inc., Secaucus (2006)
6. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
7. Chan, A.D.C., Hamdy, M.M., Badre, A., Badee, V.: Wavelet distance measure for person identification using electrocardiograms. IEEE Trans. Instrum. Meas. **57**(2), 248–253 (2008)
8. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)
9. Christ, P., Mielebacher, J., Haag, M., Rückert, U.: Detection of body movement and measurement of physiological stress with a mobile chest module in obesity prevention. In: Proceedings of the 10th Australasian Conference on Mathematics and Computers in Sport, pp. 67–74, July 2010
10. Christ, P., Werner, F., Rückert, U., Mielebacher, J.: An approach for determining linear velocities of athletes from acceleration measurements using a neural network. In: Proceedings of the 6th IASTED International Conference on Biomechanics, pp. 105–112. ACTA Press, November 2011
11. Conover, M.B.: Understanding electrocardiography. Mosby (2002)
12. Gafurov, D., Helkala, K., Søndrol, T.: Biometric gait authentication using accelerometer sensor. J. Comput. **1**(7), 51–59 (2006)
13. Green, L.S., Lux, R.L., Haws, C.W., Williams, R.R., Hunt, S.C., Burgess, M.J.: Effects of age, sex, and body habitus on QRS and ST-T potential maps of 1100 normal subjects. Circulation **71**(2), 244–253 (1985)
14. Han, J., Kamber, M.: Data mining: concepts and techniques. The Morgan Kaufmann series in data management systems. Elsevier (2006)
15. Hsu, C.W., Lin, C.J.: A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Netw. **13**(2), 415–425 (2002)
16. Mäntyjärvi, J., Lindholm, M., Vildjiounaite, E., Mäkelä, S.-M., Ailisto, H.A.: Identifying users of portable devices from gait pattern with accelerometers. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. ii/973–ii/976, March 2005
17. Møller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. **6**(4), 525–533 (1993)
18. Nixon, M.S., Tan, T., Chellappa, R.: Human identification based on gait, vol. 4. Springer-Verlag New York Inc., Secaucus (2006)
19. Rong, L., Jianzhong, Z., Ming, L., Xiangfeng, H.: A wearable acceleration sensor system for gait recognition. In: 2nd IEEE Conference on Industrial Electronics and Applications, pp. 2654–2659 (2007)
20. Schölkopf, B., Smola, A.J., Williamson, R.C., Bartlett, P.L.: New support vector algorithms. Neural Comput. **12**(5), 1207–1245 (2000)

21. Shen, T. W., Tompkins, W.J., Hu, Y.H.: One-lead ECG for identity verification. In: Proceedings of the 2nd IEEE International Joint Conference on Engineering in Medicine and Biology Society, vol. 1, pp. 62–63. IEEE (2002)
22. Simon, B.P., Eswaran, C.: An ECG classifier designed using modified decision based neural networks. Comput. Biomed. Res. **30**(4), 257–272 (1997)
23. Vidaurre, C., Sander, T.H., Schlögl, A.: BioSig: the free and open source software library for biomedical signal processing. Comput. Intell. Neurosci. **2011**, 12 (2011)
24. Weyand, P.G., Sternlight, D.B., Bellizzi, M.J., Wright, S.: Faster top running speeds are achieved with greater ground forces not more rapid leg movements. J. Appl. Physiol. **89**(5), 1991–1999 (2000)