sensors

Volume 3

Sensors for Gait, Posture, and Health Monitoring

Edited by
Thurmon Lockhart
Printed Edition of the Special Issue Published in *Sensors*

www.mdpi.com/journal/sensors

MDPI

Sensors for Gait, Posture, and Health Monitoring

Sensors for Gait, Posture, and Health Monitoring

Volume 3

Special Issue Editor

Thurmon Lockhart

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Article Number*, Page Range.

Contents

About the Special Issue Editor

Thurmon Lockhart is Professor in the Biomedical Engineering and Biological Design program in the School of Biological Health and Systems Engineering at Arizona State University, Tempe, AZ. He is also Adjunct Professor at the Barrow Neurological Institute, Research Affiliate at the Mayo Clinic College of Medicine, Division of Endocrinology, and Guest Professor at Ghent University in Belgium. His research focuses on the identification of injury mechanisms and the quantification of sensorimotor deficits and movement disorders associated with aging and neurological disorders on fall accidents. His academic grounding in biomechanical modeling, nonlinear dynamics, human postural control, gait mechanics, and wearable biosensor design underscore a fundamental capacity to provide unique clinical solutions to injury prevention utilizing both engineering and biomedical principles. He has translated research findings into practice by reaching a significant number of external organizations and individuals. His outreach efforts have impacted several organizations, including UPS, the US Navy, Los Alamos National Security, the DOE, GE, and BP. In recognition of these scientific achievements, Prof. Lockhart and co-workers were awarded the Alexander C. Williams, Jr., Design Award from the Human Factors and Ergonomics Society in 2008. His research was recently featured on the PBS NOVA ScienceNow and Good Morning America programs and in the Fortune, AgingWell, Men's Health, and Discover magazines.

Preface to "Sensors for Gait, Posture, and Health Monitoring"

The acquisition of gait and postural characteristics during active and passive movements provides important information about limb propulsion and postural control strategies and provides insight into performance and risk of injury. These measures were traditionally assessed by utilizing motion capture systems and force plates. Although modern motion capture laboratories collect precise gait and posture data, they are expensive and immobile and require serial (single person at-a-time data capture) and clustered data collection, limiting the use of motion capture in the field to obtain more realistic motion profiles that may be applicable to various interventions.

As such, in recent years, many technologies for gait and posture assessments have emerged. Wearable sensors, active and passive in-house monitors, and many combinations thereof all promise to provide accurate measures of gait and posture parameters. The objective of this Special Issue is to address and disseminate the latest gait and posture monitoring systems as well as various mathematical models/methods that characterize mobility functions.

This Special Issue explores the core scientific issues associated with the use of custom-designed, wearable, wireless sensor nodes for continuous, non-invasive gait–posture–activity monitoring and analysis in order to accurately study the relationship between these monitoring variables and physical and psychological health conditions to predict adverse medical events in a variety of populations. This type of assessment will dramatically expand the clinical usefulness of these analyses and pave the way for identifying potential adverse health conditions and appropriate interventions for those most at risk.

Thurmon Lockhart
Special Issue Editor

Context Impacts in Accelerometer-Based Walk Detection and Step Counting

Buke Ao [1], Yongcai Wang [2,*], Hongnan Liu [3], Deying Li [2], Lei Song [4] and Jianqiang Li [5]

[1] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; aobuke@hotmail.com
[2] School of Information, Renmin University of China, Beijing 100872, China; deyingli@ruc.edu.cn
[3] School of Software, Tsinghua University, Beijing 100084, China; lhn08@tsinghua.org.cn
[4] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China; leisong03@gmail.com
[5] School of Software, Beijing Institute of Technology, Beijing 100081, China; lijianqiang@bjut.edu.cn
* Correspondence: ycw@ruc.edu.cn

Abstract: Walk detection (WD) and step counting (SC) have become popular applications in the recent emergence of wearable devices. These devices monitor user states and process data from MEMS-based accelerometers and optional gyroscope sensors. Various algorithms have been proposed for WD and SC, which are generally sensitive to the contexts of applications, i.e., (1) the locations of sensor placement; (2) the sensor orientations; (3) the user's walking patterns; (4) the preprocessing window sizes; and (5) the sensor sampling rates. A thorough understanding of how these dynamic factors affect the algorithms' performances is investigated and compared in this paper. In particular, representative WD and SC algorithms are introduced according to their design methodologies. A series of experiments is designed in consideration of different application contexts to form an experimental dataset. Different algorithms are then implemented and evaluated on the dataset. The evaluation results provide a quantitative performance comparison indicating the advantages and weaknesses of different algorithms under different application scenarios, giving valuable guidance for algorithm selection in practical applications.

Keywords: walk detection; step counting; gait analysis; machine learning; signal processing

1. Introduction

We have witnessed the rapid growth and wide popularity of smart watches, smart bands and smartphones in recent years. Walk detection (WD) and step counting (SC) are fundamental applications using these smart devices. They also provide the algorithmic basis for daily activity recognition, calorie consumption estimation, body state diagnosis and indoor navigation.

WD and SC are generally carried out by processing sensing data, from an MEMS-based accelerometer and optionally gyroscope sensors. These sensors are generally tiny and low-cost and embedded in the wearable devices. They provide continuous 1D to 3D accelerator or angle rate data, so as to indicate the motion dynamics of the user who is wearing the device. Various algorithms have been proposed to process the accelerator data stream to extract temporal and frequency features from the sensing data and design recognition algorithms for walk detection and step counting. In particular, walk detection could differentiate walking from other daily activities, such as sitting, standing and running. Step counting counts the number of steps when users are in a walking state.

Algorithms were designed using different ways to process the accelerator stream data, in order to conduct walk detection and step counting. The common design methodologies include heuristic-based, signal processing and machine learning methods. However, a common feature of these algorithms

is that the design is based on recognizing the patterns of a gait cycle. In particular, a gait cycle of a human is composed of several consecutive phases, i.e., stance phase and swing phase. Algorithms recognize and extract the features of these phases to detect walking, and the distinguished periodical events are exploited for step counting.

However, the algorithms' performances are shown to be highly sensitive to the application contexts, and the following contexts are used in this paper: (1) the sensor placements; (2) the sensor orientations; (3) the user's walking patterns; (4) the preprocessing window sizes; and (5) the sensor sampling rates. How the WD and SC algorithms are impacted by the dynamic contexts has not yet been adequately investigated. Most previous works have only treated the sensor placement as the major context factor [1–3]. As in Brajdic's intensive survey [4], they only compared different algorithms and the effects of placements, including [4], who evaluated the performance of step counting algorithms in a loose way. To date, few works have studied the impact of all five application contexts simultaneously within the same experiment. In this paper, we extended the single context factor of placement to five context factors. We also employed a new criterion (Section 5.3.1) to evaluate the performance of step counting.

In addition, existing datasets do not provide enough information in evaluating WD and SC thoroughly. For example, the dataset of [5] is small, and only gestures are collected. USC-HAD [6] fixed the sensor position at the front right hip. The UCI smartphone dataset [7–9] only fixed the sensor at limited positions. The PlaceLab dataset [10] only had one subject. The work [11] lacks step data. To evaluate WD and SC thoroughly, we built our own dataset. We consider context factors including placement, orientation, walking pattern factors and also window size and sampling rate. The window size and sampling rate are determined by the potential applications and are also limited by the hardware. We collected and labeled datasets by the five factors.

In the other applications, WD and SC are fundamental algorithms. For example, [12] used SC to assist in indoor navigation; [13] recognized physical activity such as walking, running, etc.; [14] was a real-time gait analysis in walking and running. The work in [15] could detect falls among activities for elderly people. The work in [16] was a low-cost indoor/outdoor navigation system aided by GPS. Many studies have employed multiple sensors such as the accelerometer, gyroscope, heart rate sensor and a barometer to detect human activity. While placing more sensors in different locations can be cumbersome for subjects, one single accelerometer is more preferable due to the low power consumption and low cost.

From the algorithmic aspect, five algorithms from three main categories (heuristic, signal processing and machine learning) are investigated and summarized in this paper. We carefully design experiments to evaluate the impacts of different contexts thoroughly. In each experiment, testing datasets were constructed by recording multiple users' movement indoors and outdoors, using Android phones and smart watches as the testing devices. The accuracy and fragility of different algorithms provide valuable guidance for algorithm selection for the related applications. Although many commercial wristbands, pedometers and step counting applications have been developed, they usually suffer from false positives. We employ a receiver operating characteristic (ROC) curve to make a more accurate evaluation.

The remaining sections are organized as follows. The related works and context impacts are defined in Section 2. Related algorithms are introduced in Section 3. The experiment design is introduced in Section 4. Performance evaluation and comparisons are presented in Section 5. Section 6 concludes the paper.

2. Take Context into Consideration

2.1. Context Definition

Because the application contexts are diverse and time varying in real applications, they can hardly be considered in the algorithm design phase. However, they do have observable impacts on the WD

and SC performances. To evaluate the impacts of dynamic contexts on the algorithms' performances, the availability of real-time context knowledge is modeled as variables. In this paper, five context variables are considered, as shown in Table 1.

Table 1. Notations and description of context variables.

Notation	Description	Values
I_O	Orientation	known, unknown
I_P	Person info	known, unknown
R	Sampling rate	5 Hz–200 Hz
W	Window size	1.5 s–6 s
L	Wearing location	Jacket pocket (FrontPocket), trouser back pocket (BackPocket), trouser front pocket (UpPocket), foot-mounted (Foot), handheld (Hand), handheld using (HandU).

The knowledge of orientation I_O and person information I_P are modeled by binary variables. I_O indicates whether the sensor orientation is known or not, and I_P indicates whether the model is personalized (trained and tested on each subject); the sampling rate S and window size W are modeled as discrete real numbers, with data range shown in the table; the wearing location is also modeled as a discrete variable, with possible values shown in the table. The available "training" and "testing" data were changed according to the different contexts; for example, if we restrict L to being Foot, then only data collected at Foot is available.

Then, let $\mathbf{x} = \{x_1, \cdots, x_T\}$ be the accelerator data sequence collected from 1–T, $\mathbf{s} = \{s_1, \cdots, s_T\}$ be the user state ground truth during this period and n_T be the step number ground truth. With the consideration of application contexts, $\mathbf{f}_w(\mathbf{x})$ is a specific WD algorithm, which outputs the states of whether the user is walking at each time point from 1–T. $\mathbf{f}_s(\mathbf{x})$ is a specific step counting algorithm, which outputs the estimated number of steps from 1–T. Then, the walk detection error with consideration of context impacts is represented by:

$$e_{w,T} = \sum_{t=1:T} |\mathbf{s}_t - \mathbf{f}_w(\mathbf{x}, I_O, I_P, R, W, L))| \tag{1}$$

The step counting error is represented by:

$$e_{s,T} = \frac{|n_T - \mathbf{f}_s(\mathbf{x}, I_O, I_P, R, W, L))|}{n_T} \tag{2}$$

In the following sections, we introduce walk detection and step counting algorithms, i.e., $\{\mathbf{f}_w(\mathbf{x}), \mathbf{f}_s(\mathbf{x})\}$, and present the performance evaluations of these algorithms for different application contexts.

2.2. Related Works on Context Impacts

Since WD is part of human activity recognition (HAR), we surveyed the HAR works that considered the contexts such as placement and personalization, instead of purely WD. We will also introduce some typical SC works that considered contexts.

(1) Placement is the most common context and is the factor that has attracted researchers' attention. Olguin et al. placed one or two accelerometers on three different parts of the body and studied the classification accuracy of activity recognition [17]. Lester et al. studied whether a single accelerometer could generalize well on different locations and the reliability of activity recognition on a novel individual [18]. The works in [2,19] explored the influences of placements on different body parts. The work in [1,3] showed that the negative influence of various placements of the sensor could be mitigated. Cleland et al. studied the optimal placement to detect daily activities [20]. Sun et al. investigated the effects of varying positions and orientations on the accuracy of activity recognition [21].

(2) The impact of personalization was investigated in [22]. The work in [22] compared the impersonal model and personal model. The impersonal model was built using the data from many users and tested on a new user; the personal model was a personalized method built with the data from the specified user and tested on him/herself. The result showed that the personalized method was much better than the impersonal model. By using active learning and semi-supervised learning algorithms, [23,24] in fact developed a personalized model based on a original classifier and showed the significant improvement over the original classifier, which was trained on the data from many users.

(3) The impacts of window size were evaluated in [21,25,26]. Although [25] demonstrated that the accuracy was nearly the same under different window sizes, these findings contradicted those of [21,26]. The main reason is that the contexts of the three works were different, which shows the importance of conducting complete evaluations under various contexts.

(4) The impacts of multiple contexts were also investigated in some existing works. The work in [3] investigated the placements, feature selection and the window on/off on the accelerator to evaluate the accuracy of activity recognition. The research showed that the accuracy at the trouser front pocket position had lower accuracy to classify activities and also had difficulty in distinguishing normal walking and fast walking. In addition, the work showed that the classification accuracy between standing and sitting could be significantly enhanced if the sensor position were considered.

Kunze [1,2] evaluated the context impact of placements and orientations. The work considered the placement including head, trousers, torso and wrist. It showed that the displacement of sensors could harm the accuracy, but this could be mitigated by extracting placement-independent features and placement recognition. Besides, it showed that the closely related placements usually generated misclassifications.

For step counting, the Pan-Tompkins method (PTM) [27] only mounted the sensor at the foot and reached a high accuracy in SC. In [28], each subject wore the sensor on his/her waist, and then, the activities were classified and steps counted.

It can be seen that previous activity recognition studies mainly conducted the experiments under a few contexts such as classifiers, placements and orientations of the sensor; therefore, in-depth studies considering complete contexts are needed. Besides, The majority of most past research only studied HAR, which made the evaluations on both WD and SC under various contexts necessary. In our paper, we considered more comprehensive application contexts that include classifiers, placements, orientations, window size, sampling rates and personalization to make a complete comparison.

3. WD and SC Algorithms

Walk detection (WD) and step counting (SC) algorithms were generally designed following a similar routine: (1) feature extraction; (2) feature detection; and (3) state recognition. In an implementation, the design methodologies can be roughly categorized into: (1) heuristic-based; (2) signal processing based; and (3) machine learning-based. The related algorithms are briefly summarized and introduced.

Since we aim to give a comprehensive evaluation of various contexts, a proper set of algorithms that could exactly reflect the contexts' changes should be selected. Some features and algorithms that are too sensitive or not widely applied could not provide fair comparisons of different contexts. Too complicated features and some 'best' algorithms should be avoided because they may generate biased results. Therefore, we surveyed many papers on both feature extractions and classification algorithms and selected the most stable features and algorithms to conduct the experiments and evaluations.

3.1. Feature Extraction Techniques

Feature extraction from the accelerator data stream is crucial for walk detection and step counting. Various features have been presented in the literature. Statistical features in the time domain, including mean,

variance, correlation, skewness, kurtosis, energy, etc., were proposed in [29–34]. Other features such as peak interval and zero/mean-crossing rate were also proposed in [9,35]. Additionally, root mean square (RMS) and histogram were proposed in [36].

Features in frequency and transformed domains were also proposed in many works. FFT bins were used in [28,37], and wavelet coefficients were introduced in [38,39]. The peak frequency and power ratio of different frequency bands were exploited in [37]. Mel-frequency cepstral coefficients (MFCCs) and Bark-frequency cepstral coefficients (BFCCs) as complex features of frequency domains were also proposed in [36].

Besides these conventional features, principal component analysis (PCA) was proposed in [40], although it is commonly used as a feature selection method. Autoencoder networks [40] and sparse coding [41,42] have also been introduced recently. Furthermore, some manually designed features such as weightlessness features were used in [43].

We split the common features of WD and SC into three groups in Table 2.

Table 2. Feature categorization of walk detection (WD) and step counting (SC). BFCC, Bark-frequency cepstral coefficients.

Category	Features
Time Domain	mean, variance, peak, peak interval, skewness, kurtosis, energy, entropy, correlation coefficients, RMS, zero/mean crossing rate
Frequency Domain	FFT bins, wavelet coefficients, MFCCs, BFCCs, peak frequency, spectral entropy, power ratio of different frequency bands
Other	PCA, autoencoder networks, sparse coding, weightlessness feature

3.2. Related Algorithms

3.2.1. Heuristic Methods

Heuristic methods build a series of rules that leverage the cyclic patterns in the time domain to perform walk detection and step counting. The representative algorithms include the multiple threshold method (MT) and the finite state machine (FSM) method. The multiple threshold method, which was proposed by Kim et al. [44], makes use of the cyclic peaks, valleys and thresholds to count steps. The finite state machine (FSM) by Alzantot et al. [35] sets some thresholds in the magnitude to drive an FSM to count steps.

In addition, Randell et al. [45], Bylemans et al. [46] and Ailisto et al. [47] proposed algorithms to detect the step event by finding the consecutive local maxima and minima of the low-pass version of the sensor signal. Beauregard et al. [48] found the positive-going zero-crossing event that indicates the boundaries of each step cycle to count steps. Ying et al. [27] detected the negative peaks that were caused by the heel-strike event to count steps. The correspondence between a peak value and a step was shown in the study of Goyal et al. [49], which finds the peak within one zero-crossing interval when the sensor is placed at the pelvis.

3.2.2. Signal Processing Methods

Signal processing techniques were also exploited to detect walking and to count steps, generally in a transformed domain, by methods such as fast Fourier transform (FFT), short time Fourier transform (STFT) and discrete/continuous wavelet transform (DWT/CWT). Matching methods, such as auto-correlation, cross-correlation, template matching, dynamic time warping (DTW), etc., were also used.

The Pan-Tompkins method (PTM) [27] uses a series of a filter, integration and derivative module to extract step events. STFT [28,50,51] exploits the energy ratio of different frequency bands to perform walk detection, and the period information is used to perform step counting. DWT/CWT [50,52]

decompose the original signal into multiple resolutions of the frequency and time domain, which could discriminate walking activity by comparing the ratio between different wavelet coefficients.

Autocorrelation, cross-correlation, template matching and DTW all exploit the similarity between a predefined typical signal of a step cycle and the test sensor data to count steps. The autocorrelation method [53] thresholds the coefficients to detect walk activity and count steps by using the repetitiveness walk activity. Cross-correlation and template matching [27,54] threshold the high positive correlation coefficients to count steps. Ying et al. [27] extracted the first step cycle as the template and computed the normalized cross-correlation to count steps. Although these methods are accurate, the predefined typical template is different in various contexts and hard to find. Similarly, the DTW [55] method measures the similarity between a predefined typical template and the test sensor data, which is time-invariant and robust at various speeds.

3.2.3. Machine Learning Methods

Machine learning techniques such as supervised learning, unsupervised learning, online learning and transfer learning have been investigated for activity recognition and walk detection.

Supervised learning such as decision trees (DT) [56], neural networks (NN), support vector machines (SVM), Gaussian mixture models (GMMs) [32], k-nearest neighbor (KNN) [29], naive Bayes classifiers [28] and boosting methods [57] have been studied and generally achieve good detection accuracy.

Unsupervised methods have been used in activity recognition. For example, hidden Markov models [36,58–60] are powerful in sequence data analysis; they could also be used to model the walking activity. K-means clustering [61,62] clusters the data in feature space, where the activities could be identified. Although the recognition accuracy of unsupervised learning is generally lower than supervised learning [36], it exempts people from the costly work of labeling the training data.

Besides these typical learning approaches, Cheng et al. [63] investigated zero-shot learning that could recognize unseen new activities when there were no corresponding samples in the training dataset. Rebetez et al. [64] introduced growing neural gas (GNG) to build an online learning recognition system that did not require labeled data. Transfer learning [65,66] could transfer activity recognition from one domain to another domain, which adapts the changes of sensor position [33,67], activity type [68] or environment scenario [69].

3.3. Selected Algorithms for Comparison

We choose some representative algorithms in each category to evaluate the context impacts considering the complexity and practicability. In heuristic methods, threshold (THR) and FSM [35] were selected because they are intuitive and simple to implement in real systems. In the signal processing category, STFT [50], DWT [50] and PTM [27] were selected. We chose STFT and DWT because they are popular and simple to implement. We chose PTM because it has a refined processing chain. In the machine learning category, we chose KNN [29] and SVM as the simple model and complex model, respectively.

These algorithms are chosen from simpler to more complex models considering the requirements and resources of a real system. The evaluation routine is shown in Figure 1.
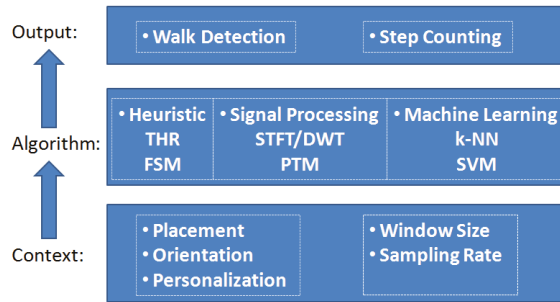
Figure 1. The method of context impacts the algorithm and output. THR, threshold; PTM, Pan-Tompkins method.

4. Experiment Design

In order to evaluate the contexts completely, we need to design the experiments carefully and correctly. First, we will give a detailed definition of walk and how to distinguish walk from other activities. Then, by simplifying the problem of complex and varying contexts, we could make sure our evaluation is reliable. Last, extensive experiments are conducted, and we will show how to set the parameters of data-preprocessing, feature extraction and model training. We also give an interpretation of false positive in step counting.

4.1. Experiment Settings

4.1.1. User Activity Categorization

Considering the periodic pattern of the walk, we categorized daily activities into two groups: periodic activities and non-periodic activities. Additionally, according to the similarity among these activities, we categorized them into four classes: walk, walk-like activities, walk-related activities and walk-unrelated activities. We show these in Figure 2. Walk-like activities are an extension of walking, which includes going upstairs and downstairs, and could be used in some undemanding systems. Walk-related activities contain those that might be misclassified by algorithms in prior studies, such as running, riding a bicycle and brushing teeth.
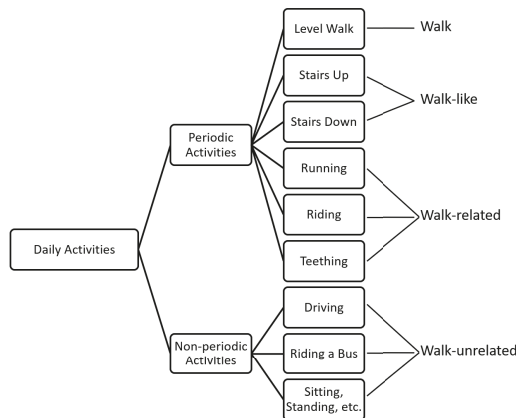
Figure 2. Hierarchical walk activity definition.

In fact, it is quite easy to distinguish walking and walk-like activities from walk-related and walk-unrelated activities. Therefore, we largely simplified the analysis and data collection of walk-related and walk-unrelated in the remainder of the article.

4.1.2. Problem Simplification

In a rigorous way to evaluate the context impacts, we should depict the performance function $f(x_1, x_2, \ldots, x_i, \ldots, x_n)$, $1 \le i \le n$, where x_i is the i-th context variable such as placement, orientation, sampling rate, subject, algorithm, age, gender, place of experiment, time of the experiment, sensor type and walking velocity. However, it is impossible to collect data under all these context variables. Instead, we should simplify the experiment by restricting some conditions:

- Sample data at 200 Hz, which is nearly the highest sampling rate of most devices, and downsample it to the evaluation.
- Carry one or two devices at one time and repeat it to cover all the six placements defined in Table 3, since the acceleration data in consecutive rounds within a same building are similar.
- Device orientation of the same placement across different subjects is the same.

Table 3. The context variables in the experiments.

Context Variable	Settings
Activity	Walk, non-walk (stairs up, stairs down, riding, etc.)
Placement	UpPocket, BackPocket, FrontPocket, Foot, Hand, HandU
Sampling rate	10 Hz, 50 Hz, 100 Hz, 200 Hz
Orientation	Free direction
Individual difference	Age, gender, height, weight, etc.

In order to evaluate the context impacts of these factors, we should collect data that traverse all the possible cases. Therefore, by exploiting those preliminaries, we conduct our experiments as follows.

4.1.3. Experiment Scheme

In order to collect data for our context impact evaluation tasks, it is necessary to have many different kinds of people participating in our experiments. A total of 15 subjects including 10 males and 5 females, with ages ranging from 18 to 28, heights ranging from 1.6 m to 1.85 m and weights ranging from 45 kg to 90 kg, participated in our experiments. Under the assumptions highlighted in Section 4.1.2, we conducted our experiments as follows. Each subject was required to carry multiple smartphones mounted at different positions on the body and walk continuously alongside an indoor track as in Figure 3. This included walking through corridors, going upstairs, going downstairs and going back to the starting point. As we did not use six devices to collect data simultaneously, multiple rounds of walking with the devices mounted at different locations were required to cover all six placements in the data collection. We also employed a camera in this indoor track to obtain the ground truth of step counting. Besides, activities such as running, riding, brushing teeth and driving, as shown in Table 4, do not need to be collected at all six placements, because the signal is similar under different positions. Hand is the subject carries the smartphone in their hand naturally, which is mainly used to simulate the position of the wrist band. Handheld using (HandU) is the subject carries the smartphone, as well as watches the screen, in order to reflect typical walking and using states. The smartphones are not limited to the left or right side; the individual just behaves naturally since we observe that the signal is similar.

Using these assumptions and experiments, we simplified the data collection, and the data could reflect and represent the real contexts well. In the comparison of each context, we only selected data that were generated under the specific context to train the model. For example, we could simply only

use the data collected at FrontPocket to train and test the model when we want to get the accuracies of the placements of FrontPocket.

Table 4. Mounted placements of activities in data collection.

Activities	Covered Placements
Level Walk, Stairs Up, Stairs Down	Hand, HandU, BackPocket, FrontPocket, UpPocket, Foot
Running	Hand, BackPocket, FrontPocket, UpPocket, Foot
Riding, Brush Teeth, Driving, Riding bus, Sitting, Standing	FrontPocket, Hand
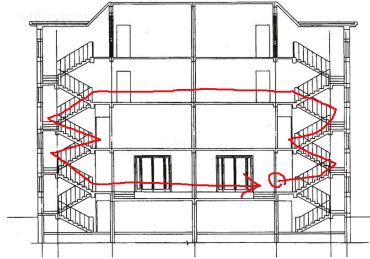
Figure 3. Designed indoor trajectory in data collection.

4.2. Data Pre-Processing

We calculate the magnitude of a tri-axis accelerometer to remove the orientation constraints in WD and SC tasks except for the evaluation of orientation.

$$d = \|Acc\| = \sqrt{x^2 + y^2 + z^2} \tag{3}$$

For WD, the raw data of the sensor are first filtered by a low-pass filter with a cut-off frequency of 15 Hz and then segmented into frames of 3 s with 0.5 s overlap. For the evaluation of window size, overlap is always one-sixth of the window size. We record the start time and end time of each activity and then label the data. The activity recognition is performed based on these frames and the labels.

For SC, contexts such as subjects, placements, movement intensities and speeds would cause the deviation of the amplitude, variance, maximum and minimum of the raw sensor data; thus, normalization is necessary. We normalize the raw data by variance since it outperforms other normalization methods such as maximum, minimum and amplitude.

4.3. Feature Extraction

Feature extraction is a crucial part of machine learning and has a great influence on the classification performance. Instead of achieving a high accuracy by some complex features in Section 3.1, we aim to evaluate the context impacts and use some popular features in our research. Mean, variation, min, max, energy, skewness, kurtosis, FFT amplitudes, mean-crossing rates and RMS were extracted from the sensor data as features.

4.4. Algorithm Parameter Setting

We empirically optimize the parameters of each algorithm and show the implementation details in this section.

THR is used for walk detection. We heuristically threshold the magnitude variance of the sensor data. The optimal threshold value is chosen by exhaustively searching within the range of minimum and maximum variance.

FSM is used for step counting. It has four thresholds that determine the state sequence of the input data stream [35]. We conduct a grid search for the four thresholds used in the model and choose the thresholds that have the best accuracy.

PTM is used for step counting. It includes a series of models such as low-pass filter, differentiator, squaring and integrator, which could find the large sloping part in the sensor data. One local maximum of PTM output means one step. We have the best accuracy when the filter order is 200, the cut-off frequency is 50 Hz and integrator window size is 0.5 s.

STFT is used for both walk detection and step counting. Walk detection by STFT in Barralon et al. [50] places the sensor on the chest, while we do not make this restriction. We first compute FFT coefficients in each frame, then detect walk by setting a threshold on the frequency energy in [0.66 Hz, 1.66 Hz], which outperforms the energy ratio method in [50].

For step counting, we also abandon the placement on the waist, as in [28], and place the sensor at various positions. We add a differentiator module after the energy-based filter that accounts for 20% of the full energy to achieve better performance.

DWT is used for both walk detection and step counting. The decomposition is eight levels by the dh10 wavelet. Instead of comparing the ratio of detail power coefficients, we first smooth the whole energy samples of seven and eight levels' details of each sample and then detect the walk by considering the mean energy of each window.

For step counting, we reconstruct the signal by $6, 7, 8$ and $2, 3$ levels' details, respectively, as two methods, which are denoted as DWT and DWT2, respectively, in the following.

SVM is used for walk detection, and the Gaussian radial basis function (RBF) kernel performs best in our experiments.

k-NN is used for walk detection, and $k = 5$ performs best in our dataset.

Besides the parameters that we mentioned here, there are also many parameters of each algorithm to be decided such as filter order, filter coefficients, differentiator coefficients, etc. In different contexts, we empirically choose these values to achieve the best accuracy.

4.5. Counting

Walking is a repetitive activity, which makes the sensor data cyclic. We could see features such as large slope changes, local maxima, local minima, peaks, valleys and mean-crossing events in the magnitude of each gait cycle, where some features could be used to identify and count step cycles. In fact, we expect to detect only one representative event in a gait cycle. However, those features are generated by physical movements, which is intrinsically not stable or uniform distributed in each gait cycle, might emerge more than once in one gait cycle and be sensitive to contexts. For example, more than one large slope change, local maxima, peaks, etc., are observed in one gait cycle because of the context impacts, as well as noises in the physical movement. Therefore, although the algorithm could detect all those features in each gait cycle, step counting is still not accurate since two or more large slope changes might exist in the signal of one gait cycle. In fact, each SC algorithm could detect one feature, for instance each peak in the output of PTM means a large slope change exists in the corresponding position of the input signal. Therefore, simple features, which might emerge more than once in one gait cycle, are inclined to have high false positives; while complex features, which might not appear in a gait cycle, tend to have low true positives. Thus, we need more approaches to balance false positive rate and true positive rate and evaluate the algorithms.

In an algorithmic view, each feature in the gait cycle has one peak in the output, so there might be multiple peaks in the output of each gait cycle. In order to remove the peaks that are generated by noises, we use not only a threshold, but also the minimum peak distance, minimum peak height and minimum peak prominence. These conditions could remove most false positives without the loss of true positives and achieve a better trade off.

5. Evaluation Results

We first give a figurative example to present the WD. We apply THR and SVM walk detection algorithms to the continuous activities in a real indoor scenario. The activities consist of a walk, using a phone, going up stairs, going down stairs and some temporal irregular activities such as pushing the door or handshake.

The results are shown in Figure 4. The windows surrounded by red and green rectangles indicate the walking state recognized from SVM and THR, while the low values out of the rectangles are non-walk activities. We saw that SVM provides better classification accuracy than the threshold method in this example. More details will be shown in the following experiments. We could find there are some jitters along the timeline in both algorithms; in addition, the estimated start and end times of walking may deviate from the ground truth.
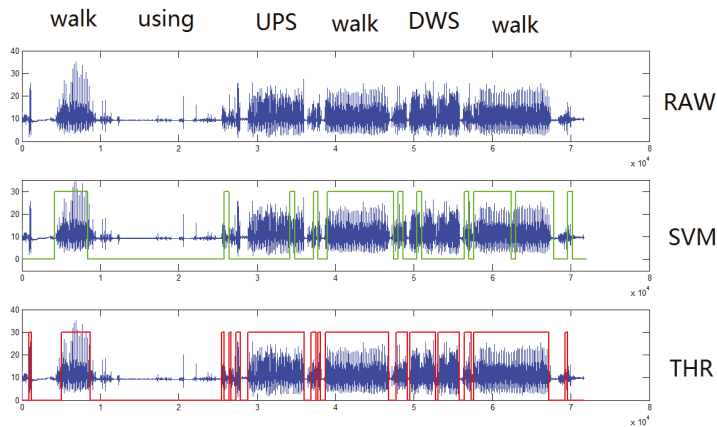
Figure 4. Example of walk detection in a real timeline.

We have defined four groups: walk, walk-like activity, walk-related activity and walk-unrelated activity. We first address a coarse-grained WD problem that distinguishes walk and walk-like activity from walk-related activity and walk-unrelated activities.

5.1. Coarse-Grained WD

Coarse-grained WD is defined to distinguish walk and walk-like activities from walk-related and walk-related activities, which is a coarse classification of walk.

In Table 5, we could find that machine learning methods could easily distinguish walk and walk-like activities from walk-related and walk-unrelated activities. STFT performs better than DWT since walk features are better discriminated in the frequency domain than the time domain.

Table 5. Accuracy of coarse-grained walk detection.

	THR	STFT	DWT	k-NN	SVM
Accuracy	77.55%	85.3%	80.7%	96.91%	97.5%

5.2. Context Impacts on Fine-Grained WD

We defined fine-grained WD in order to distinguish a walking activity from a walk-like activity and riding a bicycle. The baseline performance (accuracy) under a predefined context is used to compare with the new performances when more contexts such as placement and orientation are available. Since there are more data on walking than on walk-like activities, we balance the ratio to 1:1 by random selection.

5.2.1. Baseline Performance

The baseline performance of fine-grained WD algorithms are obtained under the sampling rate $R = 200$ Hz and window size $W = 3$ s (600 samples), while orientation I_O and personalization I_P are unknown and L is not specified. All data are trained and tested by 10-fold cross-validation, and the result is shown in the accuracy-base row in Table 6.

Table 6. Context effects: orientation and personalization.

	THR	STFT	DWT	k-NN	SVM
Baseline Accuracy	62.55%	70.96%	67.66%	82.61%	85.57%
Accuracy (I_O is known)	62.22%	62.95%	68.91%	92.27%	93.07%
Accuracy (I_P is known)	71.03%	83.15%	77.44%	97.62%	96.86%

5.2.2. Context Effects

We first evaluated the effects of orientation I_O and personalization I_P. Since we could recognize all three axes, if I_O is known, then extract features and train all of these on three axes and compare them to the baseline accuracy, which is calculated based only on the magnitude of the sensor data. For the heuristic method and the signal processing method, we chose an axis that performs best on all three axes. Additionally, the accuracies under I_P is known are obtained by averaging the accuracy of the test data on each subject. We show the results in Table 6.

If the prior information of orientation I_O and personalization I_P are independently known, then the performance of WD algorithms is shown in Table 6. If I_O is known, then the feature extraction of machine learning algorithms is performed on all three axes, which means that the classifier is trained in the designed orientation. For the heuristic method and signal processing method, we chose the axis that performs best from a gravity axis and a forwarding axis.

We found that both heuristic methods (THR) and signal processing methods (STFT, DWT) are worse than machine learning methods (k-NN, SVM). In fact, heuristic methods and signal processing methods could be viewed as features of machine learning methods. We found that either providing orientation I_O or personalization I_P could enhance the recognition accuracy, and I_P contributed more to the accuracy increasing. If we know the sensor orientation I_O, there are many techniques to employ them such as extract features in all three dimensions or reconstruct the signal into the Earth coordinate system. Here, we use the common method to extract features in three dimensions of the sensor to observe the impacts. The orientation is different from placement since the sensor could be attached freely without control while the typical placements are fixed, so we could not give comparisons in every orientation. if I_P is provided, we only use one's data to train the personalized model and test the model on him/herself.

Figure 5 shows the baseline accuracy of algorithms over different window sizes, which varies from 300 to 1200 samples (1.5 s to 6 s). The heuristic methods (THR) are nearly independent of the changing of the window size because the features are nearly time-invariant. The STFT method is best at the window size of 600 samples (3 s). The DWT method becomes slightly higher alongside the increase of window size. Machine learning methods overall become slightly lower along with the increase of window size mainly because the statistical features of different activities in a larger window size are not that easy to discriminate.

Figure 6 shows the relationship between various placements and accuracy. To evaluate the accuracies of the placements, we selected data only from the specified position to train and test the model and use the accuracies as criteria. The knowledge of placements boosts almost all the accuracies compared to the baseline accuracy, which indicates that placements make the dataset more discriminative. Although some placements slightly increase the accuracy of the heuristic method (THR), it is not sensitive to variations of placements. The STFT method and DWT method perform even worse than the baseline accuracy at some placements. The machine learning methods (k-NN and

SVM) display the largest increase in Figure 6. Lastly, the accuracies of STFT, DWT, k-NN and SVM all increased at FrontPocket.
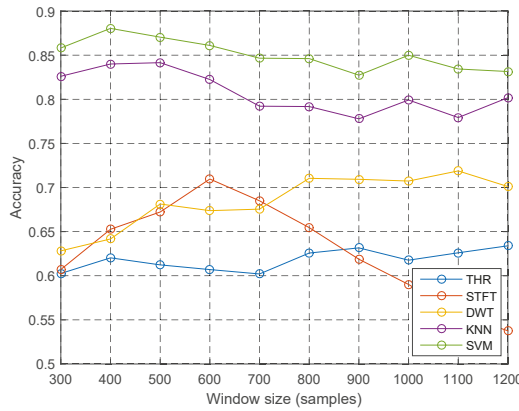
Figure 5. Context effects: window size.
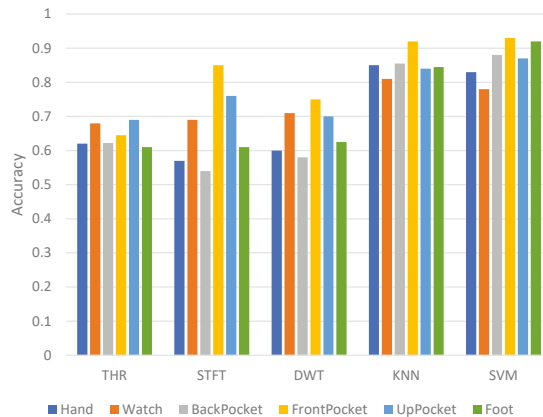
Figure 6. Context effects: placement.

Figure 7 depicts the baseline accuracies under different sampling rates R. The accuracy decreases smoothly with the increase of R. Furthermore, 20 Hz is the transition point, where the accuracies of k-NN, SVM and STFT diminish quickly. THR is nearly independent of the changes of R, since the variance is steady. DWT becomes relatively low when the sampling rate is 100 Hz.

We observed that the machine learning algorithms outperform heuristic and signal processing methods in distinguishing walking activity from other periodic and walk-like activities. This is mainly because the heuristic and signal processing methods suffer from the indistinguishable patterns of variance and spectrum between walking activity and other activities. Besides, the accuracies of heuristic methods and signal processing methods are not noticeably improved even though more contextual information is provided. Besides, the false positives and false negatives of riding a bicycle are much lower than other activities because the sensor movement in the FrontPocket is highly restricted, and false positives are difficult to recognize.

The accuracy of heuristic methods and signal processing methods is a result of a balance among true positives, true negatives, false positives and false negatives. We chose the best accuracy that was larger than 50%, under the condition of both true positives and true negatives.
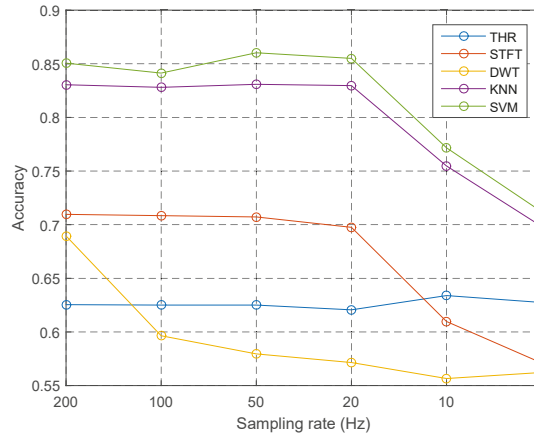
Figure 7. Context effects: sampling rate.

5.3. Context Impacts on SC

The context impacts of SC are performed only on the walking data, without other activities such as going upstairs, going downstairs and riding.

5.3.1. Definition

We propose two definitions to depict the accuracy of step counting algorithms and choose the strict definition in our evaluation.

Loose definition: The most intuitive way to evaluate step counting algorithms is to compare the estimated step counts to real step counts. This accuracy was presented by Brajdic in [4] and similar definitions are in [27,44,51]:

$$\frac{C_{est} - C_{gt}}{C_{gt}} \times 100\%$$

where C_{est} and C_{gt} are the counts of the estimated steps and ground truth steps, respectively. However, this criterion does not consider the false positives during step counting, so it is necessary to introduce a strict definition.

Strict definition: In a walking activity with uniform velocity, we manually mark each gait cycle in the sensor data. Normally, the algorithm could only count one step in one gait cycle, but this may not be true. Suppose that n steps are counted within one gait cycle, and if $n > 0$, then there is one true positive step and $n - 1$ false positive steps. Considering this situation, we introduce an ROC curve that includes both the false positive rate and true positive rate to evaluate the step counting algorithms.

$$\text{TPR} = \frac{C_{tp}}{C_{gt}} \tag{4}$$

$$\text{FPR} = \frac{C_{fp}}{C_{gt}} \tag{5}$$

where C_{tp} and C_{fp} are the true positive and false positive counts, respectively; TPR and FPR are the true positive rate and false positive rate, respectively. Note that FPR might be larger than one since $C_{fp} > C_{tp}$ is possible.

Comparison: All SC algorithms in our paper result in finding the local peaks of a signal, and one step is detected if the peaks are larger than the threshold. Figure 8 shows two examples that have four

gait cycles. Compared to the ideal case in Figure 8a that each gait cycle only has one peak, Figure 8b still counts four steps, although one step is lost in the third gait cycle. Thus, our strict definition is more comprehensive to evaluate the step counting algorithms.
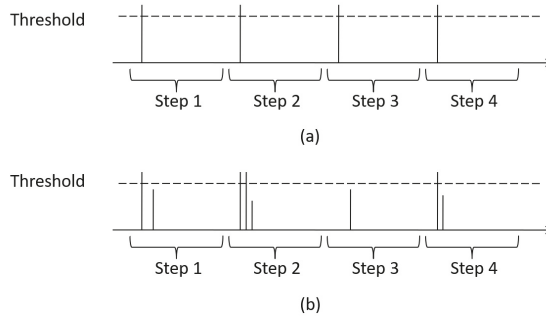
Figure 8. (**a**) Ideal Case of SC Loose definition; (**b**) Bad Case Needs SC Strict definition.

We find that the pattern of the step signal has two different groups, which depends on the placements of the sensor. When the sensor is placed at Foot, FrontPocket, BackPocket and Hand, we could observe one period in one gait cycle of one leg. When the sensor is placed at UpPocket or Hand, we could observe two periods in one gait cycle of one leg because the movement of the other leg also has a period. Based on these observations, we separate the placements into two groups: Group I and Group II.

5.3.2. Baseline Performance

The baseline performance of SC algorithms is obtained under the sampling rate $R = 200$ Hz, orientation I_O is unknown, personalization I_P is unknown, and L is in Group I or II, where Group I contains Foot, FrontPocket, BackPocket and Hand and Group II contains Hand and UpPocket. The baseline performance of Group I is in Figures 9 and 10; the baseline performance of Group II is in Table 7. For Group I, we could observe one period in one gait cycle of the leg that has a sensor; while for Group II, we could observe two periods.
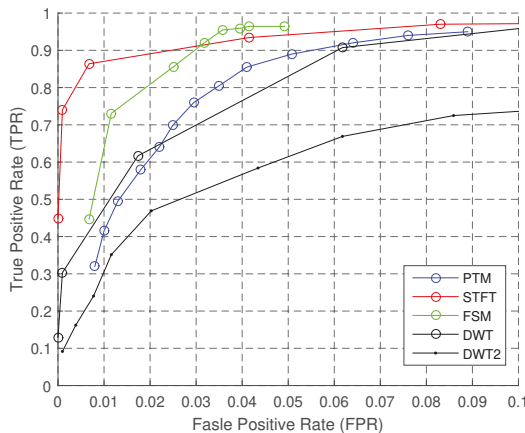
Figure 9. ROC of step counting (Group I).

Figure 9 shows the baseline performance of the algorithms under Group I. PTM, STFT and FSM have higher TPR when FPR is 5%. The overall performance of STFT is best, and it keeps improving along with the increase of FPR. FSM has the highest accuracy, while it is not robust, since it performs poorly when the FPR is low; however, this means that one could obtain better accuracy by fine-tuning the parameters. DWT2 performs poor in this case, because the details are not stable features in the signal of a gait cycle. Although PTM includes a series of elegant signal processing modules, the TPR is good only in a short interval ($fpr \approx 4.5\%$).
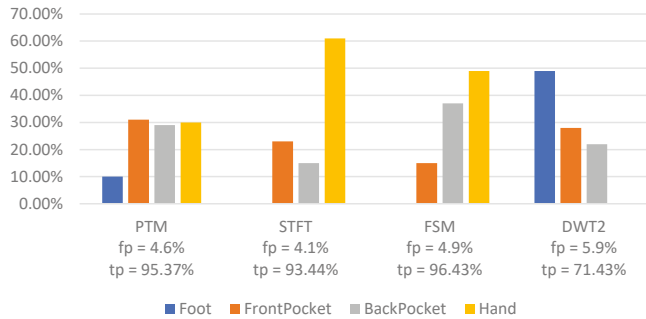
Figure 10. Error sources distribution (fp $\approx$ 0.05).

Figure 10 reveals the error proportion of each placement that accounts for the total of 5% false positives (approximately). PTM is the most stable algorithm, and most false positives of STFT and FSM happen at Hand.

Table 7 presents the baseline performance of the algorithms under Group II. We could see that these algorithms outperform Group I remarkably, mainly because the features in each gait cycle are consistent under different contexts such as placements and subjects.

Table 7. Performance of step counting (Group II).

	PTM	STFT	FSM	DWT	DWT2
TPR	91.0%	97.2%	95.7%	98.1%	80.0%
FPR	0.3%	0.52%	0.1%	0.3%	0.3%

5.3.3. Context Effects

When the placement L is known, we based the models and corresponding parameters on the data generated by the specified position. Similarly, if the personalization I_P is known, then we carried out the same process on the data of each person plus an average process. If the orientation I_O is known, we ran the algorithms on all three axes and chose the axis with the best accuracy.

Figure 11 displays the performances of algorithms when the personalization I_P is known. The ROC curve is better than the baseline performance. Compared to the baseline performance, the TPR of FSM and STFT is larger than 95%, and the FPR is lower at the same time. Although the TPR of DWT increases with FPR, the overall performance is poor.

Table 8 shows the TPR and FPR when the sensor is mounted on the foot. In this situation, all algorithms perform excellently except DWT and DWT2 because the detailed component of the sensor data is not remarkable. The overall performance ranking of algorithms is similar to the former experiment: STFT > FSM > PTM > DWT2 > DWT. Sensor data under this circumstance are very regular, which leads to a much better performance.

Figure 12 exhibits the ROC curve when the sensor is in the FrontPocket. The performance ranking is similar: STFT > FSM > PTM > DWT > DWT2. Unlike those former figures, STFT outperforms

other algorithms overwhelmingly. FSM and PTM have a high TPR within only a short interval of FPR, which means instability in real applications.
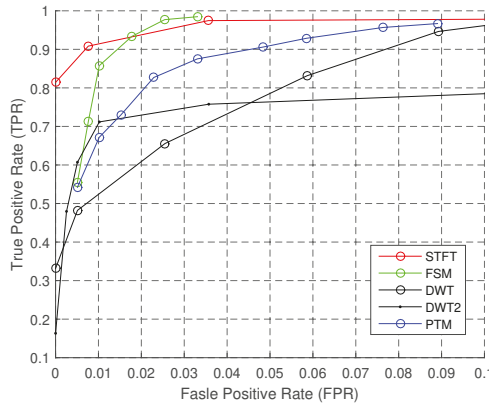
Figure 11. ROC of step counting: personalization.

Table 8. Step counting: Foot.

	PTM	STFT	FSM	DWT	DWT2
TPR	98.7%	99.57%	99.13%	92.61%	95.65%
FPR	1.73%	0%	0%	5.21%	3.47%

Figure 13 illustrates the performance of algorithms when the sensor is Hand. DWT outperforms other algorithms, and FSM is more stable than prior placements. STFT does not perform as well as other prior placements. In this case, all algorithms suffer from higher FPR compared to other prior placements, because the hand movement is diverse.
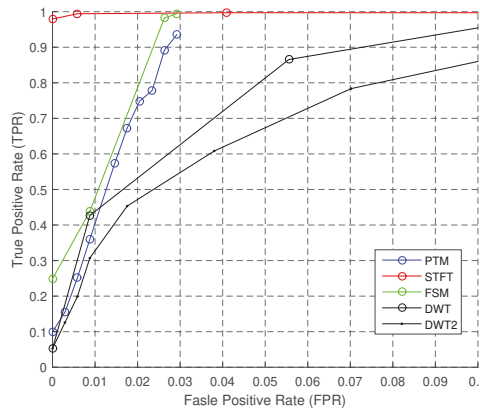
Figure 12. ROC of step counting: FrontPocket.

Figure 14 demonstrates the ROC curve when the sensor is in BackPocket. None of the algorithms display great differences under this situation, except that DWT2 performs too poorly to present. We could observe that it is nearly impossible to achieve a high TPR when FPR is low, while the TPR is acceptable at FPR $\approx$ 2%. Furthermore, the TPR is almost unchanged, although we allow larger FPR. We could find that all algorithms' TPR is high at the point of $fpr \approx$ 2%.
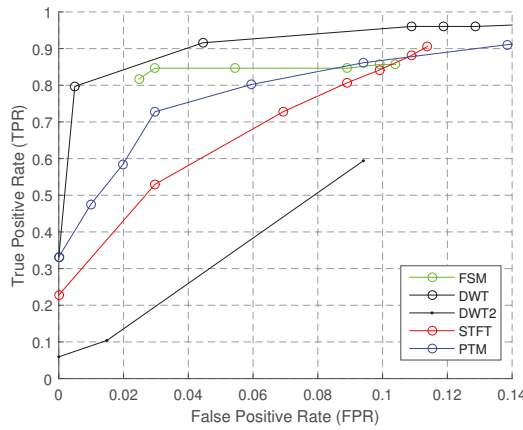
Figure 13. ROC of step counting: Hand.
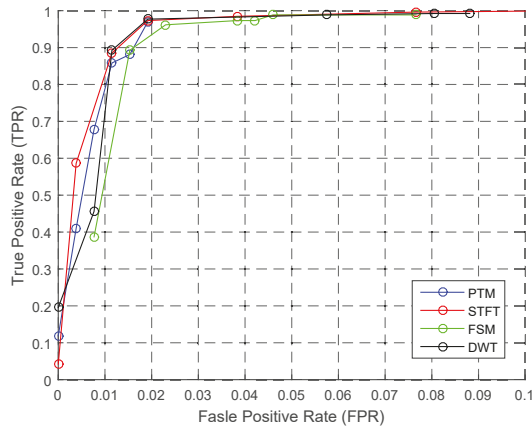
Figure 14. ROC of step counting: BackPocket.

Since it is difficult for people to do anything for a long time when walking, we abandon the evaluation of placement of HandU. Besides the influence of placements, we also investigate the influence of sensor orientation and signal sampling rate and directly gather the results into Table 9 which will be explained in the next section. If orientation is provided, we choose the axis that has the best accuracy.

Tables 9 and 10 are calculated by comparing the new performances when more contexts are available, in addition to the baseline performances.

Table 9. SC accuracy under various contexts (FPR ≈ 3%).

Group	Algorithm	Orientation	Personalization	Placement (Hand, Foot, etc.)	Sampling Rate (5 Hz to 200 Hz)
Heuristic Method	FSM	+1.5%	+9%	[−5%, +10%]	[−32%, +0%]
Signal Processing	PTM	+1%	+16%	[−49%, +17%]	[−28%, +0%]
	STFT	+3%	+5%	[−46%, +9%]	[−25%, +0%]
	DWT	+1%	−1%	[−8%, +28%]	[−30%, +0%]
	DWT2	−11%	+11%	[−42%, +31%]	[−40%, +0%]

Table 10. WD accuracy under various contexts.

Group	Algorithm	Orientation	Personalization	Placement (Hand, Foot, etc.)	Window Size (1.5 s to 6 s)	Sampling Rate (5 Hz to 200 Hz)
Heuristic Method	THR	−0.3%	+9%	[−3%, +8%]	[−2%, +5%]	[−1%, +1%]
Signal Processing	STFT	−8%	+13%	[−15%, +15%]	[−16%, +1%]	[−13%, +1%]
	DWT	+1%	+10%	[−8%, +9%]	[−5%, +4%]	[−12%, +1%]
Machine Learning	k-NN	+10%	+15%	[−1%, +10%]	[−6%, +7%]	[−12%, +3%]
	SVM	+8%	+11%	[−8%, +7%]	[−2%, +2%]	[−12%, +2%]

5.4. Design Rules

We summarize some guidelines to design the WD and SC algorithms regarding the experiments in this research:

- Although machine learning methods perform best overall, STFT could achieve an acceptable level of accuracy when we detect walk and walk-like activity from the other two activities; 20 Hz is the transition point of the sampling rate.
- Among all the contexts, personal info is most contributive where the model is trained on a specific person.
- Complex does not mean accurate: in SC, STFT and FSM perform better in most test cases; PTM is trivial; and DWT overall performs less productive comparatively except at Hand.
- If sensors are mounted on foot, then the noise is minimal, and the result is most reliable.
- Although the steps are accurate, they may suffer from miss counting and false positives; please see the strict definition of SC.

6. Conclusions

This paper introduces context factors to evaluate walk detection and step counting algorithms through a series of experiments. Additionally, to the best of our knowledge, the method that uses ROC to evaluate the step counting is new and more comprehensive.

Table 10 shows the context impacts on WD algorithms. We find that different context factors have different effects on the algorithm performance. Amongst all algorithms, heuristic methods (THR) are the most robust to various context changes, while signal processing methods are most sensitive to changes in placement, window size and sampling rate. Machine learning methods have the best performance when a predefined context is given compared to the baseline performance and could be further improved if more contexts are provided.

Table 9 shows the context impacts on SC algorithms. The contribution of orientation is not obvious, while the contribution of personalization could remarkably enhance the overall accuracy except for the DWT algorithm. Besides, all SC algorithms are sensitive to placements, and each placement has its own best algorithm. Finally, when the sampling rate is larger than 20 Hz, the performance of all algorithms remains robust.

This paper seeks to establish a connection between activity recognition and context awareness. By presenting a quantitative comparison of algorithm performance under context impacts, this paper gives valuable guidance in designing algorithms for walk detection and step counting.

References

1. Kunze, K. Compensating for on-Body Placement Effects in Activity Recognition. Ph.D. Thesis, University Library Passau, Passau, Germany, 2011.
2. Kunze, K.; Lukowicz, P. Sensor Placement Variations in Wearable Activity Recognition. *IEEE Pervasive Comput.* **2014**, *13*, 32–41. [CrossRef]
3. Martín, H.; Bernardos, A.M.; Iglesias, J.; Casar, J.R. Activity Logging Using Lightweight Classification Techniques in Mobile Devices. *Pers. Ubiquitous Comput.* **2013**, *17*, 675–695. [CrossRef]
4. Brajdic, A.; Harle, R. Walk Detection and Step Counting on Unconstrained Smartphones. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; ACM: New York, NY, USA, 2013; pp. 225–234. [CrossRef]
5. Bulling, A.; Blanke, U.; Schiele, B. A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. *ACM Comput. Surv.* **2014**, *46*, 33. [CrossRef]
6. Zhang, M.; Sawchuk, A.A. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012.
7. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition Using Smartphones. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 24–26 April 2013.
8. Bruno, B.; Mastrogiovanni, F.; Sgorbissa, A.; Vernazza, T.; Zaccaria, R. Analysis of human behavior recognition algorithms based on acceleration data. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 1602–1607.
9. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity Recognition Using Cell Phone Accelerometers. *ACM SigKDD Explor. Newsl.* **2011**, *12*, 74–82. [CrossRef]
10. Tapia, E.M.; Intille, S.S.; Lopez, L.; Larson, K. The Design of a Portable Kit of Wireless Sensors for Naturalistic Data Collection. In Proceedings of the International Conference on Pervasive Computing, Dublin, Ireland, 7–10 May 2006.
11. Collecting complex activity datasets in highly rich networked sensor environments. In Proceedings of the 2010 Seventh International Conference on Networked Sensing Systems (INSS), Kassel, Germany, 15–18 June 2010.
12. Zheng, Y.; Shen, G.; Li, L.; Zhao, C.; Li, M.; Zhao, F. Travi-navi: Self-deployable indoor navigation system. *IEEE/ACM Trans. Netw.* **2017**, *25*, 2655–2669. [CrossRef]
13. Wannenburg, J.; Malekian, R. Physical activity recognition from smartphone accelerometer data for user context awareness sensing. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 3142–3149. [CrossRef]
14. Delgado-Gonzalo, R.; Hubbard, J.; Renevey, P.; Lemkaddem, A.; Vellinga, Q.; Ashby, D.; Willardson, J.; Bertschi, M. Real-time gait analysis with accelerometer-based smart shoes. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; p. 148.
15. de la Concepción, M.Á.Á.; Morillo, L.M.S.; García, J.A.Á.; González-Abril, L. Mobile activity recognition and fall detection system for elderly people using Ameva algorithm. *Pervasive Mob. Comput.* **2017**, *34*, 3–13. [CrossRef]
16. Cheng, J.; Yang, L.; Li, Y.; Zhang, W. Seamless outdoor/indoor navigation with WIFI/GPS aided low cost Inertial Navigation System. *Phys. Commun.* **2014**, *13*, 31–43. [CrossRef]
17. Olguın, D.O.; Pentland, A.S. Human activity recognition: Accuracy across common locations for wearable sensors. In Proceedings of the 2006 10th IEEE International Symposium on Wearable Computers, Montreux, Switzerland, 11–14 October 2006.
18. Lester, J.; Choudhury, T.; Borriello, G. A practical approach to recognizing physical activities. In Proceedings of the Pervasive Computing, Dublin, Ireland, 7–10 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–16.
19. Pärkkä, J.; Cluitmans, L.; Ermes, M. Personalization Algorithm for Real-Time Activity Recognition Using PDA, Wireless Motion Bands, and Binary Decision Tree. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 1211–1215. [CrossRef] [PubMed]

20. Cleland, I.; Kikhia, B.; Nugent, C.; Boytsov, A.; Hallberg, J.; Synnes, K.; McClean, S.; Finlay, D. Optimal placement of accelerometers for the detection of everyday activities. *Sensors* **2013**, *13*, 9183–9200. [CrossRef] [PubMed]

21. Sun, L.; Zhang, D.; Li, B.; Guo, B.; Li, S. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In Proceedings of the International Conference on Ubiquitous Intelligence and Computing, Xi'an, China, 26–29 October 2010; Springer: Berlin/Heidelberg, Germany; pp. 548–562.

22. Weiss, G.M.; Lockhart, J.W. The impact of personalization on smartphone-based activity recognition. In Proceedings of the AAAI Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.

23. Longstaff, B.; Reddy, S.; Estrin, D. Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In Proceedings of the 2010 4th International Conference on Pervasive Computing Technologies for Healthcare, Munich, Germany, 22–25 March 2010; pp. 1–7. [CrossRef]

24. Hong, J.H.; Ramos, J.; Dey, A.K. Toward Personalized Activity Recognition Systems with a Semipopulation Approach. *IEEE Trans. Hum.-Mach. Syst.* **2016**, *46*, 101–112. [CrossRef]

25. Harasimowicz, A.; Dziubich, T.; Brzeski, A. Accelerometer-based Human Activity Recognition and the Impact of the Sample Size. In Proceedings of the 13th International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED '14) and 15th International Conference on Fuzzy Systems (FS '14) and 15th International Conference on Neural Networks (NN '14), Gdansk, Poland, 15–17 May 2014; pp. 130–135.

26. Wang, L.; Gu, T.; Tao, X.; Lu, J. A hierarchical approach to real-time activity recognition in body sensor networks. *Pervasive Mob. Comput.* **2012**, *8*, 115–130. [CrossRef]

27. Ying, H.; Silex, C.; Schnitzer, A.; Leonhardt, S.; Schiek, M. Automatic Step Detection in the Accelerometer Signal. In Proceedings of the 4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007), Aachen, Germany, 26–28 March 2007; Leonhardt, S., Falck, D.I.T., Mähönen, P.D.P., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 80–85.

28. Lester, J.; Hartung, C.; Pina, L.; Libby, R.; Borriello, G.; Duncan, G. Validated caloric expenditure estimation using a single body-worn sensor. In Proceedings of the 11th International Conference on Ubiquitous Computing, Orlando, FL, USA, 30 September–3 October 2009; pp. 225–234.

29. Sagha, H.; Digumarti, S.T.; Millan, J.d.R.; Chavarriaga, R.; Calatroni, A.; Roggen, D.; Troster, G. Benchmarking classification techniques using the Opportunity human activity dataset. In Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, USA, 9–12 October 2011; pp. 36–40.

30. Mortazavi, B.J.; Pourhomayoun, M.; Alsheikh, G.; Alshurafa, N.; Lee, S.I.; Sarrafzadeh, M. Determining the Single Best Axis for Exercise Repetition Recognition and Counting on SmartWatches. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks, Zurich, Switzerland, 16–19 June 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 33–38. [CrossRef]

31. Ravi, N.; Dandekar, N.; Mysore, P.; Littman, M.L. Activity recognition from accelerometer data. In Proceedings of the AAAI-05: Twentieth National Conference on Artificial Intelligence, Pittsburgh, PA, USA, 9–13 July 2005; Volume 5, pp. 1541–1546.

32. Ibrahim, R.K.; Ambikairajah, E.; Celler, B.G.; Lovell, N.H. Time-frequency based features for classification of walking patterns. In Proceedings of the 2007 15th International Conference on Digital Signal Processing, Cardiff, UK, 1–4 July 2007; pp. 187–190.

33. Casale, P.; Altini, M.; Amft, O. Transfer Learning in Body Sensor Networks using Ensembles of Randomised Trees. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks, Zurich, Switzerland, 16–19 June 2014; pp. 39–44.

34. Santhiranayagam, B.K.; Lai, D.T.; Jiang, C.; Shilton, A.; Begg, R. Automatic detection of different walking conditions using inertial sensor data. In Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, QLD, Australia, 10–15 June 2012; pp. 1–6.

35. Alzantot, M.; Youssef, M. UPTIME: Ubiquitous pedestrian tracking using mobile phones. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 1–4 April 2012; pp. 3204–3209. [CrossRef]

36. Nickel, C.; Brandt, H.; Busch, C. Benchmarking the performance of SVMs and HMMs for accelerometer-based biometric gait recognition. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (2011), Bilbao, Spain, 14–17 December 2011; pp. 281–286.

37. Parkka, J.; Ermes, M.; Korpipaa, P.; Mantyjarvi, J.; Peltola, J.; Korhonen, I. Activity classification using realistic data from wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 119–128. [CrossRef] [PubMed]

38. Ahmadi, A.; Mitchell, E.; Destelle, F.; Gowing, M.; OConnor, N.E.; Richter, C.; Moran, K. Automatic activity classification and movement assessment during a sports training session using wearable inertial sensors. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks, Zurich, Switzerland, 16–19 June 2014; pp. 98–103.

39. Wang, J.H.; Ding, J.J.; Chen, Y.; Chen, H.H. Real time accelerometer-based gait recognition using adaptive windowed wavelet transforms. In Proceedings of the 2012 IEEE Asia Pacific Conference on Circuits and Systems, Kaohsiung, Taiwan, 2–5 December 2012; pp. 591–594. [CrossRef]

40. Plötz, T.; Hammerla, N.Y.; Olivier, P. Feature learning for activity recognition in ubiquitous computing. In Proceedings of the IJCAI Proceedings-International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; Volume 22, p. 1729.

41. Bhattacharya, S.; Nurmi, P.; Hammerla, N.; Plötz, T. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive Mob. Comput.* **2014**, *15*, 242–262. [CrossRef]

42. Zhang, M.; Xu, W.; Sawchuk, A.A.; Sarrafzadeh, M. Sparse representation for motion primitive-based human activity modeling and recognition using wearable sensors. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 1807–1810.

43. He, Z.; Liu, Z.; Jin, L.; Zhen, L.X.; Huang, J.C. Weightlessness feature—A novel feature for single tri-axial accelerometer based activity recognition. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.

44. Kim, J.W.; Jang, H.J.; Hwang, D.H.; Park, C. A Step, Stride and Heading Determination for the Pedestrian Navigation System. *Positioning* **2004**, *3*, 273–279. [CrossRef]

45. Randell, C.; Djiallis, C.; Muller, H. Personal position measurement using dead reckoning. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle, UK, 18–22 June 2012.

46. Bylemans, I.; Weyn, M.; Klepal, M. Mobile Phone-Based Displacement Estimation for Opportunistic Localisation Systems. In Proceedings of the Third International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Sliema, Malta, 11–16 October 2009; pp. 113–118. [CrossRef]

47. Ailisto, H.J.; Lindholm, M.; Mantyjarvi, J.; Vildjiounaite, E.; Makela, S.M. Identifying people from gait pattern with accelerometers. In Proceedings of the Defense and Security, International Society for Optics and Photonics, Orlando, FL, USA, 28 March–1 April 2005; pp. 7–14.

48. Beauregard, S. A helmet-mounted pedestrian dead reckoning system. In Proceedings of the 3rd International Forum on Applied Wearable Computing (IFAWC 2006), Bremen, Germany, 15–16 March 2006; pp. 1–11.

49. Goyal, P.; Ribeiro, V.J.; Saran, H.; Kumar, A. Strap-down pedestrian dead-reckoning system. In Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation, Guimaraes, Portugal, 21–23 September 2011; pp. 1–7.

50. Barralon, P.; Vuillerme, N.; Noury, N. Walk detection with a kinematic sensor: Frequency and wavelet comparison. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 1711–1714.

51. Libby, R. A Simple Method for Reliable Footstep Detection in Embedded Sensor Platforms. Available online: http://ozeo.org/lib/exe/fetch.php?media=bodytrack:libby_peak_detection.pdf (accessed on 15 October 2018).

52. Nyan, M.; Tay, F.; Seah, K.; Sitoh, Y. Classification of gait patterns in the time-frequency domain. *J. Biomech.* **2006**, *39*, 2647–2656. [CrossRef] [PubMed]

53. Rai, A.; Chintalapudi, K.K.; Padmanabhan, V.N.; Sen, R. Zee: Zero-effort crowdsourcing for indoor localization. In Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, Istanbul, Turkey, 22–26 August 2012; pp. 293–304.

54. Willemsen, A.; Bloemhof, F.; Boom, H. Automatic stance-swing phase detection from accelerometer data for peroneal nerve stimulation. *IEEE Trans. Biomed. Eng.* **1990**, *37*, 1201–1208. [CrossRef] [PubMed]

55. Rong, L.; Zhiguo, D.; Jianzhong, Z.; Ming, L. Identification of individual walking patterns using gait acceleration. In Proceedings of the 2007 1st International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 6–8 July 2007; pp. 543–546.

56. Reiss, A.; Stricker, D. Personalized mobile physical activity recognition. In Proceedings of the 17th Annual International Symposium on International Symposium on Wearable Computers, Zurich, Switzerland, 8–12 September 2013; pp. 25–28.

57. Reiss, A.; Stricker, D.; Hendeby, G. Confidence-based multiclass AdaBoost for physical activity monitoring. In Proceedings of the 17th Annual International Symposium on International Symposium on Wearable Computers, Zurich, Switzerland, 8–12 September 2013; pp. 13–20.

58. Pfau, T.; Ferrari, M.; Parsons, K.; Wilson, A. A hidden Markov model-based stride segmentation technique applied to equine inertial sensor trunk movement data. *J. Biomech.* **2008**, *41*, 216–220. [CrossRef] [PubMed]

59. Mannini, A.; Sabatini, A.M. Accelerometry-based classification of human activities using markov modeling. *Comput. Intell. Nanosci.* **2011**, *2011*, 4. [CrossRef] [PubMed]

60. Mannini, A.; Sabatini, A.M. A hidden Markov model-based technique for gait segmentation using a foot-mounted gyroscope. In Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 4369–4373.

61. Chen, S.; Lach, J.; Amft, O.; Altini, M.; Penders, J. Unsupervised activity clustering to estimate energy expenditure with a single body sensor. In Proceedings of the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, MA, USA, 6–9 May 2013; pp. 1–6.

62. Choe, B.; Min, J.K.; Cho, S.B. Online gesture recognition for user interface on accelerometer built-in mobile phones. In Proceedings of the17th International Conference, ICONIP 2010, Sydney, Australia, 22–25 November 2010; pp. 650–657.

63. Cheng, H.T.; Griss, M.; Davis, P.; Li, J.; You, D. Towards zero-shot learning for human activity recognition using semantic attribute sequence model. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; pp. 355–358.

64. Rebetez, J.; Satizábal, H.F.; Perez-Uribe, A. Reducing User Intervention in Incremental Activityrecognition for Assistive Technologies. In Proceedings of the 2013 International Symposium on Wearable Computers, Zurich, Switzerland, 8–12 September 2013; ACM: New York, NY, USA, 2013; pp. 29–32. [CrossRef]

65. Mei, S. Probability Weighted Ensemble Transfer Learning for Predicting Interactions between HIV-1 and Human Proteins. *PLoS ONE* **2013**, *8*, e79606. [CrossRef] [PubMed]

66. Dai, W.; Yang, Q.; Xue, G.R.; Yu, Y. Boosting for transfer learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 193–200.

67. Calatroni, A.; Roggen, D.; Tröster, G. Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion. In Proceedings of the Eighth International Conference on Networked Sensing Systems (INSS 2011), Penghu, Taiwan, 13–15 June 2011; Volume 6.

68. Blanke, U.; Schiele, B. Remember and transfer what you have learned-recognizing composite activities based on activity spotting. In Proceedings of the 2010 International Symposium on Wearable Computers (ISWC), Seoul, Korea, 10–13 October 2010; pp. 1–8.

69. van Kasteren, T.; Englebienne, G.; Kröse, B.J. Transferring knowledge of activity recognition across sensor networks. In *Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 283–300.

The Design and Application of Simplified Insole-Based Prototypes with Plantar Pressure Measurement for Fast Screening of Flat-Foot

Wei-Chun Hsu [1,2,3,*], Tommy Sugiarto [1,2,4], Jun-Wen Chen [1] and Yi-Jia Lin [1]

1 Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan; d10622817@mail.ntust.edu.tw (T.S.); m10523102@mail.ntust.edu.tw (J.-W.C.); jiajia527@gmail.com (Y.-J.L.)
2 Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei City 10607, Taiwan
3 Department of Biomedical Engineering, National Defense Medical Center, Taipei 11490, Taiwan
4 Division of Embedded System and SoC Technology, System Integration and Application Department, Information and Communication Research Laboratory, Industrial Technology Research Institute, Hsinchu 31057, Taiwan
* Correspondence: wchsu@mail.ntust.edu.tw; Tel.: +886-2730-3741

Abstract: This study aimed to find the correlation between conventional Arch Index (AI) measurements and our prototype of a simplified insole-based plantar pressure measurement system and to find out the effective plantar pressure sensor position. Twenty-one subjects participated in this study, which was divided into two groups: 10 subjects with flatfoot and 11 subjects with normal foot. Five force sensitive resistance sensors were used on this prototype using Arduino as the data acquisition device. Two types of trials, namely static and dynamic, were conducted to validate our system against the ink-type AI measurement as a golden standard. The results showed that in the static trial, there was a high linear correlation with the medial arch sensor configuration, while in the dynamic trial, there was a high linear correlation in the medial arch sensor configuration and sensor 5 configuration. This study showed that both static and dynamic tests using the self-developed device could effectively determine most of the flatfoot subjects and suggests that in the future, it can be applied in clinical applications because of its advantages when compared to the expensive-high tech graphic input board and conventional tools, like ink-type based measurements.

Keywords: plantar pressure; flat foot; insoles; force sensors; arch index

1. Introduction

The arch of the foot helps in absorbing and transferring energy when walking or running. Thus, an abnormal arch, such as flatfoot, often leads to an abnormal gait and running pattern. Flatfoot can be divided into flexible flatfoot and rigid flatfoot, with the former, also called functional flatfoot, meaning that the arch is flat when the foot is weight bearing, or that the arch is evident when the toes of the foot point outwards while the subjects stand in non-weight bearing or through the toe-raising of the Jack Test proposed in 1953 [1]. While the latter, also called structural flatfoot, refers to when the arch is not present, whether in weight bearing status or not, and the subtalar joint movement is poor. Ninety-five percent of patients with flatfoot are flexible flatfoot [2]. Patients with rigid flatfoot often suffer from pain or potentially more severe pathological problems, such as tarsal coalition or neuromuscular flatfoot [3], etc. Rigid flatfoot mostly requires surgery rather than being treated by traditional treatments if patients are in pain [2].

Currently, the methods for judging flatfoot have been mainly focused on the medial longitude arch (MLA), and common methods include the traditional arch height, the navicular drop test [4,5], the footprint index classification [6–9], and X-ray [10]. The arch height can be quantified by the difference between the highest point of the soft tissue on the edge of the medial arch and the ground [11]. The navicular drop test (NDT) was first described as a means of representing the height change of the navicular bone under-weight bearing and non-weight bearing [5], with a moderate to good intra-reproducibility of the measurement [12]. The level of navicular drop refers to the height difference of the navicular tubercle in the standing and sitting position, and has been reported to be significantly correlated with the degree of the pronation of the foot of a runner, and has been used for judging flatfoot [4].

The footprint index classification is commonly used in research and the traditional method is mostly conducted with an ink-type instrument, which has advantages over other clinical tests, with its high reproducibility [13,14]. More recently, the plantar pressure system or the plantar scanning system has been used to measure, calculate, and determine the footprint. Although the footprints are measured in different ways, the method of judging flatfoot is the same. There are different calculations in terms of footprint classification, where the common arch index (AI) is obtained while the subjects stand evenly on the ink board one foot at a time [6]. The footprints are divided into three equal areas based on the foot length of the footprints to calculate the AI of the subjects and also quantify the severity of flatfoot [15]. The AI is highly reliable and correlates with X-ray measurements, so has been used for determining flatfoot in research [13]. The judgment for the type of arch relies on X-rays in hospital, which is reliable, but time-consuming and expensive.

Plantar pressure, also known as the pressure between the foot and the support surface, can be calculated by the force of the vertical contact divided by the contact area. Obtaining the plantar pressure information is helpful for assessing clinical problems. For example, the peak plantar pressure of the forefoot and hindfoot as well as the value of the peak forefoot pressure divided by the peak hindfoot pressure of severe diabetic neuropathic patients were significantly higher than that of patients with mild ulceration or with no history of ulceration. This was determined to be the possible reason as to why diabetic patients easily face the risk of foot ulceration. Therefore, plantar pressure can be regarded as a parameter to prevent foot ulceration for diabetic patients [16].

The average pressure has also been studied under 10 anatomical areas, including the medial and lateral heel area, the medial and lateral mid-foot area, the heads of the first, second, and fifth metatarsal area, the great toe, the head of the second metatarsal area, and the lateral metatarsal head area [15]. However, the placement of different insoles could effectively affect these pressure indices [17]. Thus, recent studies have addressed how to simplify the number of sensors in the plantar pressure insole or to place sensors at relatively important positions based on applications to reduce the development cost as well as provide relatively accurate data for launching products in the market to consumers [18]. Reduction of the spatial and temporal image resolution by placing pressure sensors at only seven locations, namely the heel, lateral midfoot, lateral forefoot, great toe, head of the first metatarsal, center midfoot, and center forefoot has been used for measuring and evaluating plantar pressure during activities of daily living for further gait analysis and interpretation of pathological foot anatomy.

A cluster type of sensor placement, being three on the metatarsal and three on the heels with a total of only six points, was developed, with the aim of further comparing the results between those obtained from state-of-the-art equipment with those obtained from sensors on the insole [19]. By placing sensors on the insole, these cost-effective and efficient wearable devices allow the monitoring and analysis of the gait anywhere from large clinics to an individual's home, and can be applied to produce graphics similar to those measured from the force-plate [20,21], where fourteen sensors of the Force Sensitive Resistance (FSR) Model 402 (Interlink Electronics, Camarillo, CA, USA) were placed on the forefoot, midfoot, and hindfoot to measure plantar pressure. In 2016, Liang et al. also developed an insole-based plantar pressure measurement with a fiber-optic sensing system where six fiber-optic sensors were used and embedded in silicone rubber. The results showed that the system could successfully identify

the four different foot types and their developed-system could reach a Pearson correlation coefficient of 0.671 compared to the *i-Step P1000* plantar pressure plate [22].

It has been suggested that the traditional visual judgment of flatfoot was too subjective, and that flatfoot should be detected by objective and accurate optical or electronic systems [23]. This study suggested how to design the insole-based plantar pressure instrument for flatfoot. However, it lacked practical clinical verification. However, there are few insole-based plantar pressure measurement systems for flatfoot, therefore, methods of measuring flatfoot with a simplified framework of software, hardware, and design to develop a low-cost wearable insole-based instrument that allows for the measurement of plantar pressure to quickly and successfully screen flatfoot and normal foot during both static and dynamic conditions are of clinical needs.

Thus, the aims of this study were first to perform the correlation between traditional AI and the value of each point of the self-made insole-based sensor to find the effective plantar pressure sensor position and to compare whether there were significant differences between the sensors at specific positions of flatfoot and normal foot from the data of 21 young subjects.

2. Materials and Methods

2.1. Subject

Twenty-one healthy subjects with shoes sized between US9/Euro 42 and US10/Euro 43–44 participated in this study. Ten subjects with flatfoot and 11 with normal foot were classified by ink-type AI. Healthy subjects had no history of affecting postural stability, or gait in the past 12 months, and/or cognitive impairment. This experiment was approved by local research ethics. Subjects who injured their lower extremity six months ago before participation and still suffered from pain, who had surgery on their hip joints, knee joints, and ankle joints of the lower extremity, and who could not walk or stand for the pain in the lower extremity or back were excluded from the study. The subject's characteristics, including their Body Mass Index (BMI) and ratio of foot length and navicular-to-toe length, are shown in Table 1.

Table 1. Subject characteristics of 21 young subjects who participated in this study.

Subject Information	Normal Foot (N = 11)		Flatfoot (N = 10)	
	Mean	±Std	Mean	±Std
Age (year)	23.4	1.0	23.1	2.2
Height (cm)	176.6	3.0	173.1	5.6
Weight (kg)	70.5	6.9	72.2	7.8
BMI (kg/m^2)	22.6	2.1	24.0	2.5
Arch Index	0.251	0.016	0.356	0.027
Shoe Size (US Size)	9.6	0.4	9.5	0.43
Ratio of foot length and navicular-to-toe length	0.63	0.01	0.64	0.01

2.2. Equipment

2.2.1. Ink-Type Footprint Printer

According to the method proposed by Cavanagh in 1987, the subjects stood statically with even force on the ink-type footprint printer, then the subjects' footprints were obtained on the paper by the ink [6].

The measured footprint was drawn from the second metatarsal bone to the center of the heel, which is called the foot axis. Next, the foot axis was divided vertically into three equal parts, namely the forefoot (A), the midfoot (B), and the hindfoot (C) (Figure 1). The AI was calculated by the midfoot area (B) being divided by the full foot area (A + B + C). If the AI was less than or equal to 0.21, it was regarded as high arch; if the AI was between 0.21 and 0.26, it was regarded as normal foot; if the AI was greater than 0.26, it was regarded as flatfoot [6].
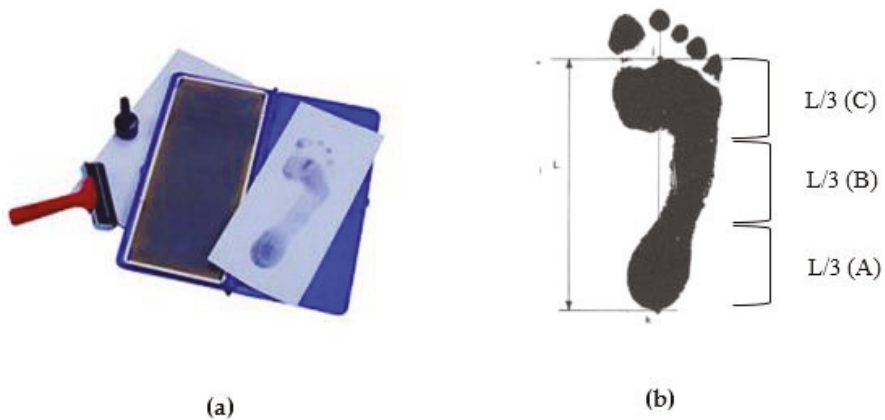
Figure 1. Ink-type footprint printer (**a**,**b**) the method of arch index calculation [6].

2.2.2. Self-Made Simplified Instrument

Combining the clinical and electronics circuit design ability, we designed a simplified self-made instrument to effectively measure flatfoot and normal foot. The hardware system design diagram is shown in Figure 2. An Arduino Micro Pro was used combined with the Force FSR placed on insoles to measure the foot plantar pressure and transmitted the data to a personal computer by Bluetooth connection.

2.2.3. Hardware System Design

The entire hardware system was designed with an Arduino Mirco Pro and the hardware system diagram is shown in Figure 2. The Arduino Micro Pro used in this study was based on ATmega32u4 as its processor. The analog input channel connected five FSR to sense the external pressure, and the data from the FSR were transmitted to a PC with the Python environment via Bluetooth connection (HC-05 Bluetooth module). This system used a rechargeable lithium battery as the power source, thus the TP-4056 module was used as a battery charger and protection module.
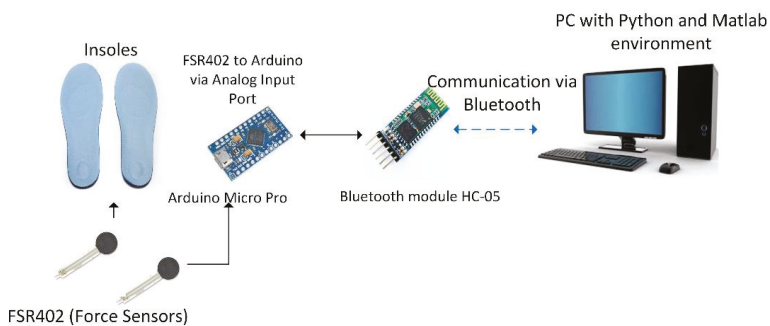
Figure 2. The hardware system diagram of the FSR-based insoles for measuring the flatfoot and normal foot.

The FSR sensor used in this study was an Interlink-402 that could measure the force values up to 100 g–10 kg. The FSR is composed of a flexible substrate with a printed semi-conductor, flexible substrate with interdigitating electrode, and spacer adhesive. FSR is quite cheap and easily used with a small size (the diameter of the sensing area is 12.7 mm) and thin thickness (thickness of 0.45). FSR is not as accurate as the load cell or the strain gauge, however, it can reduce measurement errors or

use relative comparisons through individual calibration [24]. In this study, the calibration was done before the FRS were put on the insoles, with the calibration procedure independently applied to each sensor. Data processing for this calibration procedure was done with Matlab 2017b (MathWorks Inc., Natick, MA, USA). The Interlink402-FSR is shown in Figure 3.
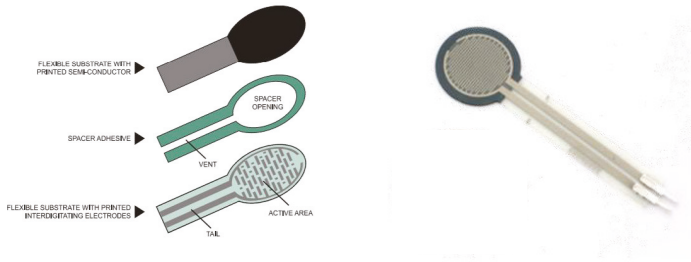
Figure 3. FSR-Interlink402 sensors (**right**) and its detailed configuration (**left**).

Special attention should be paid to the so-called force repeatability in the parameters of force sensing components as force repeatability indicates that the resistance values may change between $\pm15\%\sim\pm25\%$ with the same force in different sensors. As a result, each sensing component should be individually calibrated before use. The nickel-clad copper precision weight of the electronic scale weight was used to calibrate each FSR. Referring to the static calibration described by Flórez in 2010, the precision weight is regarded as a fixed force. In the calibration process, weights of different amounts were placed on the platform, with the force evenly applied on the FSR, before finally, the resistances were recorded under different forces by the algorithm [24].

2.2.4. Sensor Placement Design

As discussed in the literature review, flatfoot is judged mainly by whether the MLA (medial longitude arch) drops. From the perspective of plantar anatomy, the drop of the navicular bone is the main cause, thus the regular position of the navicular bone should be referred to before designing the sensor placement. According to the statistics of the experiment, the position of the navicular bone falls between 60% and 65% of the whole foot, and the length of an adult's navicular bone ranges between 1–2 cm. Therefore, we decided to set 60% as the position of the navicular bone where the sensors were placed with the distance of 1 cm on the front and back. The sensors were placed in an array and numbered from 1 to 5 from the medial to the lateral, respectively. The illustration of the sensor placement and numbering is shown in Figure 4.
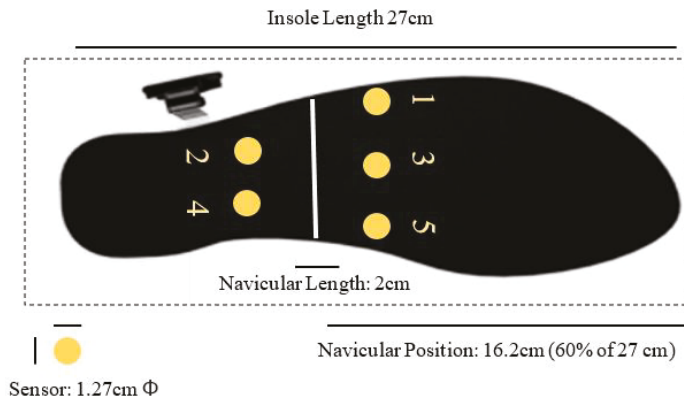
Figure 4. Sensor placement and numbering on the insoles.

2.2.5. Experimental Procedure

- Measurement of the ratio of the foot length and navicular length

In this experiment, the foot length was measured with a foot gauge. The subjects were required to stand evenly, placing one foot on the foot gauge. The heel of the foot was measured to the front position of the foot (great toe or index toe). The protruding position of the navicular bone was found and marked by palpation, then the length from the navicular bone to the front position of the foot was measured by a ruler. After recording, the value was divided by the foot length to obtain the position ratio of the corresponding navicular bone (Figure 5).
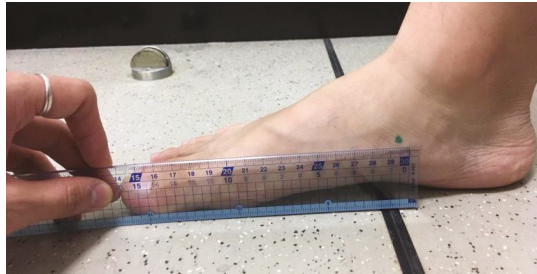
Figure 5. Illustration of the measurement procedure of the navicular length.

- Ink-type footprint measurement

As an index of determining the foot type in the current study, the measurement and calculation of the ink-type footprint refers to that mentioned in the paper by Cavanagh in 1987 [6,15], where the subjects stand statically with one foot on the ink-type footprint printer that prints the footprint of the subjects. The measurement needs to be repeated if swaying or over-imprint occurs (such as standing with one foot) (Figure 6).
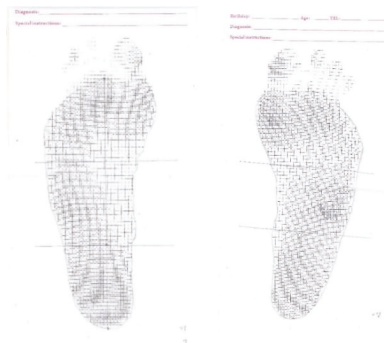
Figure 6. Typical example of ink-type footprint diagram.

2.2.6. Static Standing Insole-Type Plantar Pressure Measurement

Based on the methods proposed by Cavanagh in 1987, the movements were designed to test whether the self-made insole-type plantar was able to determine the subject's foot type. The subjects stood evenly on the insole-type plantar pressure instrument, and the designed insole-type plantar pressure instruments were placed in a normal shoe that served as the test shoe in the current study. Before using this normal shoe, all the additional support for the arch had already been removed to ensure that there was no other arch support, which may bias the result. On this statically standing

position, the subjects had to look straight ahead with their hands at their sides for five seconds with the sample rate of 100 Hz (as shown in Figure 7).

Figure 7. Static standing plantar pressure measurement position for the subject, with the insole-plantar pressure device inserted in the shoes.

2.2.7. Dynamic Insole-Type Plantar Pressure Measurement

The subjects were asked to walk in a dynamic trial while wearing the self-made insole-type plantar pressure device. The stance phase of the gaits was calculated the period where the heel makes contact with the ground to when the soles are off the ground. One sensor was placed on the heel to define the heel-strike time while the other one was placed on the forefoot to detect the toe-off time. The stance phase during the dynamic trials was started at the heel-strike and ended at the toe-off.

Each subject needed to complete three successful trials with their preferred speed. The subject's preferred speed was chosen to avoid an un-natural walking style of the subject if they walked at the specified chosen speed. The trial had to be recollected if there was any instability or if the subject did not look straight ahead.

2.3. Data Analysis

The force data from the first to fifth sensors in the static and dynamic trial were calculated, respectively. In terms of static standing, the average of the force of static standing for five seconds was divided by the body weight. During analysis, this study calculated several indices. First, the average normalized force of each sensor for five seconds. Second, this study viewed the third sensor as the center column of the arch, which divided the arch into the medial arch and lateral arch and the third sensor also contributed both the medial side and lateral, the medial arch (the sum of the average force of the fifth sensor, the fourth sensor, and the third sensor was divided by two), and the lateral arch (the sum of the average force of the first sensor, the second sensor, and the third was divided by two). Third, the first three sensors (the sum of the first, the third, and the fifth) and the five sensors were calculated.

In terms of the dynamic trial, three successful gait trials were collected, and the maximum value of each gait of each sensor, the maximum of the sum of the medial arch in Equation (1), the average force of the lateral arch in Equation (2), the maximum of the sum of the first three sensors, and the maximum force of five sensors were calculated.

$$Max \left(Sensor\ 5 + Sensor\ 4 + \frac{Sensor\ 3}{2} \right) \tag{1}$$

$$Max\left(Sensor\ 1 + Sensor\ 2 + \frac{Sensor\ 3}{2}\right) \tag{2}$$

All the force data calculated in this paper were normalized by their own body weight (kgf/kg) to minimize the effect of the individual difference.

2.4. Statistical Analysis

Statistical analysis was performed to obtain the Pearson r correlation coefficient by linear regression between the calculated data and the ink-type AI. A correlation coefficient falling between 0.7–0.99 was categorized as highly correlated, while 0.4–0.6 was categorized as a moderate correlation, 0.1–0.39 was categorized as a low correlation, and 0–0.1 was categorized as extremely uncorrelated.

For the comparison between groups, the Mann–Whitney U Test was used for the between-group comparison of pressure parameters between flatfoot and normal foot type groups with the significance level set at alpha = 0.05.

The reliability analysis for examining the reliability of the proposed system when several measurements on the same subject was performed with the Intraclass Correlation Coefficient (ICC) with a two-way mixed model (ICC $_{3,1}$) and Cronbach's alpha coefficient.

3. Results

The experimental results were divided into two parts. First, the static data results included the relationship between the individual sensors (1 to 5) with the AI as well as the relationship between the force sum of the first three sensors and all five sensors and the AI. Second, the dynamic data results included the relationship between the maximum value of one to five sensors sensing when the subjects walked and the AI.

3.1. Static Standing Trials

The Pearson correlation between the forces index calculated from the force sensors during the static and dynamic trial (with different sensors combination) with the ink-type AI value is shown in Table 2. The highest correlation was found on the sensors at the medial arch position (sensor 5 + 4 + 3/2) with the Pearson correlation coefficient equal to 0.715. Meanwhile, the lowest correlation was found on Sensor 5 with r = 0.0707. The comparison between normal and flatfoot in the static trial is shown in Table 3. There is a significant difference between the normal and flatfoot force value in most of the sensor placement positions.

Table 2. Pearson correlation between the force sensor standing value and the ink-type AI value in the static test.

Parameters	Mean (kgf/kg)	Standard Deviation	r	*p* Value
Sensor 1	0.00688	0.00424	0.406	0.001 **
Sensor 2	0.01081	0.00346	0.336	0.008 *
Sensor 3	0.00602	0.00397	0.678	0.000 **
Sensor 4	0.01562	0.0045	0.417	0.001 **
Sensor 5	0.00339	0.00471	0.0707	0.000 **
3 Sensors Point (1, 3, 5)	0.01629	0.01138	0.680	0.000 **
5 Sensors Point (1–5)	0.04272	0.01613	0.668	0.000 **
Medial Arch (5 + 4 + 3/2)	0.02202	0.00917	0.715	0.000 **
Lateral Arch (1 + 2 + 3/2)	0.02069	0.00868	0.487	0.000 **

** Correlation was significant at the 0.005 level (two-tailed).

Table 3. Comparison of normal and flatfoot pressure value in the static trials.

Parameters	Normal Foot		Flatfoot		p Value
	Mean	Standard Deviation	Mean	Standard Deviation	
Sensor 1	0.00462	0.00205	0.00929	0.00456	0.009 *
Sensor 2	0.00931	0.00261	0.0124	0.00347	0.029 *
Sensor 3	0.00342	0.00108	0.0088	0.00397	0.003 **
Sensor 4	0.01371	0.00469	0.01765	0.00311	0.049 *
Sensor 5	0.00024	0.00059	0.00676	0.00477	0.000 **
3 Sensors Point (1, 3, 5)	0.00827	0.00223	0.02485	0.01078	0.000 **
5 Sensors Point (1–5)	0.03129	0.00648	0.05491	0.01401	0.001 **
Medial Arch (5 + 4 + 3/2)	0.01566	0.00519	0.02881	0.00728	0.000 **
Lateral Arch (1 + 2 + 3/2)	0.01563	0.00412	0.02609	0.00886	0.006 **

* $p < 0.05$, significant difference, ** $p < 0.005$, significant difference.

3.2. Dynamic Trials

The Pearson correlation between the maximum force index of the force sensor in the dynamic test and the ink-type AI value is shown in Table 4. The highest correlation in the dynamic trial was found on Sensor 5 with r = 0.801 while the less correlated sensor position in the dynamic trial was found in the Sensor 1 position with r = 0.063. The comparison of the normal and flatfoot pressure value and Mann–Whitney U test results in the dynamic trial is shown in Table 5. There was a significant difference between the normal and flatfoot pressure value in six of the sensor configurations: Sensor 3, Sensor 4, Sensor 5, 3 Sensor point (1, 3, 5), the medial arch sensor position, and the lateral arch sensor position.

The linear regression result from the AI vs. Sensor 5 placement in the dynamic trial is shown in Figure 8a. Those results had the highest correlation with r = 0.801 in the dynamic trial. Figure 8b also showed the worst correlation, which was obtained from the AI vs. Sensor 1 placement in the dynamic trial.
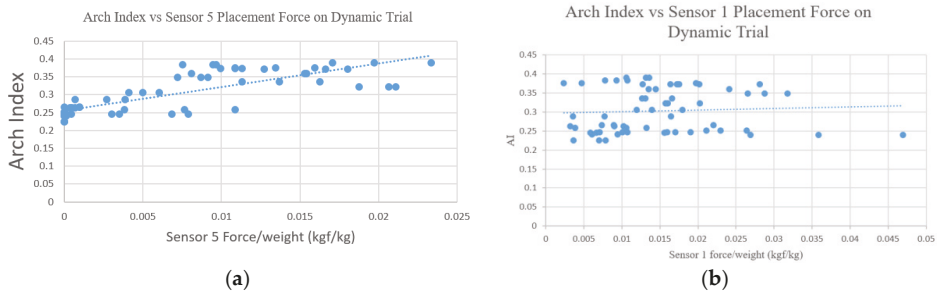
Figure 8. Arch index vs. sensor placement normalized force index in the dynamic trial. (**a**) The best correlation, which was obtained from the placement of Sensor 5; and (**b**) the worst correlation, which was obtained from the placement of Sensor 1.

Reliability analysis for the system measuring different trials on the same subject was done with an ICC two-way mixed model and Cronbach's alpha coefficient. The result showed that the proposed measurement system had high reliability for measuring different trials on the same subject with an ICC = 0.812 and Cronbach's alpha = 0.93.

Table 4. Pearson correlation between the force sensor standing value and the ink-type AI value in the dynamic test.

Parameters	Mean (kgf/kg)	Standard Deviation	r	*p* Value
Sensor 1	0.03655	0.00861	0.063	0.627
Sensor 2	0.03127	0.00543	0.343	0.006 *
Sensor 3	0.02922	0.00692	0.541	0.000 **
Sensor 4	0.04123	0.00747	0.505	0.000 **
Sensor 5	0.02017	0.00695	0.801	0.000 **
3 Sensors Point (1, 3, 5)	0.056	0.01529	0.604	0.000 **
5 Sensors Point (1–5)	0.11451	0.02282	0.587	0.000 **
Medial Arch (5 + 4 + 3/2)	0.06643	0.01319	0.784	0.000 **
Lateral Arch (1 + 2 + 3/2)	0.06069	0.01374	0.319	0.011 *

** Correlation was significant at the 0.005 level (two-tailed).

Table 5. Comparison of the normal and flatfoot pressure value in the dynamic trials.

Parameters	Normal Foot		Flatfoot		*p* Value
	Mean	Standard Deviation	Mean	Standard Deviation	
Sensor 1	0.01351	0.00859	0.01605	0.00555	0.105
Sensor 2	0.01883	0.00374	0.02223	0.00529	0.139
Sensor 3	0.01104	0.00421	0.01798	0.00633	0.007 *
Sensor 4	0.02304	0.00727	0.02906	0.00558	0.035 *
Sensor 5	0.00168	0.00249	0.01291	0.00466	0.000 **
3 Sensors Point (1, 3, 5)	0.02574	0.00967	0.04532	0.01071	0.000 **
5 Sensors Point (1–5)	0.06191	0.01307	0.08878	0.01926	0.067
Medial Arch (5 + 4 + 3/2)	0.02812	0.00758	0.04735	0.00912	0.001 **
Lateral Arch (1 + 2 + 3/2)	0.03461	0.0111	0.04401	0.00961	0.006 *

* $p < 0.05$, significant difference; group effects were analyzed using the Mann–Whitney U test.

4. Discussion

4.1. Placements of the Sensors

To the best of our knowledge, the majority of previous papers have suggested the placement of the force sensitive resistance (FSR) for some specific reference sites of the foot itself, such as the lateral arch and forefoot [18,21]. The current study proposed a more general guideline for the placement of the FSR relative to a given anatomy site, like the commonly targeted medial arch taken in the current study, by describing the placement site in terms of percentage of the insole along the length of the insole itself, which copes with the size of the participant's foot. The established FSR placement indication on the insole could also help provide guidelines regarding the placements of the FSR from a single one to numerous ones according to the needs for clinical application during the current and future development of the prototype. Based on this repeatability consideration, the author could then perform the human testing that aimed to first perform the correlation between traditional AI and the value of each point, which is a new way to define the placement of the sensor in this study to find out the position of the effective plantar pressure sensor. In addition, unlike the aims of those previous research and development studies reported in the literature, the current study aimed to compare whether there were significantly different values obtained from sensors at specific positions between the flatfoot and normal foot using the data of 21 young subjects.

4.2. Static Trials

The results of the static standing trial showed that there was a high linear correlation between the medial arch sensor (5 + 4 + 3/2) with the AI obtained from the static ink ($r > 0.7$). Meanwhile, five

sensor configurations (sensor 1, sensor 3, 3-sensors point (1, 3, 5), 5-sensors point (1–5), and the lateral arch sensor (1 + 2 + 3/2) showed a moderate correlation with 0.4 < r < 0.6.

The ink-type AI has been reported to also be correlated with the navicular height [13,25], the medial arch angle analyzed by X-ray [14,26,27], and the pressure of the hindfoot [28–30] and midfoot [31,32] measured during gait. The results of the current study indicated that the force value measured by the insole-type sensor alone was also associated with the AI measured by the traditional ink one, which may be introduced as another method to distinguish between a flatfoot and normal foot. Although significant differences between normal foot and flatfoot were also found ($p < 0.05$), the current method should be applied with caution for individuals whose AI value falls on the transition area (from normal to flat foot).

The results showed that the force of the center midfoot of the flatfoot was obviously higher than the normal foot, which was consistent with the results of the current study. There were obvious differences between the medial arch of the flatfoot and normal foot. Thus, in the static experiment, it was found that the sensors on the arch could be placed in the position of 3 or 5 for the instruments to judge between flatfoot and common foot in the static design. There was a high linear correlation with the AI from the conventional ink-type measurement, and the comparison between the groups showed that there was a significant difference between the sensing measurements of the position of 3 or 5 when judging between normal foot and flatfoot ($p < 0.05$).

4.3. Dynamic Trials

The results of the dynamic walking trials pointed out that the medial sensor and Sensor 5 and the medial arch sensor position had high linear correlations (r > 0.7) with the ink-type AI, and different calculation methods were designed mainly with MLA. The linear correlation between Sensor 4 and MLA was merely moderate, speculated to be the same as the results observed when static.

For the comparison of data measured in the walking trials between the flatfoot and normal foot, the calculated force sum of five sensors, three forefoot arch sensors or forefoot arch and Sensor 3, Sensor 4, and Sensor 5, all had significant differences ($p < 0.05$). The force in the midfoot of the flatfoot was greater than that of the normal foot, suggesting the drop of flatfoot arch may result in a poor buffer or propulsive mechanism, as is what has been reported regarding the alternation of the peak value of the ground reaction force between the flatfoot and normal foot population. Without using an expensive force plate to measure the ground reaction force, the results of the pressure sensing test reported in the current study can also provide evidence for the significant differences between the medial and lateral arches of the flatfoot and normal feet, which can be further used to distinguish between a flatfoot and normal foot.

There have been various methods and instruments used to diagnose flatfoot in the past few decades [33,34], but there are still no unified methods for determining flatfoot by referring to both the static and dynamic data. Several parameters measured by the smart insole developed in the current study were compared with the AI and proved to be correlated. For better classification of the type of foot arch, Sensor 5 was sufficient to measure whether the foot was flat, which can be another alternative to facilitate and simplify the procedure of assessing the arch of the foot.

4.4. Research Limitation and Future Works

The effectiveness of using AI as a measurement of judging the foot has been questioned as the AI of 24 either overweight or obese elders were related to the body mass index (BMI) [35,36] and the AI measurement could reflect the "fatness" of the foot rather than the "flatness" of the foot. As it has been indicated that obesity may influence the middle third of the foot, especially in overweight or obese subjects [37], the BMI should be considered as a potentially confounding factor when comparing differences between groups of the AI.

In terms of sensor placement, it has been suggested that future studies place sensors on the insole in other remote positions from the arch, such as the great toe or the heel, to complete the design concept

by developing a sensor map of the entire foot with more index obtained from different combinations of measured pressure. The guideline of the placements of the sensors in the current study could serve as one of the options for future studies or for future smart insole development that involves the monitoring of the arch-relative index in the industry.

In terms of the algorithm, it is suggested that an artificial intelligence algorithm with a mass of available data is applied to find out the importance of each sensing point in the application to improve the efficiency and accuracy of judgment. As correlations were noted between the AI and musculoskeletal diseases in previous studies, such as plantar fasciitis [38], midfoot osteoarthritis [39], and medial knee osteoarthritis [37], the application of the current method to a variety of clinical patients is also suggested in future studies.

5. Conclusions

An Arduino-based pressure-sensor insole was prototyped for assessing flatfoot by choosing a suitable software and hardware architecture from measuring to receiving and designing a simplified self-made measurement plantar pressure instrument. After defining the position of the arch sensors relative to the length of the insole, which served as the placement rule of the sensors at the arch, the effectiveness of different sensor placements and flatfoot type assessment were tested during static and dynamic trials. By performing a correlation between traditional AI and the value of the self-made insole-based sensors with different sensor-placement options, and by comparing whether there were significant differences between the sensors at specific positions of flatfoot and normal foot from the data of 21 young subjects, it was found that during both the static and dynamic tests, the self-developed prototype could determine most of the flatfoot subjects and the most effective plantar pressure sensor position was also determined. Thus, in addition to the traditional subjective judgment of doctors, expensive X-ray detection, or judgments using a conventional ink-type area, the low-price insole-type measurement, including both static standing and dynamic waking, testing could be another wearable tool that reflects the medial longitude arch of the foot. It is suggested that in the future, this device can be applied in the insoles to examine the AI of the foot in clinical purposes due to its advantages when compared to the expensive-high tech graphic input board and conventional tools.

References

1. Jack, E.A. Naviculo-cuneiform fusion in the treatment of flat foot. *J. Bone Jt. Surg. Br. Vol.* **1953**, *35*, 75–82. [CrossRef]
2. Carr, J.B.; Yang, S.; Lather, L.A. Pediatric pes planus: A state-of-the-art review. *Pediatrics* **2016**, *137*, e20151230. [CrossRef] [PubMed]
3. Lee, M.S.; Vanore, J.V.; Thomas, J.L.; Catanzariti, A.R.; Kogler, G.; Kravitz, S.R.; Miller, S.J.; Gassen, S.C. Diagnosis and treatment of adult flatfoot. *J. Foot Ankle Surg.* **2005**, *44*, 78–113. [CrossRef] [PubMed]
4. Aenumulapalli, A.; Kulkarni, M.M.; Gandotra, A.R. Prevalence of flexible flat foot in adults: A cross-sectional study. *J. Clin. Diagn. Res.* **2017**, *11*, AC17–AC20. [CrossRef] [PubMed]

5.   Brody, D.M. Techniques in the evaluation and treatment of the injured runner. *Orthop. Clin. N. Am.* **1982**, *13*, 541–558.

6.   Cavanagh, P.R.; Rodgers, M.M. The arch index: A useful measure from footprints. *J. Biomech.* **1987**, *20*, 547–551. [CrossRef]

7.   Rao, U.B.; Joseph, B. The influence of footwear on the prevalence of flat foot. A survey of 2300 children. *Bone Jt. J.* **1992**, *74*, 525–527. [CrossRef]

8.   Riddiford-Harland, D.; Steele, J.; Storlien, L. Does obesity influence foot structure in prepubescent children? *Int. J. Obes.* **2000**, *24*, 541. [CrossRef]

9.   Hernandez, A.J.; Kimura, L.K.; Laraya, M.H.F.; Fávaro, E. Calculation of staheli's plantar arch index and prevalence of flat feet: A study with 100 children aged 5–9 years. *Acta Ortopédica Brasileira* **2007**, *15*, 68–71. [CrossRef]

10. Chen, C.-H.; Huang, M.-H.; Chen, T.-W.; Weng, M.-C.; Lee, C.-L.; Wang, G.-J. The correlation between selected measurements from footprint and radiograph of flatfoot. *Arch. Phys. Med. Rehabil.* **2006**, *87*, 235–240. [CrossRef] [PubMed]

11. Hawes, M.R.; Nachbauer, W.; Sovak, D.; Nigg, B.M. Footprint parameters as a measure of arch height. *Foot Ankle* **1992**, *13*, 22–26. [CrossRef] [PubMed]

12. Vauhnik, R.; Turk, Z.; Pilih, I.A.; Mičetić-Turk, D. Intra-rater reliability of using the navicular drop test for measuring foot pronation. *Hrvatski Športskomedicinski Vjesnik* **2006**, *21*, 8–11.

13. Menz, H.B.; Munteanu, S.E. Validity of 3 clinical techniques for the measurement of static foot posture in older people. *J. Orthop. Sports Phys. Ther.* **2005**, *35*, 479–486. [CrossRef] [PubMed]

14. Queen, R.M.; Mall, N.A.; Hardaker, W.M.; Nunley, J.A. Describing the medial longitudinal arch using footprint indices and a clinical grading system. *Foot Ankle Int.* **2007**, *28*, 456–462. [CrossRef] [PubMed]

15. Cavanagh, P.R.; Rodgers, M.M.; Liboshi, A. Pressure distribution under symptom-free feet during barefoot standing. *Foot Ankle* **1987**, *7*, 262–278. [CrossRef] [PubMed]

16. Armstrong, D.G.; Peters, E.J.; Athanasiou, K.A.; Lavery, L.A. Is there a critical level of plantar foot pressure to identify patients at risk for neuropathic foot ulceration? *J. Foot Ankle Surg.* **1998**, *37*, 303–307. [CrossRef]

17. Lavery, L.A.; Vela, S.A.; Fleischli, J.G.; Armstrong, D.G.; Lavery, D.C. Reducing plantar pressure in the neuropathic foot: A comparison of footwear. *Diabetes Care* **1997**, *20*, 1706–1710. [CrossRef] [PubMed]

18. Saito, M.; Nakajima, K.; Takano, C.; Ohta, Y.; Sugimoto, C.; Ezoe, R.; Sasaki, K.; Hosaka, H.; Ifukube, T.; Ino, S. An in-shoe device to measure plantar pressure during daily human activity. *Med. Eng. Phys.* **2011**, *33*, 638–645. [CrossRef] [PubMed]

19. Shu, L.; Hua, T.; Wang, Y.; Li, Q.; Feng, D.D.; Tao, X. In-shoe plantar pressure measurement and analysis system based on fabric pressure sensing array. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 767–775. [PubMed]

20. Eguchi, R.; Yorozu, A.; Fukumoto, T.; Takahashi, M. Ground Reaction Force Estimation Using Insole Plantar Pressure Measurement System from Single-Leg Standing. In Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Baden-Baden, Germany, 19–21 September 2016; pp. 109–113.

21. Howell, A.M.; Kobayashi, T.; Hayes, H.A.; Foreman, K.B.; Bamberg, S.J.M. Kinetic gait analysis using a low-cost insole. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3284–3290. [CrossRef] [PubMed]

22. Liang, T.-C.; Lin, J.-J.; Guo, L.-Y. Plantar pressure detection with fiber bragg gratings sensing system. *Sensors* **2016**, *16*, 1766. [CrossRef] [PubMed]

23. Navarro, L.A.; García, D.O.; Villavicencio, E.A.; Torres, M.A.; Nakamura, O.K.; Huamaní, R.; Yabar, L.F. Opto-electronic system for detection of flat foot by using estimation techniques: Study and approach of design. In Proceedings of the 2010 Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina, 31 August–4 September 2010; pp. 5768–5771.

24. Flórez, J.; Velasquez, A. Calibration of force sensing resistors (fsr) for static and dynamic applications. In Proceedings of the 2010 IEEE ANDESCON, Bogota, Colombia, 15–17 September 2010; pp. 1–6.

25. McCrory, J.; Young, M.; Boulton, A.; Cavanagh, P. Arch index as a predictor of arch height. *Foot* **1997**, *7*, 79–81. [CrossRef]

26. Kanatli, U.; Yetkin, H.; Cila, E. Footprint and radiographic analysis of the feet. *J. Pediatr. Orthop.* **2001**, *21*, 225–228. [CrossRef] [PubMed]

27. Yalcin, N.; Esen, E.; Kanatli, U.; Yetkin, H. Evaluation of the medial longitudinal arch: A comparison between the dynamic plantar pressure measurement system and radiographic analysis. *Acta Orthop. Traumatol. Turc.* **2010**, *44*, 241–245. [CrossRef] [PubMed]

28. Menz, H.B.; Morris, M.E. Clinical determinants of plantar forces and pressures during walking in older people. *Gait Posture* **2006**, *24*, 229–236. [CrossRef] [PubMed]

29. Jonely, H.; Brismée, J.-M.; Sizer, P.S., Jr.; James, C.R. Relationships between clinical measures of static foot posture and plantar pressure during static standing and walking. *Clin. Biomech.* **2011**, *26*, 873–879. [CrossRef] [PubMed]

30. Xiong, S.; Goonetilleke, R.S.; Witana, C.P.; Weerasinghe, T.W.; Au, E.Y.L. Foot arch characterization: A review, a new metric, and a comparison. *J. Am. Podiat. Med. Assoc.* **2010**, *100*, 14–24. [CrossRef]

31. Kernozek, T.W.; Ricard, M.D. Foot placement angle and arch type: Effect on rearfoot motion. *Arch. Phys. Med. Rehabil.* **1990**, *71*, 988–991. [PubMed]

32. Elvira, J.; Vera-Garcia, F.; Meana, M. Subtalar joint kinematic correlations with footprint arch index in race walkers. *J. Sports Med. Phys. Fitness* **2008**, *48*, 225. [PubMed]

33. Shrader, J.A.; Popovich, J.M., Jr.; Gracey, G.C.; Danoff, J.V. Navicular drop measurement in people with rheumatoid arthritis: Interrater and intrarater reliability. *Phys. Ther.* **2005**, *85*, 656–664. [PubMed]

34. Murley, G.S.; Menz, H.B.; Landorf, K.B. A protocol for classifying normal-and flat-arched foot posture for research studies using clinical and radiographic measurements. *J. Foot Ankle Res.* **2009**, *2*, 22. [CrossRef] [PubMed]

35. Wearing, S.C.; Hills, A.P.; Byrne, N.M.; Hennig, E.M.; McDonald, M. The arch index: A measure of flat or fat feet? *Foot Ankle Int.* **2004**, *25*, 575–581. [CrossRef] [PubMed]

36. Aurichio, T.R.; Rebelatto, J.R.; De Castro, A.P. The relationship between the body mass index (bmi) and foot posture in elderly people. *Arch. Gerontol. Geriatr.* **2011**, *52*, e89–e92. [CrossRef] [PubMed]

37. Levinger, P.; Menz, H.B.; Fotoohabadi, M.R.; Feller, J.A.; Bartlett, J.R.; Bergman, N.R. Foot posture in people with medial compartment knee osteoarthritis. *J. Foot Ankle Res.* **2010**, *3*, 29. [CrossRef] [PubMed]

38. Pohl, M.B.; Hamill, J.; Davis, I.S. Biomechanical and anatomic factors associated with a history of plantar fasciitis in female runners. *Clin. J. Sport Med.* **2009**, *19*, 372–376. [CrossRef] [PubMed]

39. Menz, H.B.; Munteanu, S.E.; Zammit, G.V.; Landorf, K.B. Foot structure and function in older people with radiographic osteoarthritis of the medial midfoot. *Osteoarthr. Cartil.* **2010**, *18*, 317–322. [CrossRef] [PubMed]

MDPI

Article

A Unified Deep-Learning Model for Classifying the Cross-Country Skiing Techniques Using Wearable Gyroscope Sensors

Jihyeok Jang [1,†], Ankit Ankit [2,†], Jinhyeok Kim [1], Young Jae Jang [1], Hye Young Kim [3], Jin Hae Kim [4] and Shuping Xiong [1,*]

[1] Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea; natejjh2@kaist.ac.kr (J.J.); mbnmbckbs@kaist.ac.kr (J.K.); yjang@kaist.edu (Y.J.J.)

[2] Department of Mechanical and Industrial Engineering, Indian Institute of Technology Roorkee, Uttarakhand 247667, India; ankit1@me.iitr.ac.in

[3] Division of Liberal Arts and Science, Korea National Sport University, Seoul 05541, Korea; hykim@knsu.ac.kr

[4] Department of Physical Education, Korea National Sport University, Seoul 05541, Korea; kimjhski@knsu.ac.kr

* Correspondence: shupingx@kaist.ac.kr; Tel.: +82-42-350-3132

† Both authors contributed equally to this work.

Abstract: The automatic classification of cross-country (XC) skiing techniques using data from wearable sensors has the potential to provide insights for optimizing the performance of professional skiers. In this paper, we propose a unified deep learning model for classifying eight techniques used in classical and skating styles XC-skiing and optimize this model for the number of gyroscope sensors by analyzing the results for five different configurations of sensors. We collected data of four professional skiers on outdoor flat and natural courses. The model is first trained over the flat course data of two skiers and tested over the flat and natural course data of a third skier in a leave-one-out fashion, resulting in a mean accuracy of ~80% over three combinations. Secondly, the model is trained over the flat course data of three skiers and tested over flat course and natural course data of one new skier, resulting in a mean accuracy of 87.2% and 95.1% respectively, using the optimal sensor configuration (five gyroscope sensors: both hands, both feet, and the pelvis). High classification accuracy obtained using both approaches indicates that this deep learning model has the potential to be deployed for real-time classification of skiing techniques by professional skiers and coaches.

Keywords: sports analytics; deep learning; classification; inertial sensor; cross-country skiing; classical style; skating style

1. Introduction

1.1. Problem Definition

Cross-country (XC) skiing is a whole-body exercise endurance sport, which requires prolonged complex cyclical motions performed using skis and poles on the snow [1]. There are two main styles in XC-skiing: the classical and the skating style. The classical style can be performed both on prepared trails with pairs of parallel grooves cut into the snow or on natural undisturbed snow whereas the skating style is generally performed on firm and smooth snow surfaces. Each of the classical and skating styles have four techniques or gears. These are diagonal stride (DS), double poling (DP), push-off (P-Off), and kick-double poling (KDP) for the classical style, and V2 skate (V2), V2A skate (V2A), V1 skate (V1), and free skate (FS) for the skating style, respectively. For ease of understanding

and reference, we will henceforth refer to these techniques using the abbreviations mentioned after the name of the technique.

In XC-skiing, the performance of the skiers depends on the biomechanical and the physiological aspects of the motions of the body parts and the sequence in which the skiing techniques are performed on the uphill and downhill tracks (commonly known as a natural course) and flat tracks (flat course). As the results of the skiing races can be determined by time steps as small as a few milliseconds, it becomes imperative for the professional coaches to understand both these aspects of XC-skiing to recommend an improved set of techniques for optimizing the performance of the skiers. Traditionally, these analyses have been performed using video-based systems [2–6] and/or force measurement systems [7–9]. However, the utilization of equipment in both these systems interferes with the natural movements of the body parts and the heavy cost involved limits their practical usage to only a few researchers across the world [10].

Body worn sensors, particularly the inertial sensors have recently emerged as a convenient substitute for such systems due to their small size, light weight, and low cost. Inertial sensors are sensors based on inertia and relevant measuring principles. In general, inertial sensors include gyroscopes used for measurements of the sensor's angular velocity and accelerometers for measurements of linear acceleration. These sensors can sample at high frequencies and are easily attached to the skier's body without interfering with the natural motion during skiing. This ease of use has made it possible to carry out experiments that require sensor data outside the controlled environment of the laboratory and provide a more realistic analysis of the task at hand. Marshland et al. [11] were the first to demonstrate this potential of body worn microsensors in the identification of XC-skiing techniques by plotting acceleration and angular velocity curves for eight athletes for both the classical and skating techniques. By visual inspection of the cyclical patterns in these plots, they concluded that all the classical and skating techniques can be clearly identified for each skier, with certain variations unique to each skier.

1.2. Literature Review and Proposed Work

Traditionally, studies of XC-skiing techniques have been limited to the kinematical and biomechanical analysis of various techniques [12,13]. These studies aim to determine numerous hard rules, like cycle time, poling/pushing time, recovery time [5,14,15], number of recovery motions, the sign of forearm angular velocity [16,17], correlation of the angular velocity of arms and legs [18], and figures showing identifiable cyclic patterns in the gyroscope and accelerometer data for the classification of techniques. These approaches, however, are extremely time-consuming as the derivation of the classification rules requires manual analysis of the gyroscope and accelerometer data from multiple sensors. Recently, many researchers have analyzed and classified techniques of XC-skiing using algorithms, like markov chains of multivariate distributions, and more advanced machine learning techniques. Stoggl et al. [19] utilized an accelerometer attached to the chest of professional skiing skiers to classify skating techniques. They collected data of 11 skiing skiers on a treadmill and developed a classification model based on the markov chains of multivariate distributions. Their model achieved an accuracy of 86% ± 8.9% on the test set when the training data included data from all the skiers, which rose to 90.3% ± 4.1% when separate classification models were developed for each skier. Rindal et al. [1] utilized a neural network for the classification of skating techniques by utilizing two sensors-a gyroscope on the arm for cycle identification and an accelerometer on the chest for technique classification. They achieved an accuracy of 93.9% on the test set. In both studies, the raw data was passed through a gaussian filter for the removal of ringing effects and undesirable time shifts at different frequencies. Ristner [20] implemented a Markov model and a k-nearest neighbors (KNN) algorithm for classifying XC skiing techniques using a 3D accelerometer attached on the chest of the skiers. The comparison showed that the KNN algorithm showed much lower error rates (0.19%) than the Markov model (7.22%). All these studies are impressive and show promise for the automatic and reliable classification of XC-skiing techniques using inertial sensors.

However, these studies suffer from many limitations. In the study performed by [19], the data is collected in the controlled environment of the laboratory, which will be different from the actual on-field data. The neural network model used by [1] takes in data of each cycle after flattening it into a single vector. This leads to information loss as the spatial and temporal patterns in the data are lost. Both studies develop models that classify either only the classical or skating techniques and do not obtain a single model, which could be employed for both the styles. Table 1 summarizes the relevant details of the aforementioned studies.

In this paper, we propose a unified convolutional neural network (CNN) and long-short term memory (LSTM) based deep learning classification model, which can be used to classify both the classical and skating style techniques of the skiers simultaneously. The first novelty of our approach lies in using convolutional layers for merging the local interactions among the time-series data obtained from each sensor and recurrent layers (long-short term memory layers) for extracting the temporal patterns. In this way, the model is able to extract important features for the classification of various techniques automatically from the raw data, thus eliminating the need for manually designed features required by machine learning algorithms. To prove this point, we present a comparison of the results obtained from our model with a KNN model developed by manually extracting features on the same training, validation, and test datasets. We collected the flat and natural course data of four professional skiers in total and pose a working hypothesis that the generalization accuracy of the proposed deep learning model increases as the amount of training data is increased. To prove this point, the model is first trained on the flat course data of two professional skiers and tested on the flat and natural course data of a third skier in a leave-one-out fashion. Secondly, the model is trained over the flat course data of three skiers and tested on the flat and natural course data of the fourth skier. An increase in the accuracy of classification when the size of the training data is increased confirms this hypothesis.

The second novelty of our approach lies in developing a unified model, which can be used for the classification of both classical and skating techniques simultaneously. We present strong evidence in favor of using only the flat course data for training the model and using it to classify XC-skiing techniques both on flat and natural courses, thus eliminating the need for collecting natural course data for training, which is extremely difficult to procure. Finally, the comparison of accuracies among five different combinations of sensors, which establishes the sports biomechanics configuration (both hands, both feet, and the pelvis sensors) as the optimal set of sensors, provides empirical evidence to researchers to base their future studies on this optimal configuration.

Table 1. Previous studies on XC-skiing technique classification with wearable sensors.

Related Work	Number of Sensors/Locations	Subjects	Number of Classes (XC-Skiing Style)	Used Classification Method	Classification Accuracy	Data Acquisition Details
[18]	6 inertial measurement units (IMUs)/upper back, lower back, left and right arm, left and right ankle; 1 sensor unit/chest (including 1 accelerometer, 1 gyroscope, 1 skin temperature sensor, 1 heart rate sensor)	11 (10 males, 1 female)	3 (classical style)	An algorithm based on the correlations of angle values of arms and legs	Sensitivity: 99–100%	Data collected in different types of tracks and snow conditions
[19]	1 accelerometer/chest	11 (7 males, 4 females)	5 (skating style)	Machine Learning Model (Markov Chain of multivariate Gaussian distribution)	86% ± 8.9% for collective data	Data collected on treadmill using roller skies (different speeds and inclines)
[20]	1 accelerometer/chest	3 skiers for the test set	Both classical style and skating style	Machine Learning Models (a Markov model and a KNN model)	The error rates for tests: 7.22% for the Markov model; 0.19% for KNN model	The test set for the cross-validation consists of 30 cycles from each gear from three different skiers in classical style and from three different skiers in skating style
[1]	2 inertial sensors: 1 accelerometer/chest; 1 gyroscope/arm	10 (9 males, 1 female)	8 (classical style)	Neural Networks	93.9% ± 3% for the test data	Data collected on treadmill and real competition course on snow
[21]	1 accelerometer/chest	No details	2 (skating style)	Neural Networks	CNN error rate: 2.4% LSTM error rate: 1.6%	No detailed information

2. Materials and Methods

2.1. Inertial Sensors, Synchronization and Calibration

XSens MVN motion capture system (Xsens Technologies B.V., Enschede, The Netherlands) consisting of 17 body-wired inertial motion trackers (Figure 1) was used to record the participants' kinematical data at a frequency of 240 Hz. Each motion tracker is 36 × 24.5 × 10 mm in dimension, weighs 10 g, and is mounted at a specific body location with the help of a wearable lycra suit. The wired motion trackers are connected to an on-body data hub (known as bodypack), which is responsible for synchronization and gathering data on its internal memory (Xsens MVN Technical Report, March 2018).
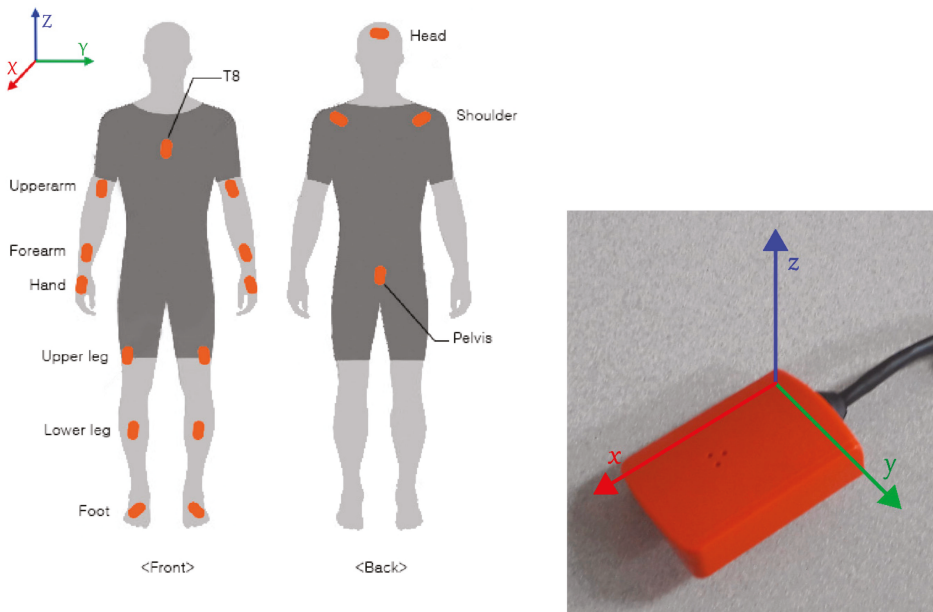
Figure 1. (**Left**) Front and back view of the position of 17 inertial motion trackers that are attached to the Xsens bodypack worn by the skiers. (**Right**) One Xsens inertial motion tracker depicting the local *x*, *y*, and *z* axes.

Calibrations were performed by the system software, *Xsens MVN Analyze*, prior to data collection, which requires the subject's height and foot length to estimate the dimensions and proportions of the person being tracked. After subject calibration, he/she is asked to stand still in a T-pose and walk a few meters back and forth for the purpose of the sensor to segment calibration and development of a biomechanical human model. This is followed by a slight forward movement for defining the positive *X*-axis. The *Y*-axis is perpendicular to the *X*-axis in the horizontal plane while the *Z*-axis is perpendicular to the horizontal surface. Each motion tracker has 3 sensors: accelerometer, gyroscope, and magnetometer, which provide raw recorded data of linear acceleration, angular velocity, and magnetic field intensity along the *x*, *y*, and *z* axis local to each sensor, respectively.

The on-body recording function was used for data collection, which allows recording of the subject's motions without the need for a laptop or PC by storing the motion trackers' data on the body pack. After finishing the data recording, the body pack was connected with a laptop to import the recordings for further processing and analysis.

2.2. Data Acquisition

2.2.1. Training and Validation Data Acquisition

For the purpose of training and validating the trained model, XC-skiing data from 3 professional skiers (Table 2) from the Korea National Sport University was collected. They performed the classical and skating XC-skiing techniques on outdoor flat and natural courses in Pyeongchang, South Korea, where the 2018 Winter Olympic Games took place. All of them were informed of the purpose of this study and they participated voluntarily in the experiment after reading the research guidelines and signing consent forms. The study was ethically approved by the Korea National Sport University Institutional Review Board (IRB Number 20170424-004).

Table 2. General characteristics of the three professional skiers that participated in the study.

Attribute Skier	Gender	Age (in Years)	Weight (in kg)	Height (in cm)
1	Female	24	51	163
2	Female	22	51	162
3	Male	23	69	176

In the training dataset, each skier performs only one technique of either the classical or skating styles on the flat course repeatedly. Each subject performs 5–6 laps of each technique on a 500 m long track. As there are 8 techniques (4 classical and 4 skating) and 3 skiers, a total of 24 such files are obtained. In the validation dataset, each skier is allowed to perform either all the 4 classical or skating techniques on either a 2.5 km long flat or natural course, similar to what he/she would perform under competitive conditions. The skiers are free to make transitions from one technique to the other; however, the skiing style remains the same. A total of 11 files are obtained in this manner (1 file could not be obtained due to unavailability of the tracks). For both the datasets, a video recording of the skiers while performing the skiing techniques is also shot. Table 3 lists the type of data collected for each subject. In order to examine whether our developed model could classify skiing techniques for skiers with different skill levels, we allowed the skiers to freely choose their own preferred skiing speeds and exercise intensities during the data collection.

Table 3. Training and validation data collected for three professional skiers characterized by the type of course (flat/natural) and the number and type of skiing techniques (classical/skating) that the subject is allowed to perform simultaneously.

Skier	Training Dataset	Validation Dataset			
	Flat Course; 1 Technique of Skating or Classical Style Repeatedly	Flat Course		Natural Course	
		Classical Style (DS, P-Off, KDP, DP)	Skating Style (V2, V2A, V1, FS)	Classical Style (DS, P-Off, KDP, DP)	Skating Style (V2, V2A, V1, FS)
Skier 1	√	√	√	X	√
Skier 2	√	√	√	√	√
Skier 3	√	√	√	√	√

X: The classical style data on the natural course for skier 1 is not available.

Each file in the training data exclusively contains the data of one of the techniques of one of the skiing styles, and hence does not require any labelling. However, the 4 techniques of each XC-skiing style in validation data are performed in a combined way and hence have to be labelled after data collection. The ground truth labels for the validation data are developed by professional cross-country skiing players from the Korea National Sport University by simultaneously watching the recorded video from a digital camera, human model video from the *XSens MVN Analyze*, and marking the frames corresponding to each technique in the raw data files. Labelling follows a 0–9 convention for each file: 0: start/end of the recording, 1: DS, 2: P-Off, 3: KDP, 4: DP, 5: V2, 6: V2A, 7: V1, 8: FS, 9: descending). The labels have been double checked by a professional XC skiing coach at the Korea National Sport University.

2.2.2. Test Data Acquisition

For the purpose of testing the generalization of the trained model, two test datasets of a new skier (skier 4) were collected. The test subject (gender: Female, age: 24 years, weight: 55 kg, height: 156 cm) is also a professional XC-skiing player from the Korea National Sport University and prior information about the purpose of the study was provided to her followed by signing of consent forms (IRB Number 20170424-004).

The 2 types of test data are as follows:

(i) Test set 1: In the first type of test data, the subject is allowed to perform only one of the techniques of one of the XC-skiing styles on the flat course. As there are 4 classical and 4 skating techniques, this test set consists of 8 files.

(ii) Test set 2: In the second type of test data, the subject performs all the skating style techniques on a natural course simultaneously. The subject is allowed to make transitions between the various skating techniques similar to what would be performed during a competition. One data file is obtained in this manner.

2.3. Data Selection and Preprocessing

In this study, each training instance represents one cycle of one of the techniques of either the classical or skating styles. Thus, it becomes extremely important to select the data that can represent these cyclic patterns most clearly. Figure 2 represents the typical data patterns in linear acceleration and angular velocity data for each of the 4 classical and 4 skating techniques. These figures were plotted after filtering the raw data using a low pass butterworth filter of fourth order and a cutoff frequency of 0.007 Hz. As is clear from the figure, angular velocity data, which is obtained via the gyroscope, shows more easily identifiable cyclic patterns as compared to linear acceleration data, which comes from the accelerometer. Due to the ease of identifying the cycles and low computational cost from the less sensor data, only the gyroscope data is used in this study for developing the classification models.
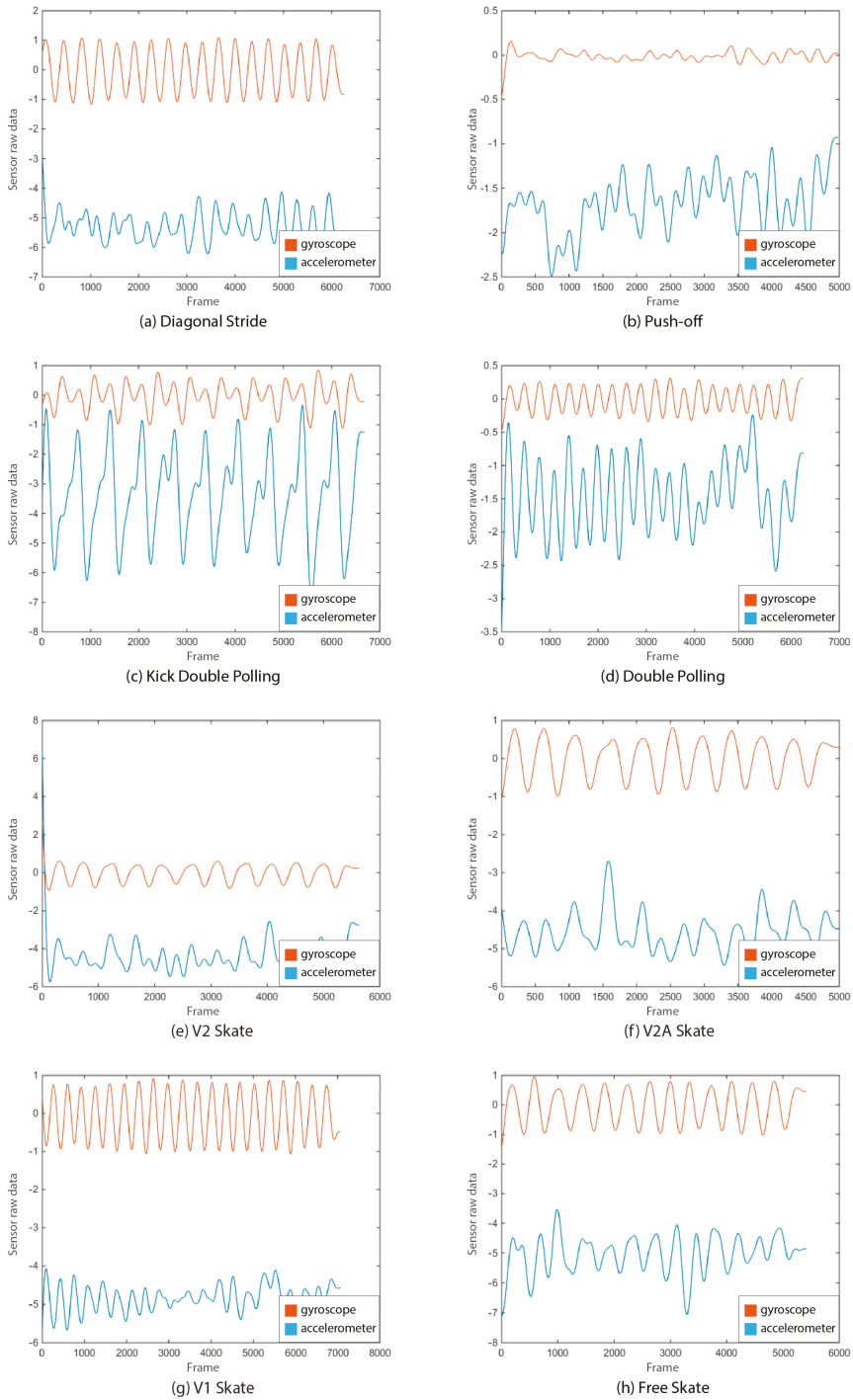
Figure 2. Comparison of the cyclic patterns in the *z*-axis (anteroposterior direction: normal to frontal plane) angular velocity and linear acceleration data (from a motion tracker on the flat surface of the shin bone of left leg) for the classical (**a–d**) and skating (**e–h**) XC- skiing techniques.

2.3.1. Training Dataset

Each of the 24 files in the training dataset, in which each skier performs only 1 technique of either the classical or skating style on the flat course repeatedly, has turning points at the end of a lap each time the track is traversed. During the turning points, the skier is not performing any technique and hence these points are removed from the data of each skier by manually identifying the frames corresponding to such durations using the Xsens human model videos and the recorded videos from a digital camera. This gives data files with continuous repetitions of the same technique. A low-pass butterworth filter of fourth order and a cutoff frequency 0.007 Hz is applied to smoothen the raw data, following which the z-axis angular velocity of the gyroscope on the left leg is used for finding the locations of peaks. The distance between two consecutive peaks in the filtered data is variable and represents 1 cycle. Each input to the CNN-LSTM model must have the same dimensions. Thus, the locations of the peaks in the filtered data is used to resample the raw data (after removing the turning points) to a fixed cycle-length of 333 time-steps (which is the mean number of time-steps for all the techniques in the training data) using an anti-aliasing finite impulse response low pass filter, which resamples the data at $(333/n) \times 240$ Hz, where n is the number of time steps in a given cycle before resampling and 240 Hz is the original sample rate. Resampling is performed over the raw data and not over the filtered data. It is because a neural network works best with raw data from which it automatically extracts the features required for the classification task. These resampled cycles are arranged into tensors, which make it suitable for passing them to a CNN layer and later to an LSTM layer as a multivariate time series. We thus obtain a total of 24 tensors. The number of cycles of each technique of each XC-skiing style performed by each skier on the flat course is shown in Table 4.

Table 4. Number of cycles of the classical (DS, P-Off, KDP, DP) and skating (V2, V2A, V1, FS) techniques performed by each skier in the training dataset.

Technique / Skier	Classical Style				Skating Style				Sum
	DS	P-Off	KDP	DP	V2	V2A	V1	FS	
Skier 1	153	123	107	128	85	104	157	103	960
Skier 2	103	83	94	68	38	56	65	61	568
Skier 3	83	64	64	78	48	59	86	57	539
Sum	339	270	265	274	171	219	308	221	2067

The dimensions of each matrix in a tensor is 333×51 (17 gyroscopes, 3 axes each, hence 51 columns).

2.3.2. Validation Dataset

In the validation dataset, each skier performs either all the 4 classical or skating techniques on either a flat course or a natural course simultaneously. Each of the 11 files in the validation dataset has starting and end points of the recording (labelled as 0), descending points (labelled as 9), and transition points (not labelled), which are classified as noise. The start/end and descending points are deleted from the data and the transition points are removed by manually identifying the frames corresponding to such time durations using the Xsens human model videos. The filtering of this data for cycle detection followed by resampling and arrangement into tensors is in accordance with what was performed for the training dataset. A total of 11 tensors are obtained in this manner and the dimensions of each matrix in a tensor is 333×51 (17 gyroscopes, 3 axes each, hence 51 columns). Table 5 provides further information about this dataset.

Table 5. Number of cycles of each technique for the classical and skating style performed by each skier in the validation dataset on the flat and natural course.

Skiing Course & Technique	Flat Course		Natural Course	
	Classical Style (DS, P-Off, KDP, DP)	Skating Style (V2, V2A, V1, FS)	Classical Style (DS, P-Off, KDP, DP)	Skating Style (V2, V2A, V1, FS)
Skier 1	33, 13, 39, 0	44, 60, 7, 21	X	60, 31, 197, 0
Skier 2	49, 0, 54, 26	13, 24, 15, 10	190, 0, 7, 63	98, 34, 602, 0
Skier 3	16, 27, 34, 46	23, 32, 19, 19	145, 2, 5, 85	89, 30, 44, 0

X: The classical style data on the natural course for skier 1 is not available.

As can be observed from Table 5, different skiers have different preferences in terms of the techniques they use on a particular course. For example, skier 1 does not perform double poling (DP) on the flat course at all whereas skier 2 performs it 26 times and skier 3 performs it 46 times (highest among all classical techniques) on the same course. However, it is clear that there is a somewhat more even distribution among the usage of skating techniques as compared to classical techniques.

On the natural course, all the skiers do not use free skate (FS) and seldom use push off (P-Off) and kick double poling (KDP). Thus, it is clear that the skiers have certain preferences on the techniques they utilize on different courses (Table 5). This preference makes this dataset highly imbalanced and a perfect validation set for validating the performance of the trained deep learning model.

2.3.3. Test Dataset

In test set-1, the test subject (skier 4) performs only one of the techniques of one of the XC-skiing styles repeatedly on the flat course. It has 8 files (one file for each technique of the classical and skating styles) and resembles the training data. Thus, its preprocessing is performed analogously to the training data. Similarly, the test set-2 resembles the validation data and its preprocessing is performed analogously to the validation data. Thus, 8 tensors are obtained for test set-1 and 1 for the test set-2. Table 6 contains more information about the 2 types of test sets.

Table 6. Number of cycles of the classical (DS, P-Off, KDP, DP) and skating (V2, V2A, V1, FS) techniques performed by skier 4 in the test set-1 and test set-2.

Test Set-1 Flat Course								Test Set-2 Natural Course			
Classical Style				Skating Style				Skating Style			
DS	P-Off	KDP	DP	V2	V2A	V1	FS	V2	V2A	V1	FS
649	511	378	569	366	447	631	480	72	13	219	0

2.4. Architecture of the Deep Network

The deep learning model developed to classify the XC-skiing techniques is motivated from Deepsense proposed by Yao et al. [22], which in the authors' words "provides a general signal estimation and classification framework [for regression and classification problems] that accommodate a wide range of applications." The training, validation, and testing data of our problem is arranged into 3D tensors, where each matrix corresponds to one training (or testing) example and has 333 rows and 51 columns. Each column contains data along one of the axes as recorded by one sensor and represents a time-series with 333 time steps. Thus, the classification problem is posed as a multivariate time-series sequence classification task. To capture the interactions among these time series, they are passed through convolutional layers. Deepsense first performs fourier transformation on the raw data of each sensor, passes the frequency data of each sensor through convolutional layers individually, and then combines the data of all the sensors to pass it through another convolutional layer. This approach requires a greater number of convolutional layers and selecting the number of frequencies that must be

passed to the network, which introduces an element of human decision-making. We solve this problem by making 2 simple modifications to our model: (i) We pass the raw sensor data to the convolutional layers instead of the frequency data, and (ii) we convolve the raw data of all the sensors in a single step, which reduces the total number of convolutional layers required for training.

As the skiers may perform the same skiing techniques at varying speeds and intensities, it is necessary that the deep network layers be robust to the scale of the data and be able to capture features that may be found at different time-steps in the time-series. CNNs are very powerful in extracting local spatial coherence and dependencies in the data, and the scale invariance introduced by the max-pooling layers allows them to learn hidden features regardless of the position of the feature or its scale. We pass the raw data through two 1D convolutional layers with 64 filters each and 20 and 10 kernels, respectively, followed by a max-pooling layer with a pool-size of 4. This convolved data is then passed through 2 long-short term memory (LSTM) layers, with 300 and 200 units, respectively, to capture long-term temporal dependencies in the time series. As LSTMs are highly prone to overfitting, a dropout layer with a dropout probability of 0.2 is added after each LSTM layer. The network is trained over 12 epochs with a batch size of 40. Various steps involved in data preprocessing and the architecture of the deep network are summarized in Figure 3.
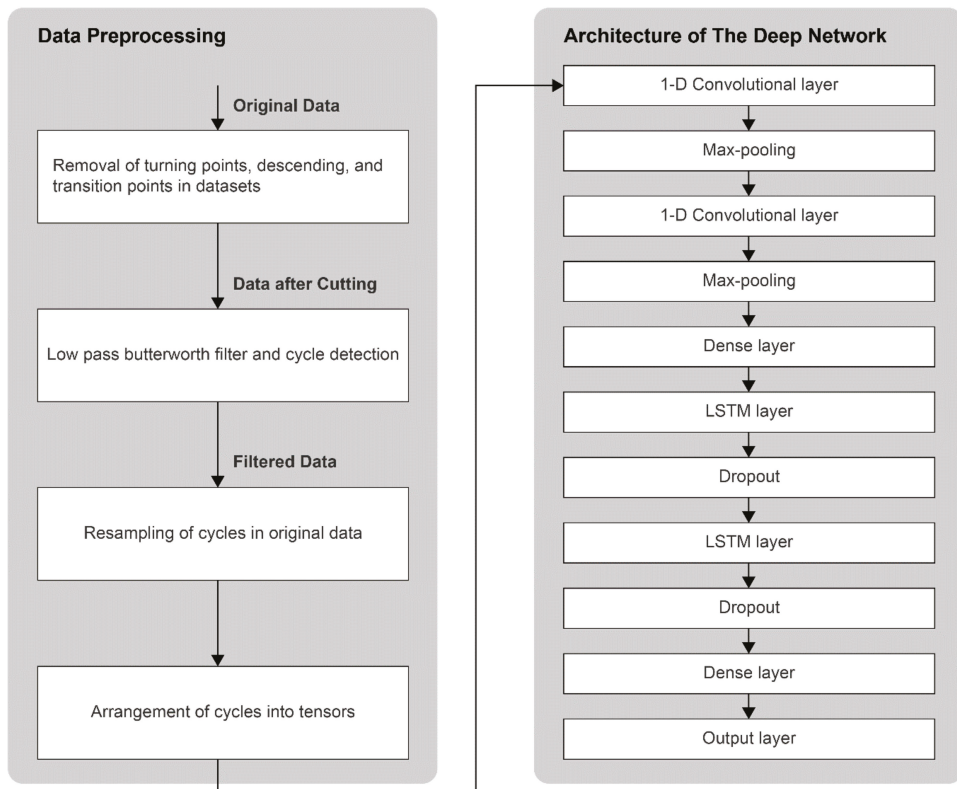
Figure 3. (**Left**) A schematic view of the steps involved in preprocessing of the raw data, and (**Right**) the architecture of the deep neural network.

3. Results

We now present results for the training, validation, and test sets obtained by training the proposed CNN-LSTM based deep learning model. To compare the results with a traditional machine learning

algorithm, we also trained a k-nearest neighbor (KNN) classifier by extracting manually designed features from the training data. Five different combinations of sensors are used for training the model as shown in Table 7. We will compute and compare the results for each of the sensor configurations and come up with the best subset of the 17 sensors, which should be used for the analysis of XC-skiing techniques in future studies.

Table 7. Five combinations of the sensors for which the CNN-LSTM based deep learning model is trained.

Sensor Configuration	Number of Sensors	Locations of Sensors
1	All 17 sensors (Whole-body)	Pelvis, chest, head, right and left shoulders, right and left upper arms, right and left forearms, right and left hands, right and left upper legs, right and left lower legs, right and left feet
2	11 sensors (Upper body only)	Pelvis, chest, head, right and left shoulders, right and left upper arms, right and left forearms, right and left hands
3	7 sensors (Lower body only)	Pelvis, right and left upper legs, right and left lower legs, right and left feet
4	5 sensors (Sports biomechanics configuration)	Pelvis, right and left hands, right and left feet
5	1 sensor (Pelvis only)	Pelvis

3.1. Training and Validation Set Results Using Deep Learning

The deep learning model, trained on the training data of the three skiers according to the architecture described in Section 2.4, resulted in a training dataset accuracy of at least 97.8% for the whole body, upper body, lower body, and sports biomechanics sensors configuration, and 79.4% when only the pelvis sensor is used (Table 8). Table 9 provides the confusion matrix for the training dataset for the sports biomechanics configuration.

Table 8. Training dataset accuracies for the five different sensor configurations.

Sensor Configuration	17 Sensors (Whole Body)	11 Sensors (Upper Body)	7 Sensors (Lower Body)	5 Sensors (Sports Biomech.)	1 Sensor (Pelvis)
Accuracy	99.39%	97.96%	98.03%	97.82%	79.4%

Table 9. Confusion matrix for the training dataset for the sports biomechanics configuration of sensors.

True \ Predicted	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	339	0	0	0	0	0	0	0
P-Off	0	242	0	28	0	0	0	0
KDP	0	0	265	0	0	0	0	0
DP	0	9	1	264	0	0	0	0
V2	0	1	0	1	168	0	1	0
V2A	0	0	0	0	1	218	0	0
V1	0	0	0	0	3	0	305	0
FS	0	0	0	0	0	0	0	221

As is clear from Table 9, the model is able to classify all the techniques almost perfectly except the classical push-off and double poling techniques. Twenty-eight push-off techniques have been wrongly classified as double poling techniques and nine double poling techniques as push-off techniques.

The validation dataset accuracies for the five different configurations of sensors are shown in Table 10.

Table 10. Classification accuracies for the validation dataset for five different sensor configurations.

Sensor Configuration / Skier	Course & Technique	Natural Course		Flat Course		Mean Accuracy
		Classical Style	Skating Style	Classical Style	Skating Style	
Whole body sensors (17)	Skier 1	X	89.58%	78.82%	76.52%	81.64%
	Skier 2	97.31%	97.68%	89.15%	79.03%	90.79%
	Skier 3	93.67%	82.21%	91.06%	89.25%	89.04%
	Mean Accuracy	95.49%	89.82%	86.34%	81.60%	~87.00%
Upper body sensors (11)	Skier 1	X	87.85%	74.12%	71.97%	77.98%
	Skier 2	95.38%	95.78%	85.27%	72.58%	87.25%
	Skier 3	66.24%	82.21%	73.17%	88.17%	77.45%
	Mean Accuracy	80.81%	88.61%	77.52%	77.57%	~80.00%
Lower body sensors (7)	Skier 1	X	54.86%	78.82%	52.27%	61.98%
	Skier 2	92.31%	25.34%	85.27%	69.35%	68.07%
	Skier 3	85.23%	66.87%	90.24%	80.65%	80.75%
	Mean Accuracy	88.77%	49.02%	84.78%	67.42%	~70.00%
Sports biomechanics configuration (5)	Skier 1	X	76.04%	88.23%	71.96%	78.74%
	Skier 2	94.61%	96.87%	91.47%	82.26%	91.30%
	Skier 3	96.62%	82.21%	88.62%	92.47%	89.98%
	Mean Accuracy	95.61%	85.04%	89.44%	82.23%	~87.00%
Pelvis sensor only (1)	Skier 1	X	74.20%	80.21%	59.30%	71.24%
	Skier 2	56.37%	78.79%	74.44%	52.36%	64.81%
	Skier 3	59.32%	64.73%	68.80%	31.23%	56.02%
	Mean Accuracy	57.85%	72.57%	74.48%	47.63%	~64.00%

X: The classical style data on the natural course for skier 1 is not available.

The mean classification accuracy with the pelvis, the upper, and the lower body sensors is 64%, 80%, and 70% respectively, which increases to approximately 87% for the 17 sensors (the whole body) and the sports biomechanics configuration. As the accuracies while using all the 17 sensors and the five sensors in sports biomechanics configuration are the same, the sports biomechanics configuration of sensors is the optimal set due to much smaller number of sensors. The confusion matrices for the validation sets for skier 3 for the sports biomechanics configuration are shown in Tables 11–14. The confusion matrices for the validation sets of skier 1 and skier 2 can be found in Appendices A.1 and A.2.

Table 11. Confusion matrix for the natural course, classical style validation set of skier 3 when using the sports biomechanics configuration of sensors.

True / Predicted	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	145	0	0	0	0	0	0	0
P-Off	0	1	0	1	0	0	0	0
KDP	0	0	4	0	1	0	0	0
DP	1	1	0	79	1	3	0	0
V2	0	0	0	0	0	0	0	0
V2A	0	0	0	0	0	0	0	0
V1	0	0	0	0	0	0	0	0
FS	0	0	0	0	0	0	0	0

Table 12. Confusion matrix for the natural course, skating style validation set of skier 3 when using the sports biomechanics configuration of sensors.

Predicted True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	0	86	3	0	0
V2A	0	0	0	0	1	5	23	1
V1	0	0	0	0	1	0	43	0
FS	0	0	0	0	0	0	0	0

Table 13. Confusion matrix for the flat course, classical style validation set of skier 3 when using the sports biomechanics configuration of sensors.

Predicted True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	16	0	0	0	0	0	0	0
P-Off	0	22	0	5	0	0	0	0
KDP	0	6	27	1	0	0	0	0
DP	0	0	1	44	1	0	0	0
V2	0	0	0	0	0	0	0	0
V2A	0	0	0	0	0	0	0	0
V1	0	0	0	0	0	0	0	0
FS	0	0	0	0	0	0	0	0

Table 14. Confusion matrix for the flat course, skating style validation set of skier 3 when using the sports biomechanics configuration of sensors.

Predicted True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	1	0	20	1	1	0
V2A	0	0	0	1	0	30	0	1
V1	0	0	0	0	1	0	17	1
FS	0	0	0	0	0	0	0	19

It is interesting to note that although the model has been trained on classical and skating styles data simultaneously, it has almost perfectly learnt to differentiate between these two styles. In Table 12, three V2 techniques have been incorrectly classified as V2A, and 23 out of the 28 V2A techniques have been incorrectly classified as V1. In Table 13, five push-off techniques have been incorrectly classified as DP and 6 KDP techniques as push-off. These are certain areas of misclassification errors, which the model is not robust to. However, despite these small misclassification errors, the mean classification accuracy achieved for skier 3 is approximately 90%, which is a very high value considering that our model is trained only on the flat course data and has never observed natural course data.

3.2. Leave-One-Out Testing Results

To assess the generalization accuracy of the model, we perform a leave-one-out type of testing in which the flat course data of two out of the initial three skiers is used for training and the flat and natural course data of the remaining third skier is used for testing. As the third skier can be chosen in $^{3}C_{1} = 3$ ways, we have a total of three combinations of training and test sets. For example, combination

1 includes subject 2 and subject 3's flat course data as the training set, and subject 1's flat and natural course data as the test set. We present the results of leave-one-out testing for both the proposed deep learning model as well as a traditional k-nearest neighbors (KNN) machine learning algorithm. For the KNN algorithm, the feature vector corresponding to a cycle in the training data consists of the pairwise correlation values between the time-series represented by each axis of each sensor in a cycle. For example, while using the whole-body configuration of sensors (17 sensors), there are a total of 51 time series in each cycle, which correspond to a feature vector of length 1276 ($=^{51}C_2 + 1$). Table 15 presents the results of the proposed deep learning model and Table 16 for the KNN machine learning model when a leave-one-out type of testing is performed using the sports biomechanics configuration of sensors. The results for the other four configurations of sensors are available in the Appendix A.3.

Table 15. Classification accuracies for the first three skiers using the proposed deep learning model when a leave-one-out type of testing is performed using the sports biomechanics configuration of sensors.

Course & Technique / Skier	Natural Course		Flat Course		Mean Accuracy
	Classical Style	Skating Style	Classical Style	Skating Style	
Skier 1	X	90.35%	76.48%	74.50%	80.44%
Skier 2	76.20%	97.70%	68.65%	85.68%	82.06%
Skier 3	90.28%	78.50%	79.70%	55.00%	75.87%
Mean Accuracy	83.24%	88.85%	74.94%	71.73%	79.69%

X: The classical style data on the natural course for skier 1 is not available.

Table 16. Classification accuracies for the first three skiers using the k-nearest neighbors machine learning model when a leave-one-out type of testing is performed using the sports biomechanics configuration of sensors.

Course & Technique / Skier	Natural Course		Flat Course		Mean Accuracy
	Classical Style	Skating Style	Classical Style	Skating Style	
Skier 1	X	86.05%	57.25%	58.33%	67.21%
Skier 2	75.09%	87.76%	50.38%	53.33%	66.64%
Skier 3	83.71%	56.98%	57.25%	37.86%	58.95%
Mean Accuracy	79.40%	76.93%	54.96%	49.84%	65.28%

X: The classical style data on the natural course for skier 1 is not available.

The overall mean accuracy for the leave-one-out type of testing for the three skiers using k-nearest neighbors' algorithm is ~65%, which increases to ~80% when the proposed deep learning model is used. Also, the mean accuracy values for each skier is higher in the case of the deep learning model as compared to the KNN model.

3.3. Test Set Result Using Deep Learning

To further assess the generalization performance of the model, it is trained on the flat course datasets of the first three subjects and tested on the two test datasets obtained from subject 4. The classification accuracies for test set-1 and test set-2 for all five sensor configurations are as shown in Table 17. Again, the sports biomechanics configuration with five sensors has maximum accuracy for both the test sets, reaffirming the aforementioned proposition that this configuration is the optimal set of sensors.

Table 17. Classification accuracies for test set-1 and test set-2 for the five different sensor configurations.

Sensor Configuration	17 Sensors (Whole Body)	11 Sensors (Upper Body)	7 Sensors (Lower Body)	5 Sensors (Sports Biomech.)	1 Sensor (Pelvis)
Test Set-1	84.21%	84.80%	84.35%	87.20%	58.54%
Test Set-2	92.18%	84.83%	64.58%	95.10%	29.65%
Mean Accuracy	88.19%	84.82%	74.47%	91.15%	44.10%

Tables 18 and 19 show the confusion matrices for test set-1 and test set-2 for the sports biomechanics configuration, respectively. In test set-1, all the techniques except the classical push-off (P-Off) and double poling (DP) have been classified almost perfectly. One hundred and seventy-two (out of total 241) push-off techniques have been incorrectly classified as double poling and 74 (out of total 295) double poling as push-off, leading to a low classification accuracy for the push-off and the double poling. These are the same two techniques that were confused by the model in the training set, and hence some misclassification in the test set was also expected. For test set-2, a very high overall accuracy of 95.1% is obtained for the sports biomechanics configuration of sensors. Thus, we achieve an overall mean accuracy of 91.15% on the test set of skier 4.

Table 18. Confusion matrix for the test set-1, in which the test subject performs one of the techniques of one of the XC-skiing styles repeatedly on a flat course at a time for the sports biomechanics configuration of sensors.

True \ Predicted	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	310	0	0	0	0	0	0	0
P-Off	0	63	1	172	5	0	0	0
KDP	0	0	113	0	0	0	0	0
DP	0	74	0	221	0	0	0	0
V2	0	1	0	0	193	0	1	0
V2A	0	0	0	1	4	221	2	0
V1	0	0	0	0	1	3	319	0
FS	0	0	0	0	0	0	0	259

Table 19. Confusion matrix for the test set-2 in which the test subject performs all skating techniques on a natural course simultaneously for the sports biomechanics configuration of sensors.

True \ Predicted	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	1	66	2	3	0
V2A	0	0	0	0	0	13	0	0
V1	0	0	0	0	6	1	210	2
FS	0	0	0	0	0	0	0	2

It should be emphasized that the overall mean accuracy when a leave-one-out type of testing is performed over the first three skiers is ~80%, which increases to ~91.1% when testing is performed over skier 4 using the same deep learning model. This is due to the fact that the deep learning model tested over skier 4 has been trained over the data of three skiers whereas the same model when tested over each of the first three skiers has been trained over the data of two skiers in a leave-one-out fashion. Thus, the generalization accuracy of the proposed model increases as the size of the training dataset is

increased. These results provide strong evidence in favor of our hypothesis that the accuracy of the deep learning model increases as the training datasets become larger.

3.4. Validation and Test Set Results Using K-Nearest Neighbors Algorithm

To compare the results of the deep learning model with a traditional machine learning algorithm, we trained a k-nearest neighbors classifier on the training data of three subjects and tested it on the fourth subject. The feature vector corresponding to a cycle in the training data consists of the pairwise correlation values between the time-series represented by each axis of each sensor in a cycle.

Table 20 represents the validation set and Table 21 represents the test set-1 and test set-2 accuracies for all five configurations of the sensors.

Table 20. Classification accuracies for the validation dataset for five different sensor configurations using the k-nearest neighbors algorithm.

Sensor Configuration	Course & Technique / Skier	Natural Course		Flat Course		Mean Accuracy
		Classical Style	Skating Style	Classical Style	Skating Style	
Whole body sensors (17)	Skier 1	X	60.54%	61.59%	60.54%	60.89%
	Skier 2	91.70%	97.50%	83.67%	80.00%	88.22%
	Skier 3	88.26%	83.72%	80.15%	83.50%	83.91%
	Mean Accuracy	89.98%	80.59%	75.14%	74.68%	80.10%
Upper body sensors (11)	Skier 1	X	66.33%	47.83%	66.32%	60.16%
	Skier 2	84.43%	93.95%	77.56%	88.00%	85.99%
	Skier 3	83.34%	81.40%	71.76%	81.55%	79.51%
	Mean Accuracy	83.89%	80.56%	65.72%	78.62%	77.20%
Lower body sensors (7)	Skier 1	X	68.37%	58.70%	68.37%	65.15%
	Skier 2	82.70%	38.55%	77.56%	64.00%	65.70%
	Skier 3	85.98%	53.49%	81.68%	61.17%	70.58%
	Mean Accuracy	84.34%	53.47%	72.65%	64.51%	68.74%
Sports biomechanics configuration (5)	Skier 1	X	50.68%	52.17%	50.68%	51.18%
	Skier 2	86.85%	95.92%	83.67%	81.34%	86.95%
	Skier 3	82.58%	76.74%	81.00%	83.50%	80.96%
	Mean Accuracy	84.72%	74.45%	72.28%	71.84%	75.82%
Pelvis sensor only (1)	Skier 1	X	2.72%	25.36%	2.08%	10.05%
	Skier 2	31.83%	31.18%	15.74%	10.67%	22.36%
	Skier 3	34.09%	7.56%	18.32%	13.59%	18.39%
	Mean Accuracy	32.96%	13.82%	19.81%	8.78%	18.84%

X: The classical style data on the natural course for skier 1 is not available.

Table 21. Classification accuracies for test set-1 and test set-2 for the five different sensor configurations using the k-nearest neighbors algorithm.

Sensor Configuration	17 Sensors (Whole Body)	11 Sensors (Upper Body)	7 Sensors (Lower Body)	5 Sensors (Sports Biomech.)	1 Sensor (Pelvis)
Test Set-1	48.06%	33.58%	65.54%	68.82%	38.13%
Test Set-2	86.82%	71.96%	71.28%	78.04%	7.43%

4. Discussion

We developed a unified CNN-LSTM based deep learning model for classifying both the classical and skating style techniques simultaneously using the gyroscope data. Even though our model was trained only on the outdoor flat course data, it achieved an accuracy of 87.2% and 95.1% on the flat and natural course test sets, respectively, leading to an overall mean accuracy of 91.15%, using the optimal gyroscope sensor configuration (five sensors: both hands, both feet, and the pelvis). This presents strong evidence in favor of using only the flat course data for training the model and using it to classify XC-skiing techniques both on flat and natural courses, thus eliminating the need for collecting natural

course data for training, which is extremely difficult to procure. To the best of our knowledge, we are the first ones to propose a unified deep learning model for classifying classical and skating techniques simultaneously with high accuracy. A KNN model with manually designed features for the skiing technique classification on the same datasets was further used as a benchmark for evaluating the performance of our unified deep learning model. The KNN algorithm was chosen since it is preferable due to less error rates in classifying the classical style and skating style simultaneously when compared with a Markov model according to an earlier study [20]. The comparison between the accuracies obtained from these two approaches for the validation dataset (Figure 4) and two test sets (Table 22) clearly showed that deep learning is more effective and has higher classification accuracy than the KNN. This result is in line with the findings from a recent study[23], which used a 3D accelerometer to classify only two free skating style techniques (gear 2, gear 3) and reported that the deep learning had the highest accuracy among all investigated classification models.
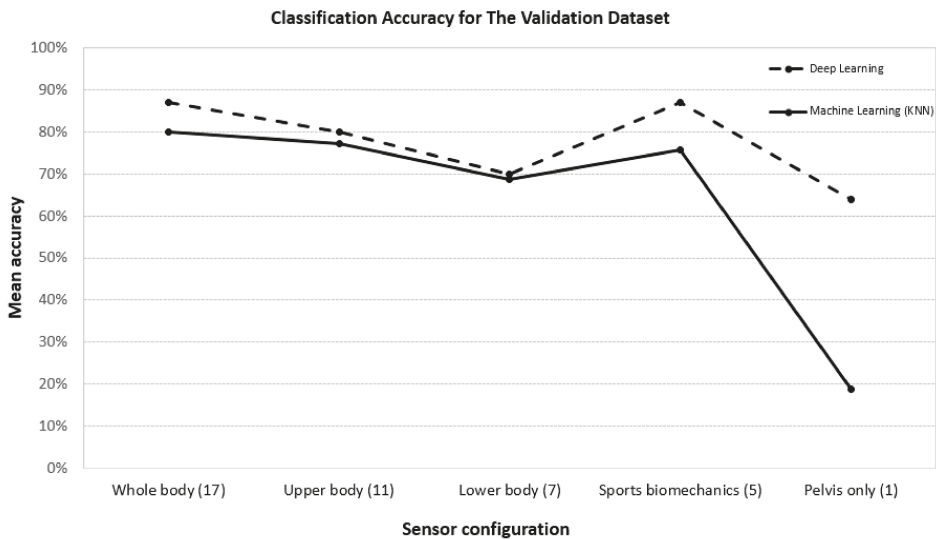
Figure 4. Comparison between the machine learning (KNN) and deep learning methods on classification accuracies for the validation dataset for the five different sensor configurations.

Table 22. Comparison between the machine learning (ML-KNN) and deep learning (DL) methods on classification accuracies for test set-1 and test set-2 for the five different sensor configurations.

Sensor Configuration	17 Sensors (Whole Body)		11 Sensors (Upper Body)		7 Sensors (Lower Body)		5 Sensors (Sports Biomech.)		1 Sensor (Pelvis Body)	
Modeling method	ML-KNN	DL	ML-KNN	DL	ML-KNN	DL	ML-KNN	DL	ML-KNN	DL
Test set-1	48.1%	84.2%	33.6%	84.8%	65.5%	84.4%	68.8%	87.2%	38.1%	58.5%
Test set-2	86.8%	92.2%	72.0%	84.8%	71.3%	64.6%	78.0%	95.1%	7.4%	29.7%

Even though the developed deep learning model achieved high overall classification accuracy for eight skiing techniques simultaneously, in-depth analysis of the confusion matrices showed that most incorrect classifications occurred for classical push-off and double poling techniques. The classical push-off and double poling techniques have identical motions of upper body and pelvis, the only differences between these two techniques are that a classical push-off begins with a slight jump for the propulsive force and the body movements are faster and exaggerated as compared to double poling. Such exceedingly similar physiological and biological characteristics are the cause for the confusion of the model and lead to misclassifications. In addition, some misclassifications occurred for V2A,

which is a typical technique used in level terrain up to moderate uphill inclines or during transitions between V2 and V1. In V2A skate, the timing sequence for pole push is the same as V2 skate, but it employs one double pole with every second skate, which is different from one double pole with every skate in V2 skate**.** V1 skate is an uphill technique, which employs an asymmetrical poling with every second skate [3,24]. Transitions between similar techniques, V2A and V2, V2A and V1 lead to high classifications errors on V2A. This finding is consistent with the result from a previous study [19].

In order to provide empirical evidence to researchers to base their future studies on the optimal sensor configuration for analysis of XC-skiing techniques, we compared classification accuracies among five different combinations of sensors on the training, validation, and test datasets. The five combinations include whole body with 17 sensors, upper body with 11 sensors, lower body with 7 sensors, sports biomechanics configuration with 5 sensors, and the pelvis configuration with 1 sensor only. Collective results (Tables 10, 20 and 22) show that the sports biomechanics configuration (both hands, both feet, and the pelvis sensors) can achieve a very similar accuracy as the whole body with 17 sensors. The classification accuracy from the sport biomechanics configuration is much higher than the accuracies from the pelvis, the upper, and the lower body sensors. A low classification accuracy for the pelvis sensor indicates that this sensor alone is not sufficient to capture the complex motions of all the body segments during the XC-skiing. Moderate, but not high, classification accuracies for the upper and lower body configuration of sensors are not surprising because in both the configurations, the data of only half of the body segments is available for training the model. In the sports biomechanics configuration, only five sensors, those on the hands, on the feet, and the pelvis, are used. Out of these five body segments, four body segments, both hands and feet, are at the extremes of the body where the motions of the segments are most exaggerated and vigorous, and the pelvis is close to the centre of the mass of the body, which represents an overall motion of the body segments. As the results while using all 17 sensors and five sensors in sports biomechanics configuration are very close to each other, we infer that the other 12 sensors are almost inconsequential and provide no additional information. Thus, the sports biomechanics configuration of sensors is the optimal set and future studies of XC-skiing classification can be based on the data obtained from this set with strong experimental proof.

Several previous studies have attempted to classify XC-skiing techniques by numerous hard rules or machine learning algorithms. Seeberg et al. [18] classified the classical XC-skiing techniques by deriving hard rules based on the data of 11 skiers and achieved an overall sensitivity of 99~100%. They, however, classified only the diagonal stride, double poling, and kick double poling techniques while leaving out push-off from the classification. Among the classical XC-skiing techniques, push-off and double poling are the only two techniques that are substantially misclassified by our algorithm, as is evident from the test set-1 confusion matrix in Table 18. One hundred and seventy-two (out of total 241) push-off techniques have been incorrectly classified as double poling and 74 (out of total 295) double poling as push-off. Moreover, they utilized six IMUs for classification of XC-skiing techniques and a total of 18 sensors were used, since each IMU contains one accelerometer, one gyroscope, and one magnetometer, whereas our model performs classification only with five gyroscope sensors. In addition, our model development does not require expert domain knowledge and a tedious process to derive the hard rules for classification. Rindal et al. [1] classified classical XC-skiing techniques by utilising two sensors, one accelerometer and one gyroscope, on the data of 10 participants and achieved an overall accuracy of 93.9% $\pm$ 3%. We achieved an overall mean accuracy of 91.1% by utilising data of four subjects and five gyroscope sensors. Our results rivals the results of [1] in terms of accuracy, but at an additional cost of three extra sensors. However, the data in [1] is a combination of data obtained from outdoor tracks and that obtained on a treadmill in the controlled environment of the laboratory whereas our data is obtained only from natural outdoor tracks. Additionally, they classified only the classical XC-skiing techniques whereas our model classifies both the classical as well as the skating techniques simultaneously. At the same time, our model shows considerable improvement in classification accuracy when the size of the training data is increased. Stoggl et al. [19] classified skating techniques by utilising a single accelerometer on the data of 11 skiers obtained on a

treadmill in the controlled environment of the laboratory, and achieved an accuracy of 86% ± 9% on the test set. As the accuracy achieved by our model is higher and our model has additional advantages in terms of performing classification of both classical and skating techniques simultaneously, we conclude that our model has higher potential of being deployed as a real time classification model for XC-skiing techniques.

Despite the inherent advantages in terms of automatic selection of features, high accuracy, and simultaneous classification of classical and skating XC-skiing techniques, this study suffers from certain limitations. First, due to practical constraints, such as the unavailability of the skiing tracks and tight training schedule of professional skiers, we only obtained the experimental data from a relatively small sample size (four professional skiers) for training, validating, and testing our models, further study should be carried out with a larger sample size for the verification of these results. Second, although the CNN-LSTM network promises high accuracy, the model is slow to train as compared to a model developed using a traditional machine learning algorithm due to the large size of training data that is fed to it. Traditional machine learning approaches rely on manually designed features, compact the raw data into a small number of features after pre-processing, and are much faster to train. Thus, there is a compromise between the time spent in data pre-processing in the case of traditional algorithms and training a deep learning model. However, by utilising computer systems with good software configurations, the deep learning model can be trained in a reasonable time to remain suitable for real time deployment. In this study, we assumed turning points in flat course data and descending and transition points in natural course data as noise and removed them manually by finding frames corresponding to them. These points, however, can be treated as dummy techniques and passed to the model, the study and analysis of which should be taken up as future research work. Last, but not the least, we utilized only the angular velocity due to clear cyclic patterns and for achieving a higher test time efficiency. The development of classification models using linear acceleration and magnetic fields is left as a future research work.

5. Conclusions

We utilized a novel CNN-LSTM based deep learning approach to develop a unified model for the classification of eight techniques used in classical and skating styles for XC-skiing. Overall, we achieved an accuracy of 87.2% and 95.1% on the flat and natural course test sets using the optimal sensor configuration (five gyroscope sensors: both hands, both feet, and the pelvis). High classification accuracy on both the test sets indicates that this deep learning based approach is very promising for automatic identification and classification of different XC-skiing techniques. The essence of our approach lies in eliminating the need of manually designed features required for traditional machine learning approaches and substituting the video-based and force measurement systems for classification of the XC-skiing techniques. Our model has the potential to be trained in the wild and as data of more skiers is made available, the fine tuning of the parameters will improve the accuracy as well as the scope of generalization continuously. This increases the practical value of our model and makes it suitable for real-time deployment by sports professionals. We optimized for the number of sensors and obtained the sports biomechanics configuration with five sensors as the optimal set, providing empirical evidence to researchers to base their future studies on this optimal configuration.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1

Tables A1–A4 show the confusion matrices for the test set of skier 2 when using the sports biomechanics configuration of sensors.

Table A1. Confusion matrix for the natural course, classical style validation set of skier 2 when using the sports biomechanics configuration of sensors.

Predicted / True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	190	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	7	0	0	0	0	0
DP	0	2	9	49	1	2	0	0
V2	0	0	0	0	0	0	0	0
V2A	0	0	0	0	0	0	0	0
V1	0	0	0	0	0	0	0	0
FS	0	0	0	0	0	0	0	0

Table A2. Confusion matrix for the natural course, skating style validation set of skier 2 when using the sports biomechanics configuration of sensors.

Predicted / True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	2	93	3	0	0
V2A	0	1	0	8	0	25	0	0
V1	0	1	1	0	4	0	593	3
FS	0	0	0	0	0	0	0	0

Table A3. Confusion matrix for the flat course, classical style validation set of skier 2 when using the sports biomechanics configuration of sensors.

Predicted / True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	48	1	0	0	0	0	0	0
P-Off	3	50	1	0	0	0	0	0
KDP	2	3	20	1	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	0	0	0	0	0
V2A	0	0	0	0	0	0	0	0
V1	0	0	0	0	0	0	0	0
FS	0	0	0	0	0	0	0	0

Table A4. Confusion matrix for the flat course, skating style validation set of skier 2 when using the sports biomechanics configuration of sensors.

	Predicted							
True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	0	10	2	1	0
V2A	0	0	0	1	2	20	0	1
V1	0	0	1	0	2	0	12	0
FS	0	0	0	0	0	1	0	9

Appendix A.2

Tables A5–A7 show the confusion matrices for the test set of skier 1 when using the sports biomechanics configuration of sensors.

Table A5. Confusion matrix for the natural course, skating style validation set of skier 1 when using the sports biomechanics configuration of sensors.

	Predicted							
True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	2	34	15	8	1
V2A	0	0	0	0	2	27	2	0
V1	0	0	0	0	5	30	158	4
FS	0	0	0	0	0	0	0	0

Table A6. Confusion matrix for the flat course, classical style validation set of skier 1 when using the sports biomechanics configuration of sensors.

	Predicted							
True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	30	1	2	0	0	0	0	0
P-Off	0	10	1	2	0	0	0	0
KDP	1	3	35	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	0	0	0	0	0
V2A	0	0	0	0	0	0	0	0
V1	0	0	0	0	0	0	0	0
FS	0	0	0	0	0	0	0	0

Table A7. Confusion matrix for the flat course, skating style validation set of skier 1 when using the sports biomechanics configuration of sensors.

	Predicted							
True	DS	P-Off	KDP	DP	V2	V2A	V1	FS
DS	0	0	0	0	0	0	0	0
P-Off	0	0	0	0	0	0	0	0
KDP	0	0	0	0	0	0	0	0
DP	0	0	0	0	0	0	0	0
V2	0	0	0	0	19	21	4	0
V2A	0	0	0	0	3	54	0	3
V1	0	0	0	0	0	0	7	0
FS	0	0	0	1	0	2	3	15

Appendix A.3

Table A8. Classification accuracies for the first three skiers using the proposed deep learning model when a leave-one-out type of testing is performed using four different sensor configurations of sensors.

Sensor Configuration	Course & Technique / Skier	Natural Course		Flat Course		Mean Accuracy
		Classical Style	Skating Style	Classical Style	Skating Style	
Whole body sensors (17)	Skier 1	X	89.4%	82.3%	76.5%	82.7%
	Skier 2	97.7%	96.2%	84.5%	88.7%	91.8%
	Skier 3	92.4%	61.4%	82.9%	54.8%	72.9%
	Mean Accuracy	95.0%	82.3%	83.2%	73.3%	~83.0%
Upper body sensors (11)	Skier 1	X	82.4%	67.6%	74.2%	74.7%
	Skier 2	73.5%	68.8%	76.5%	64.1%	70.7%
	Skier 3	66.2%	42.3%	78.5%	44.1%	57.8%
	Mean Accuracy	69.9%	64.5%	74.2%	60.8%	~68%
Lower body sensors (7)	Skier 1	X	24.2%	74.9%	66.3%	55.1%
	Skier 2	94.6%	24.2%	86.8%	67.7%	68.3%
	Skier 3	91.1%	68.7%	86.2%	50.5%	74.1%
	Mean Accuracy	92.9%	39.0%	82.6%	61.5%	~66%
Pelvis sensor only (1)	Skier 1	X	24.0%	77.6%	56.1%	52.6%
	Skier 2	64.6%	85.3%	51.1%	38.7%	59.9%
	Skier 3	62.4%	20.2%	82.1%	21.5%	46.5%
	Mean Accuracy	63.5%	43.1%	70.3%	38.8%	~53%

X: The classical style data on the natural course for skier 1 is not available.

Table A9. Classification accuracies for the first three skiers using the k-nearest neighbors algorithm when a leave-one-out type of testing is performed using four different sensor configurations of sensors.

Sensor Configuration	Course & Technique / Skier	Natural Course		Flat Course		Mean Accuracy
		Classical Style	Skating Style	Classical Style	Skating Style	
Whole body sensors (17)	Skier 1	X	89.1%	65.2%	65.4%	73.2%
	Skier 2	69.5%	82.6%	60.8%	78.7%	72.9%
	Skier 3	81.4%	74.4%	70.2%	56.3%	70.6%
	Mean Accuracy	75.5%	82.0%	65.4%	66.8%	~72%
Upper body sensors (11)	Skier 1	X	50.3%	46.4%	61.1%	52.6%
	Skier 2	66.8%	51.7%	61.3%	58.7%	59.6%
	Skier 3	64.4%	73.3%	49.6%	55.3%	60.7%
	Mean Accuracy	65.6%	58.4%	52.4%	58.4%	~59%
Lower body sensors (7)	Skier 1	X	81.3%	56.5%	50.7%	62.8%
	Skier 2	49.8%	51.7%	48.8%	38.7%	47.3%
	Skier 3	81.1%	38.9%	62.6%	36.9%	54.9%
	Mean Accuracy	65.4%	57.3%	56.0%	42.1%	~55%
Pelvis sensor only (1)	Skier 1	X	2.4%	30.4%	2.1%	11.7%
	Skier 2	21.8%	10.9%	15.7%	9.3%	14.5%
	Skier 3	40.1%	2.9%	24.4%	7.8%	18.8%
	Mean Accuracy	30.9%	5.4%	23.5%	6.4%	~15%

X: The classical style data on the natural course for skier 1 is not available.

References

1. Rindal, O.M.H.; Seeberg, T.M.; Tjønnås, J.; Haugnes, P.; Sandbakk, Ø. Automatic classification of sub-techniques in classical cross-country skiing using a machine learning algorithm on micro-sensor data. *Sensors* **2017**, *18*, 75. [CrossRef] [PubMed]

2. Andersson, E.; Pellegrini, B.; Sandbakk, O.; Stuggl, T.; Holmberg, H.C. The effects of skiing velocity on mechanical aspects of diagonal cross-country skiing. *Sports Biomech.* **2014**, *13*, 267–284. [CrossRef] [PubMed]

3. Bilodeau, B.; Boulay, M.R.; Roy, B. Propulsive and gliding phases in four cross-country skiing techniques. *Med. Sci. Sports Exerc.* **1992**, *24*, 917–925. [CrossRef] [PubMed]

4. Bilodeau, B.; Rundell, K.W.; Roy, B.; Boulay, M.R. Kinematics of cross-country ski racing. *Med. Sci. Sports Exerc.* **1996**, *28*, 128–138. [CrossRef] [PubMed]

5. Lindinger, S.J.; Gopfert, C.; Stoggl, T.; Muller, E.; Holmberg, H.C. Biomechanical pole and leg characteristics during uphill diagonal roller skiing. *Sports Biomech.* **2009**, *8*, 318–333. [CrossRef] [PubMed]

6. Nilsson, J.; Tveit, P.; Eikrehagen, O. Effects of speed on temporal patterns in classical style and freestyle cross-country skiing. *Sports Biomech.* **2004**, *3*, 85–107. [CrossRef] [PubMed]

7. Ohtonen, O.; Lindinger, S.; Lemmettylä, T.; Seppälä, S.; Linnamo, V. Validation of porTable 2d force binding systems for cross-country skiing. *Sports Eng.* **2013**, *16*, 281–296. [CrossRef]

8. Stoggl, T.; Kampel, W.; Muller, E.; Lindinger, S. Double-push skating versus v2 and v1 skating on uphill terrain in cross-country skiing. *Med. Sci. Sports Exerc.* **2010**, *42*, 187–196. [CrossRef] [PubMed]

9. Stoggl, T.; Muller, E.; Lindinger, S. Biomechanical comparison of the double-push technique and the conventional skate skiing technique in cross-country sprint skiing. *J. Sports Sci.* **2008**, *26*, 1225–1233. [CrossRef] [PubMed]

10. Fasel, B.; Favre, J.; Chardonnens, J.; Gremion, G.; Aminian, K. An inertial sensor-based system for spatio-temporal analysis in classic cross-country skiing diagonal technique. *J. Biomech.* **2015**, *48*, 3199–3205. [CrossRef] [PubMed]

11. Marsland, F.; Lyons, K.; Anson, J.; Waddington, G.; Macintosh, C.; Chapman, D. Identification of cross-country skiing movement patterns using micro-sensors. *Sensors* **2012**, *12*, 5047–5066. [CrossRef] [PubMed]

12. Smith, G.A. Biomechanics of crosscountry skiing. *Sports Med.* **1990**, *9*, 273–285. [CrossRef] [PubMed]

13. Marsland, F.; Mackintosh, C.; Holmberg, H.C.; Anson, J.; Waddington, G.; Lyons, K.; Chapman, D. Full course macro-kinematic analysis of a 10 km classical cross-country skiing competition. *PLoS ONE* **2017**, *12*, e0182262. [CrossRef] [PubMed]

14. Marsland, F.; Mackintosh, C.; Anson, J.; Lyons, K.; Waddington, G.; Chapman, D.W. Using micro-sensor data to quantify macro kinematics of classical cross-country skiing during on-snow training. *Sports Biomech.* **2015**, *14*, 435–447. [CrossRef] [PubMed]

15. Myklebust, H.; Nunes, N.; Hallén, J.; Gamboa, H. Morphological analysis of acceleration signals in cross-country skiing. In Proceedings of the Biosignals-International Conference on Bio-Inspired Systems and Signal Processing (BIOSTEC 2011), Rome, Italy, 26–29 January 2011; pp. 26–29.

16. Sakurai, Y.; Fujita, Z.; Ishige, Y. Automated identification and evaluation of subtechniques in classical-style roller skiing. *J. Sports Sci. Med.* **2014**, *13*, 651–657. [PubMed]

17. Sakurai, Y.; Fujita, Z.; Ishige, Y. Automatic identification of subtechniques in skating-style roller skiing using inertial sensors. *Sensors (Basel)* **2016**, *16*, 473. [CrossRef] [PubMed]

18. Seeberg, T.M.; Tjonnas, J.; Rindal, O.M.H.; Haugnes, P.; Dalgard, S.; Sandbakk, O. A multi-sensor system for automatic analysis of classical cross-country skiing techniques. *Sports Eng.* **2017**, *20*, 313–327. [CrossRef]

19. Stoggl, T.; Holst, A.; Jonasson, A.; Andersson, E.; Wunsch, T.; Norstrom, C.; Holmberg, H.C. Automatic classification of the sub-techniques (gears) used in cross-country ski skating employing a mobile phone. *Sensors (Basel)* **2014**, *14*, 20589–20601. [CrossRef] [PubMed]

20. Ristner, M. A Comparison of a k-Nearest Neighbour and a Markov Model for Classification of Gears in Cross-Country Skiing. Available online: https://www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2015/sammanf15/ristner_moa_eng.html (accessed on 4 October 2018).

21. Holst, A.; Jonasson, A. *Classification of Movement Patterns in Skiing*; SCAI: Aalborg, Denmark, 2013; pp. 115–124.

22. Yao, S.; Hu, S.; Zhao, Y.; Zhang, A.; Abdelzaher, T. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland; pp. 351–360.
23. Rassem, A.; El-Beltagy, M.; Saleh, M. Cross-country skiing gears classification using deep learning. *arXiv* **2017**, arXiv:1706.08924.
24. Andersson, E.; Supej, M.; Sandbakk, O.; Sperlich, B.; Stoggl, T.; Holmberg, H.C. Analysis of sprint cross-country skiing using a differential global navigation satellite system. *Eur. J. Appl. Physiol.* **2010**, *110*, 585–595. [CrossRef] [PubMed]

A Batteryless, Wireless Strain Sensor Using Resonant Frequency Modulation

Kyeong Jae Lee [1], Namsun Chou [2] and Sohee Kim [1,*]

[1] Department of Robotics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Korea; kj.lee@dgist.ac.kr
[2] Center for BioMicroSystems, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea; nschou@kist.re.kr
*   Correspondence: soheekim@dgist.ac.kr; Tel.: +82-53-785-6217

Abstract: In this study, we demonstrated the feasibility of a wireless strain sensor using resonant frequency modulation through tensile impedance test and wireless sensing test. To achieve a high stretchability, the sensor was fabricated by embedding a copper wire with high conductivity in a silicone rubber with high stretchability, in which the resonant frequency can be modulated according to changes in strain. The characteristics of the sensor and the behavior of wireless sensing were calculated based on equations and simulated using finite element method. As the strain of the sensor increased, the inductance increased, resulting in the modulation of resonant frequency. In experimental measurement, as the strain of the sensor increased from 0% to 110%, its inductance was increased from 192 nH to 220 nH, changed by 14.5%, and the resonant frequency was shifted from 13.56 MHz to 12.72 MHz, decreased by 6.2%. It was demonstrated that using the proposed sensor, strains up to 110% could be detected wirelessly up to a few centimeters.

Keywords: batteryless strain sensor; wireless strain sensor; resonant frequency modulation; Ecoflex

1. Introduction

Stomach cancer is one of the common cancers in Northeast Asian male population [1,2]. The general treatment is gastrectomy, which removes the stomach or the small intestine partially or totally [3–5]. After surgery, the stomach motility is observed to check normal gastric emptying functions using several clinical methods such as gastric emptying time measurement, ultrasound, scintigraphy, and gastroscopy [6–8]. Recently, methods such as electrogastrography (EGG) and magnetogastrography (MGG) have been introduced to quantitatively measure the stomach motility. However, in case of EGG, the electrodes must be attached onto the operated skin of the patient [9]. In case of MGG, it can be measured only in the place where the huge facility such as Superconducting Quantum Interference Device (SQUID) is located [10]. Both methods are not adequate to continuously measure the stomach condition right after the surgery.

As a potential alternative to those methods, we propose a sensor that can be attached to the stomach directly and detect the stomach deformation wirelessly. Figure 1 illustrates the schematic of monitoring the stomach motility using the proposed sensor. The change in resonant frequency of the sensor could be measure wirelessly by a transceiver with a sufficiently large diameter to minimize the misalignment effect. The wireless strain sensor is attached to the stomach during surgery by stitching, for instance. When the stomach is active, the sensor deforms. The applied strain to the sensor causes the change in inductance, thereby shifting the resonant frequency of the sensor. The transceiver located outside the body senses the change in resonant frequency wirelessly by inductive coupling, and therefore the stomach deformation can be detected.
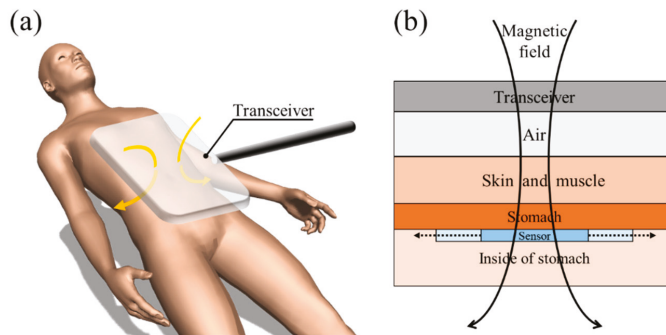
Figure 1. Schematic of the proposed sensor in a wireless sensing application. (**a**) Schematic illustration of the transceiver covering the entire abdomen to wirelessly measure the strain of the sensor attached to the stomach and (**b**) its cross-sectional view.

Several studies have reported sensors that demonstrated the feasibility of measuring the strain wirelessly. Kim et al. developed an LC tag-based RFID tag using silver nano-ink [11]. It operates in the 1.5 to 1.6 GHz frequency range, showing that as the tensile strain increases, the resonant frequency decreases within 4% of strain change. Song et al. developed a microstrip patch antenna made of silver nanowires and polydimethylsiloxane (PDMS), which operates at 2.9 to 3.1 GHz and the resonant frequency increases linearly with increasing strain within 15% of change [12]. Lazarus et al. proved that the modulation of resonant frequency caused by changes in strain can be measured wirelessly using a single-turn coil made of galistan and silicone rubber, which operates at frequencies of 0.97 to 1.29 MHz within 80% of strain change [13].

Unlike these previously developed sensors, the proposed sensor aims to eventually measure the stomach motility, which can stretch up to 100% and more [14]. To develop such a highly stretchable and wirelessly functioning strain sensor, the conducting patterns must have a low DC resistance enough to operate as a coil and at the same time, should have sufficient stretchability [15]. These two requirements are however contradictory, as outlined in many previous studies that the high stretchability can only be achieved by employing very thin conducting lines on very thin substrates [16–18]. Thus, we developed the wireless strain sensor that is patterned with a copper wire in a serpentine shape, based on a flexible substrate to achieve the higher stretchability that can cover the range of stomach motility as well as the wireless functionality.

2. Materials and Methods

2.1. Fabrication of the Sensor

Figure 2 illustrates the fabrication process of the sensor. In the whole process, no specific semiconductor fabrication equipment was required. Ecoflex was chosen as the substrate material since it is flexible, highly stretchable and biocompatible [19–21]. Ecoflex mixture of part A and part B (Ecoflex 00-30, Smooth-On Inc., Easton, PA, USA) was prepared at a ratio of 1:1. The 15 g of mixture was poured into a petri dish to make a 1 mm thick substrate and left for 10 min at room temperature to remove air bubbles and make the surface even. The substrate was then partially cured in an oven at 60 °C for 3 min. Partially cured substrates enabled copper wires to be easily embedded in them (Figure 2a). Prior to the transfer step, the wire pattern was prepared using a custom-made winding mold. A copper wire with a diameter of 160 μm was wounded manually on the pillars printed by a 3D printer (Projet 3500, 3D systems, Rock Hill, SC, USA) to make a serpentine pattern (Figure 2b). The coil pattern was transferred to the top of the partially cured substrate and then 15 g of silicone rubber mixture was poured again onto the substrate with the serpentine pattern transferred (Figure 2c). It was then cured completely in the 60 °C oven for 10 min. The resulted coil had a single turn in a serpentine

pattern embedded in silicone rubber in a thickness of 2 mm (Figure 2d). The outer dimension of the fabricated coil was 35 mm × 35 mm, the radius and angle of a single arc were 1.57 mm and 216°, respectively, taken from the previous study to maximize the stretchability [22] (Figure 2e). Figure 3 shows the fabricated sensor in non-stretched and stretched states. When the sensor was stretched by 100% in one direction, the coil pattern was stretched by 63% in the same direction and shrunk by 30% in the orthogonal direction. The resistance and the inductance of the sensor was 1.45 ± 0.04 Ω and 192 ± 3 nH, respectively. The Young's modulus of the sensor was measured to be 40.5 ± 1.1 kPa.
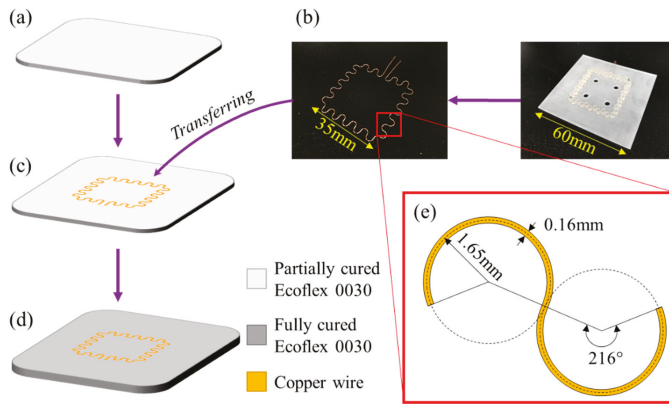
Figure 2. Fabrication process of the wireless strain sensor: (**a**) curing the substrate partially, (**b**) patterning the copper wire in serpentine shape, (**c**) transferring the pattern and (**d**) curing the device completely. (**e**) Detailed design of the serpentine pattern of the coil.
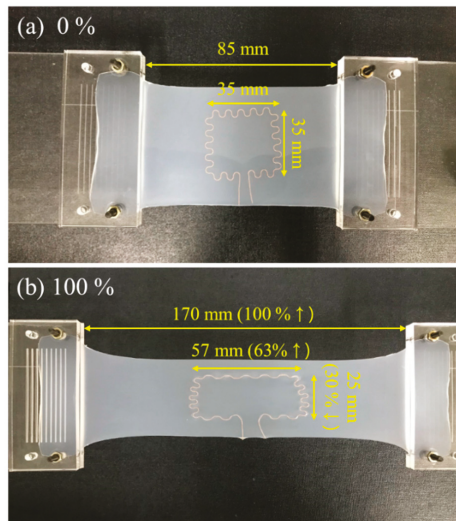
Figure 3. Photographs of the wireless strain sensor in (**a**) non-stretched and (**b**) 100% stretched states.

To make the sensor resonate at a specific frequency, additional capacitors were connected in parallel to the sensor. Ceramic capacitors were used as they have a low parasitic inductance [23]. 13.56 MHz was selected as the resonating frequency of the sensor since it belongs to the industrial, scientific and medical (ISM) bands. The capacitance value to tune the sensor to resonate at 13. 56 MHz was 708 pF on average.

2.2. Calculation of Sensor Inductance

The behavior of the proposed sensor was first predicted by simulations based on calculation. Due to the practical limitation that finite element method (FEM) simulation of electromagnetic fields is time consuming, we simplified the serpentine pattern as a straight line, so the shape change of the serpentine pattern could be simplified to the change in aspect ratio of a rectangle as illustrated in Figure 4. The inductance of the simplified model was calculated as the sum of self-inductances and mutual inductances of all line segments, based on the Greenhouse method [24]. In the Greenhouse method, the inductance is expressed as the sum of self-inductances and mutual inductances of each segments. The inductance L is given by

$$L = \Sigma L_s + \Sigma M_+ - \Sigma M_- \tag{1}$$

where L_s is the self-inductance of each segment, and M_+ and M_- are the positive and negative mutual inductance between segments, respectively. Since there is no segment combination in which current flows in the same direction, the sum of the positive mutual inductance is zero. The four segments used in the calculation are illustrated in Figure 4: left, right, top, and bottom segment.
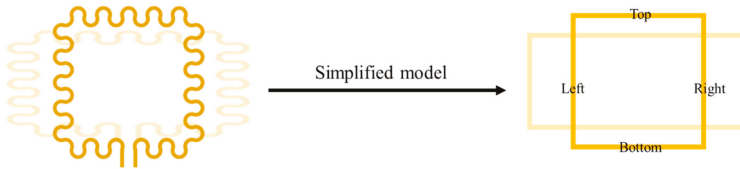
Figure 4. Simplification of the serpentine sensor coil as a rectangle to calculate the inductance.

Based on Equation (1), the inductance of the 1-turn rectangular coil has only two components of self-inductance L_s and negative mutual inductance M_-, and it can be calculated with the following equation:

$$L = L_{s,\ Bottom} + L_{s,Left} + L_{s,Top} + L_{s,Right} + M_{-,\ Top,Bottom} + M_{-,Left,Right} \tag{2}$$

where $L_{s,\ Bottom}$, $L_{s,Left}$, $L_{s,\ Top}$, and $L_{s,Right}$ are the self-inductances of the bottom, left, top and right segment, respectively, and $M_{-,\ Top,Bottom}$ is the negative mutual inductance between the top and bottom segments, and $M_{-,Left,Right}$ is the negative mutual inductance between the left and right segments. L_s and M_- can be calculated by the following equations:

$$L_s = 0.002l \times \ln\left(\frac{2l}{l_1 + l_2} + 0.50049 + \frac{l_1 + l_2}{3l}\right) \tag{3}$$

$$M_- = 2l\left[\ln\left\{\frac{l}{d} + \left(1 + \frac{l^2}{d}\right)^{\frac{1}{2}}\right\} - \left(1 + \frac{d^2}{l^2}\right)^{\frac{1}{2}} + \left(\frac{d}{l}\right)\right] \tag{4}$$

where l is the length of the selected segment, l_1 is the horizontal length of the coil, l_2 is the vertical length of the coil, and d is the distance between the segment centers. The relationship between the length of the coil segments and the strain applied to the sensor is based on the measurement that when the sensor was stretched 100%, the length of the top and bottom segments were increased by 63% and the length of the left and right segments were decreased by 30% as illustrated in Figure 3. Based on this, l_1 and l_2 of the coil when a strain ε is applied to the whole sensor are expressed as below:

$$l_1 = 1.63\varepsilon l_{1,0\%} \tag{5}$$

$$l_2 = 0.7\varepsilon l_{2,0\%} \tag{6}$$

where $l_{1,0\%}$ is the horizontal length at 0% strain and $l_{2,0\%}$ is the vertical length at 0% strain. Thus, the inductance of the sensor when a strain is applied can be calculated using Equation (2), substituted by Equations (3) to (6). This calculation was implemented in MATLAB (MATLAB 2016b, Mathworks Inc., Natick, MA, USA). Based on this method, the relationship between the inductance and strain could be obtained by linear fitting based on the calculated inductance values, expressed by

$$L \approx L_0 + \alpha_0 \varepsilon \tag{7}$$

where L is the inductance of the sensor, α_0 is the stretching coefficient, L_0 is the non-stretched inductance, and ε is the applied strain. As the sensor consisted of an LC circuit to resonate at a specific frequency, the resonant frequency f of which is calculated as:

$$f \approx \frac{1}{2\pi\sqrt{(L_0 + \alpha_0\varepsilon)C}} \tag{8}$$

where C is the additional tuning capacitance.

2.3. Finite Element Simulation of Wireless Sensing

Before measuring the maximum distance that the sensor could operate wirelessly, finite element simulation using an electromagnetic simulator ANSYS HFSS (ANSYS Inc., Canonsburg, PA, USA) was performed to estimate the wireless working distance. Although the inductance of the sensor can be relatively easily calculated based on equations, it is not easy to mathematically derive the coupling coefficient between the sensor and the transceiver and thus, we simulated the wireless sensing using FEM. The simulation model is illustrated in Figure 5. The simplified geometry of the sensor was used, same as the model used in inductance calculation. Both the transceiver and the sensor were set to resonate at 13.56 MHz. The sensitivity of frequency shifting was defined as the gauge factor as follows:

$$\text{Gauge Factor} = \frac{\Delta f / f_0}{\varepsilon} \tag{9}$$

where f_0 is the resonant frequency at 0% strain, Δf is the difference between f_0 and the measured resonant frequency, and ε is the applied strain.
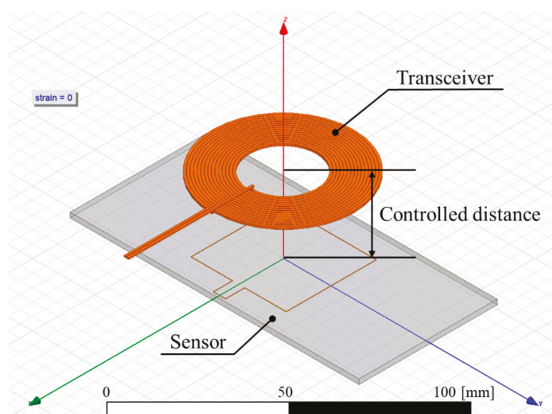
Figure 5. Simulation model of the sensor and the transceiver for wireless sensing. The distance between the sensor and the transceiver was changed from 10 mm to 50 mm, and the strain applied to the sensor was changed from 0% to 100%. In this figure, the transceiver is 20 mm away from the non-stretched sensor.

2.4. Experimental Setup

To verify whether the change in sensor strain results in resonant frequency shifting, the experimental test bench was constructed as shown in Figure 6a. An impedance analyzer (4294A, Agilent, Santa Clara, CA, USA) was directly connected to the sensor to measure the impedance, and the strain applied to the sensor was controlled by an XY stage (CROSS 130-HSM, OWIS®, Staufen, Germany). Since the applied strain and the impedance should be measured at the same time, a LabVIEW-based program (LabVIEW 2016, National Instrument, Austin, TX, USA) was used to control the XY stage and the impedance analyzer simultaneously. Because the available travel range of the XY stage was limited, the sensor was stretched up to 110% at maximum. From each impedance measurement, the resistance, inductance, and capacitance of the sensor were evaluated using the equivalent circuit model, in which the resistance and inductance are connected in series and capacitance connected to those in parallel [25]. Figure 6b shows the experimental setup to observe whether the change in strain of the sensor can be measured wirelessly and to investigate the maximum distance of wireless detection. The deformation of the sensor was measured wirelessly using a transceiver, which was a circular planar coil with 15 turns, an outer diameter of 65 mm and a wire diameter of 1 mm, resulting in a self-inductance of 10 μH and the quality factor of 71. The impedance of the transceiver was measured as the strain applied to the sensor varied from 0% to 100%. The distance between the sensor and the transceiver was varied from 10 mm to 50 mm to observe the maximum distance for wireless sensing. The centers of the sensor and the transceiver were aligned manually at every measurement step. Since the purpose of the wireless sensing setup was to observe the modulation of resonant frequency at the transceiver with varying the vertical distance, the influence of coil misalignment was excluded.
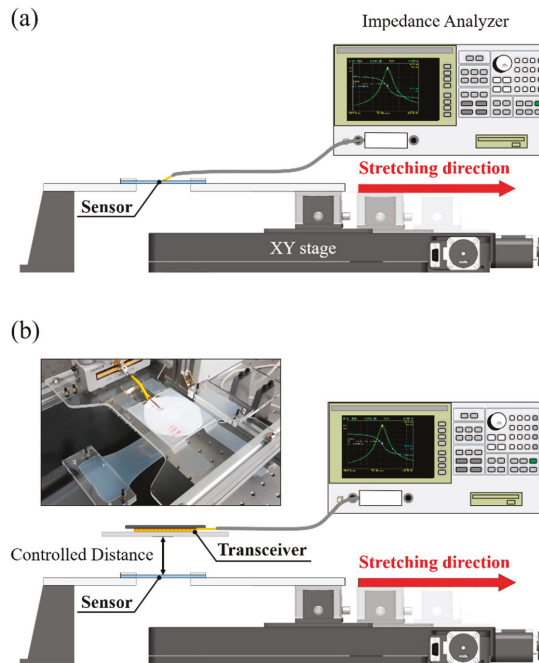
Figure 6. Experimental setup configurations for (**a**) tensile impedance testing and (**b**) wireless sensing.

3. Results

3.1. Tensile Impedance Test

The behavior of the proposed sensor was predicted by simulations based on calculation, and compared to measurements. As shown in Figure 7a, the calculated and measured results show that the inductance increased as the strain increased. The calculated inductance increased by 16.6% at 110% strain, and the measured inductance increased by 14.5%. The measured values were slightly higher than the calculated values, which was speculated to be caused by the parasitic inductances due to the serpentine structure and the additional measuring leads connected to the impedance analyzer. In Figure 7b, the resonant frequencies according to different strains are presented by using the calculated inductance, and compared with measurements. The calculated resonant frequency decreased by 6.7% with the change in strain of 110%, and the measured resonant frequency decreased by 6.2%. Both the measured and calculated resonant frequency decreased at a similar rate, as the inductance increased at a similar rate in both the measured and calculated cases. Figure 7c shows the quality factor (Q) of the sensor as a function of strain. The quality factor of the sensor was measured to be 10.9 ± 0.3 at 0% strain to 11.6 ± 0.3 at 110% strain, increased by 6.3%. For comparison with other studies, we calculated the Q of the wireless strain sensors reported in [12,13] based on the provided values such as resonant frequency, bandwidth, resistance, and inductance as those studies did not provide the Q explicitly. The fabricated sensor exhibited the Q similar to those of any stretchable strain sensors reported in the literature, up to 15% strain [12,13], but maintained a much higher Q for strains greater than 15%. This is thanks to the fabrication method of embedding the copper wire in thin silicone rubber sheet.
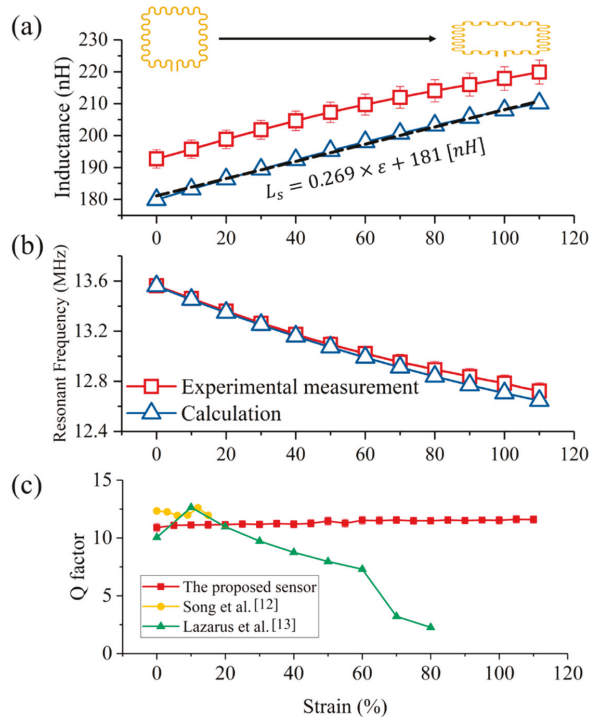
Figure 7. (**a**) Inductance change and (**b**) shift in resonant frequency of the sensor according to changes in strain, calculated using the Greenhouse method and measured experimentally. (**c**) Quality factor of the sensor as a function of strain, and compared with other sensors reported in previous studies [12,13].

3.2. Wireless Sensing Test

Next, we tested the wireless strain sensing, and compared with simulations using electromagnetic simulator. Figure 8 compares the wireless sensing results from the measurement and simulation when the transceiver was 10 mm away from the sensor. Among the two peaks, the right peak was chosen as the resonant frequency of the transceiver, because the changes in frequency of the right peak showed a higher sensitivity than that of the left peak. The measured and simulated results show that the resonant frequency decreased as the sensor strain increased. Overall, the simulated impedance magnitudes were higher than the measured values. It appears to be due to the relatively high DC resistance of the sensor in measurements, caused by the serpentine shape, which however was not included in the simplified simulation model. Nevertheless, the degree of changes in the peak frequency of the simulation was similar to that of the measured results, so that the simulation results could be used to predict the maximum wireless working distance of the sensor.
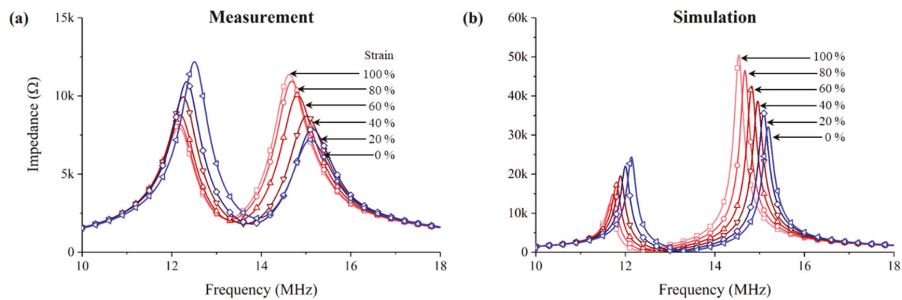
Figure 8. Experimentally measured and simulated impedance magnitudes of the transceiver at 10 mm away from the sensor. The frequency of the right peak was shifted (**a**) from 15.01 MHz at 0% strain to 14.59 MHz at 100% strain in experiment, and (**b**) from 15.31 MHz at 0% strain to 14.47 MHz at 100% strain in simulation.

The shifted resonant frequencies and the gauge factors of strain sensing according to different strains and different distances are summarized in Figure 9. Figure 9a shows the measured resonant frequency modulation detected at the transceiver wirelessly, and Figure 9c shows the calculated gauge factors. As the strain of the sensor increased, the resonant frequency of the transceiver decreased gradually. This decreasing trend could be seen only up to 22.5 mm distance. When measured at farther distances, the frequency converged to the resonant frequency of the transceiver itself, indicating that no measurable changes were detected. At 100% strain, the gauge factor decreased from 0.028 to 0.011 at 10 mm to 22.5 mm distance, converging to 0 at farther distances. Figure 9b shows the simulated resonant frequency at the transceiver, and the calculated gauge factors are shown in Figure 9d. The simulated results also showed that as the distance increased, the frequency converged to the self-resonant frequency of the transceiver. The detectable distance was observed to be between 20 mm and 30 mm, in good agreement with the experimental results. However, unlike the measured results, the simulated results did not show that the resonant frequency first increased at strains up to 20% at distances greater than 20 mm. This appears to be because the simplified simulation model did not include the additional DC resistance caused by the serpentine pattern and therefore did not fully reflect the decrease of inductive coupling. Nonetheless, the simulation results showed that the used FEM model could predict the wirelessly measurable distance reasonably well.
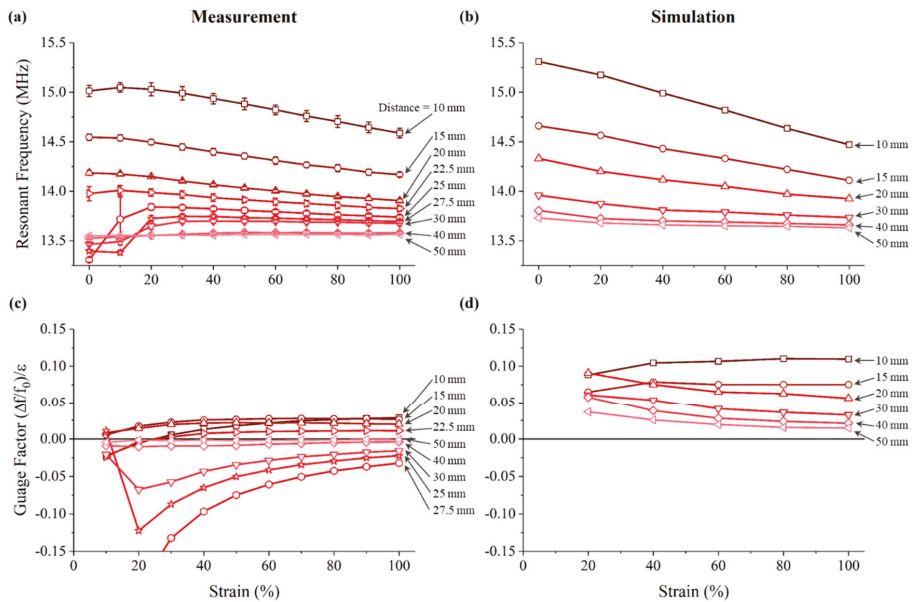
Figure 9. Experimentally measured and simulated results of wireless strain sensing. (**a**) Measured and (**b**) simulated resonant frequency modulation as the strain applied to the sensor increased, and the gauge factor of the transceiver calculated based on (**c**) the measured and (**d**) simulated resonant frequency.

4. Discussion

To measure the motility of internal organs such as the stomach, a stretchable, batteryless and wireless strain sensor using resonant frequency modulation was investigated. This sensor can stretch up to 110% from its original length, shifting the resonant frequency by 6.2% at maximum around 13.56 MHz, which exceeds the strain ranges that have been measured wirelessly in previous studies [11–13]. Previous studies have shown that up to 80% strain can be measured wirelessly. However, to measure the stomach motility, the sensor must be able to detect a stretch of up to more than 100%, and thus, we have developed a sensor that can wirelessly measure strains up to 110%. The conductivity of the copper wire used in the present study (5.96×10^5 S/cm) [26] is higher than the conductivity of the materials used in previous studies, by about 9 to 4000 times (silver nano-ink: 1.5×10^2 S/cm [11], silver nanowire: 8.13×10^3 S/cm [12], and eGaIn: 3.4×10^4 S/cm [27]). In addition, since the Q increases as the coil resistance decreases [15], the sensor made of a copper wire developed in the present study has the advantage of maintaining a high Q for a wider range of strains than those reported in previous studies. This high Q of the proposed sensor was obtained through embedding the copper wire in silicone rubber sheet, unlike the other commonly employed methods to fabricate stretchable strain sensors based on thin metal films on thin substrates.

The sensor could detect the strain wirelessly at up to 22.5 mm distance by the used transceiver, although there is room for further improvement to increase the operating distance. Thus, our proposed sensor demonstrated that it satisfied the two contradictory requirements for wireless sensing of strains, namely, the high conductivity of the coil sensor and the high stretchability. Moreover, the maximum strain of the developed sensor was not inherently limited, but instead, originated from the limitation of the used experimental setup. One additional point to note is that we only investigated the resonant frequency modulation in response to strains in only one direction. For actual applications, however, bi-axial strains and bending would need to be applied. Previous studies show that there is a correlation

between the diameter of the coil pattern and the wireless sensing performance due to bending. The wireless sensing performance is less affected by the degree of bending of the sensor when the diameter of the coil sensor is smaller [13,28,29]. Thus, to minimize the bending effect caused by the curved surface of the stomach, the outer length of the coil pattern would need to be minimized. Also, to increase the distance of wireless sensing and the sensitivity, parameter optimizations would be needed, which include the reduction of the sensor thickness, the increase in the number of turns of the sensor to increase the degree of inductive coupling between the sensor and the transceiver, and the increase of the outer diameter of the transceiver.

To apply the developed sensor in a real application, misalignment between the sensor and the transceiver should be overcome. To minimize the misalignment effect, the transceiver is suggested to be large enough to cover the entire abdomen as illustrated in Figure 1a. By creating a magnetic field as uniformly as possible throughout the abdomen, the sensor can keep relatively constant coupling with the external transceiver even if the exact location of the sensor is unknown. As shown in Figure 1b, the distance between the sensor and the transceiver can be determined by knowing the distance from the abdominal surface to the sensor, which could be measured by a physician during the surgery or by a medical imaging method. Therefore, it would be possible to detect the deformation of internal organs such as the stomach. To minimize the variations in stomach conditions and posture, the patient would be asked to stay with a constant posture, e.g., laying down on a bed, during the detection. Also, checking the stomach motility would be performed a few days after the gastrectomy, in which duration the food intake is generally restricted and thus, the stomach condition is expected to be quite constant. However, all such aspects need further investigation including animal experiments.

References

1. Jung, K.-W.; Won, Y.-J.; Kong, H.-J.; Oh, C.-M.; Cho, H.; Lee, D.H.; Lee, K.H. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2012. *Cancer Res. Treat.* **2015**, *47*, 127–141. [CrossRef] [PubMed]

2. Chen, W.; Zheng, R.; Baade, P.D.; Zhang, S.; Zeng, H.; Bray, F.; Jemal, A.; Yu, X.Q.; He, J. Cancer Statistics in China. *CA Cancer J. Clin.* **2016**, *66*, 115–132. [CrossRef] [PubMed]

3. Mochiki, E.; Kamiyama, Y.; Aihara, R.; Nakabayashi, T.; Asao, T.; Kuwano, H. Laparoscopic assisted distal gastrectomy for early gastric cancer: Five years' experience. *Surgery* **2005**, *137*, 317–322. [CrossRef] [PubMed]

4. Choi, S.H.; Noh, S.H.; Min, J.S.; Lee, K.S.; Kim, C.K. Clinical Analysis According to Reconstructive Type after Total Gastrectomy for Gastric Cancer. *Ann. Surg. Treat. Res.* **1991**, *41*, 734–743.

5. Noh, S.H.; Yoo, C.H.; Kim, Y.; Kim, C.B.; Min, J.S.; Lee, K.S. Results after a Gastrectomy of 2,603 Patients with Gastric Cancer: Analysis of Survival Rate and Prognostic Factor. *Ann. Surg. Treat. Res.* **1998**, *55*, 206–213.

6. Jung, H.J.; Kim, D.H.; Kim, D.H. Proximal Gastrectomy with Double Tract Reconstruction Using the Remnant Antrum in Early Upper Gastric Cancer. *Ann. Surg. Treat. Res.* **2008**, *74*, 261–266.

7. Jeong, I.-U.; Song, Y.-J.; Yun, H.-Y. Gastric-Emptying Patterns after Gastroduodenal Reconstruction. *Ann. Surg. Treat. Res.* **2000**, *59*, 46–53.

8. Kim, W.; Jeon, H.M.; Hur, H.; Lee, J.H.; Won, J.M. Jejunal Pouch Interposition (JPI) after Distal Gastrectomy. *J. Korean Gastric Cancer Assoc.* **2004**, *4*, 242–251. [CrossRef]

9. Komorowski, D.; Pietraszek, S.; Tkacz, E.; Provaznik, I. The extraction of the new components from electrogastrogram (EGG), using both adaptive filtering and electrocardiographic (ECG) derived respiration signal. *Biomed. Eng. Online* **2015**, *14*. [CrossRef] [PubMed]

10. Estombelo-Montesco, C.A.; de Araujo, D.B.; Roque, A.C.; Moraes, E.R.; Barros, A.K.; Wakai, R.T.; Baffa, O.; Davies, M.E.; James, C.J.; Abdallah, S.A.; et al. Extraction of gastric electrical response activity from magnetogastrographic recordings by DCA. *Indep. Compon. Anal. Signal Sep. Proc.* **2007**, *4666*, 585–592.

11. Kim, J.; Wang, Z.; Kim, W.S. Stretchable RFID for wireless strain sensing with silver nano ink. *IEEE Sens. J.* **2014**, *14*, 4395–4401. [CrossRef]

12. Song, L.; Myers, A.C.; Adams, J.J.; Zhu, Y. Stretchable and reversibly deformable radio frequency antennas based on silver nanowires. *ACS Appl. Mater. Interfaces* **2014**, *6*, 4248–4253. [CrossRef] [PubMed]

13. Lazarus, N.; Meyer, C.D.; Turner, W.J. A microfluidic wireless power system. *RSC Adv.* **2015**, *5*, 78695–78700. [CrossRef]

14. Egorov, V.I.; Schastlivtsev, I.V.; Prut, E.V.; Baranov, A.O.; Turusov, R.A. Mechanical properties of the human gastrointestinal tract. *J. Biomech.* **2002**, *35*, 1417–1425. [CrossRef]

15. Neagu, C.R.; Jansen, H.V.; Smith, A.; Gardeniers, J.G.E.; Elwenspoek, M.C. Characterization of a planar microcoil for implantable microsystems. *Sens. Actuators A Phys.* **1997**, *62*, 599–611. [CrossRef]

16. Rogers, J.A.; Someya, T.; Huang, Y. Materials and mechanics for stretchable electronics. *Science* **2010**, *327*, 1603–1607. [CrossRef] [PubMed]

17. Fan, J.A.; Yeo, W.H.; Su, Y.; Hattori, Y.; Lee, W.; Jung, S.Y.; Zhang, Y.; Liu, Z.; Cheng, H.; Falgout, L.; et al. Fractal design concepts for stretchable electronics. *Nat. Commun.* **2014**, *5*. [CrossRef] [PubMed]

18. Huang, X.; Liu, Y.; Cheng, H.; Shin, W.J.; Fan, J.A.; Liu, Z.; Lu, C.J.; Kong, G.W.; Chen, K.; Patnaik, D.; et al. Materials and designs for wireless epidermal sensors of hydration and strain. *Adv. Funct. Mater.* **2014**, *24*, 3846–3854. [CrossRef]

19. *Ecoflex Series*; MSDS No. 844; Smooth-On Inc.: Easton, PA, USA, 2013.

20. Salvatore, G.A.; Sülzle, J.; Dalla Valle, F.; Cantarella, G.; Robotti, F.; Jokic, P.; Knobelspies, S.; Daus, A.; Büthe, L.; Petti, L.; et al. Biodegradable and Highly Deformable Temperature Sensors for the Internet of Things. *Adv. Funct. Mater.* **2017**, *27*, 1–10. [CrossRef]

21. Park, G.; Chung, H.J.; Kim, K.; Lim, S.A.; Kim, J.; Kim, Y.S.; Liu, Y.; Yeo, W.H.; Kim, R.H.; Kim, S.S.; et al. Immunologic and tissue biocompatibility of flexible/stretchable electronics and optoelectronics. *Adv. Healthc. Mater.* **2014**, *3*, 515–525. [CrossRef] [PubMed]

22. Chou, N.; Lee, J.; Kim, S. Large-sized out-of-plane stretchable electrodes based on poly-dimethylsiloxane substrate. *Appl. Phys. Lett.* **2014**, *105*, 241903. [CrossRef]

23. Horowitz, P.; Winfield, H. *The Art of Electronics*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2015; pp. 15–20, ISBN 0521809266.

24. Greenhouse, H.M. Design of Planar Rectangular Microelectronic Inductors. *IEEE Trans. Parts Hybrids Packag.* **1974**, *10*, 101–109. [CrossRef]

25. Agilent Technologies. *Agilent 4294A Precision Impedance Analyzer Operation Manual*, 7th ed.; Agilent Technologies: Santa Clara, CA, USA, 2003; pp. 249–253.

26. So, J.H.; Thelen, J.; Qusba, A.; Hayes, G.J.; Lazzi, G.; Dickey, M.D. Reversibly deformable and mechanically tunable fluidic antennas. *Adv. Funct. Mater.* **2009**, *19*, 3632–3637. [CrossRef]

27. Hayes, G.J.; So, J.-H.; Qusba, A.; Dickey, M.D.; Lazzi, G. Flexible Liquid Metal Alloy (EGaIn) Microstrip Patch Antenna. *IEEE Trans. Antennas Propag.* **2012**, *60*, 2151–2156. [CrossRef]

28. Shin, G.; Gomez, A.M.; Al-Hasani, R.; Jeong, Y.R.; Kim, J.; Xie, Z.; Banks, A.; Lee, S.M.; Han, S.Y.; Yoo, C.J.; et al. Flexible Near-Field Wireless Optoelectronics as Subdermal Implants for Broad Applications in Optogenetics. *Neuron* **2017**, *93*, 509–521.e3. [CrossRef] [PubMed]

29. Kim, J.; Banks, A.; Xie, Z.; Heo, S.Y.; Gutruf, P.; Lee, J.W.; Xu, S.; Jang, K.I.; Liu, F.; Brown, G.; et al. Miniaturized Flexible Electronic Systems with Wireless Power and Near-Field Communication Capabilities. *Adv. Funct. Mater.* **2015**, *25*, 4761–4767. [CrossRef]

Article

Classification of Human Daily Activities Using Ensemble Methods Based on Smartphone Inertial Sensors

Ku Nurhanim Ku Abd. Rahim [1], I. Elamvazuthi [1,*], Lila Iznita Izhar [1] and Genci Capi [2]

[1] Smart Assistive and Rehabilitative Technology (SMART) Research Group,
Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS,
32610 Bandar Seri Iskandar, Malaysia; ku_g03269@utp.edu.my (K.N.K.A.R.); lila.izhar@utp.edu.my (L.I.I.)

[2] Assistive Robotics Laboratory, Department of Mechanical Engineering, Faculty of Science and Engineering,
HOSEI University, Tokyo 184-8584, Japan; capi@hosei.ac.jp

[*] Correspondence: irraivan_elamvazuthi@utp.edu.my; Tel.: +605-3687882

Abstract: Increasing interest in analyzing human gait using various wearable sensors, which is known as Human Activity Recognition (HAR), can be found in recent research. Sensors such as accelerometers and gyroscopes are widely used in HAR. Recently, high interest has been shown in the use of wearable sensors in numerous applications such as rehabilitation, computer games, animation, filmmaking, and biomechanics. In this paper, classification of human daily activities using Ensemble Methods based on data acquired from smartphone inertial sensors involving about 30 subjects with six different activities is discussed. The six daily activities are walking, walking upstairs, walking downstairs, sitting, standing and lying. It involved three stages of activity recognition; namely, data signal processing (filtering and segmentation), feature extraction and classification. Five types of ensemble classifiers utilized are Bagging, Adaboost, Rotation forest, Ensembles of nested dichotomies (END) and Random subspace. These ensemble classifiers employed Support vector machine (SVM) and Random forest (RF) as the base learners of the ensemble classifiers. The data classification is evaluated with the holdout and 10-fold cross-validation evaluation methods. The performance of each human daily activity was measured in terms of precision, recall, F-measure, and receiver operating characteristic (ROC) curve. In addition, the performance is also measured based on the comparison of overall accuracy rate of classification between different ensemble classifiers and base learners. It was observed that overall, SVM produced better accuracy rate with 99.22% compared to RF with 97.91% based on a random subspace ensemble classifier.

Keywords: gait; human activity recognition; smartphone; wearable sensor; human daily activity; ensemble method

1. Introduction

Recently, increasing interest has been shown in analysing human gait using wearable sensors, which is known as Human Activity Recognition (HAR). HAR with automatic recognition of physical activities is increasingly being studied and applied in Human-Computer Interaction (HCI), mobile and pervasive computing. One of the objectives of HAR is to offer information that enables computing frameworks to productively aid users in their tasks [1]. A few feasible applications that can be used with HAR which improves the service are on-request data systems, monitoring and surveillance system of smart homes, interactive interfaces for mobile services and games, and medical care service applications for both inpatient and outpatient treatments [2–4]. In addition, other applications include bilateral links targeted to advertising, entertainment, games, and multimedia visualization guidance [5,6].

Typically, human daily activities are divided into three categories, namely, gestures, low-level activities and high-level activities. Gestures involve simple activities such as the opening-closing of hands and bending of arms. The low-level activities are standing, sitting, walking, cycling and jogging, whereas, the high-level activities are cooking, dancing, eating, drinking and talking. A number of researchers have explored machine vision systems in gesture and activity recognition from video and still images in various settings [7–9].

Advances in sensor innovation in on-body wearable sensors and smartphones have enabled them to be used effectively for HAR systems. However, the difference between on-body wearable sensor-based HAR systems and the smartphone-based HAR systems is that, smartphones have a few integrated sensors that are equipped to provide an extensive variety of selections embedded into one cohesive device. Moreover, smartphones have computing ability, although not as capable as the devoted control units of wearable-sensor systems. In addition, smartphones have become an essential gadget in human's daily life and the usage of a smartphones greatly exceeds that of on-body wearable sensor-based systems. Hence, smartphones have become a prominent tool for assisting and supporting patients undergoing health rehabilitation and treatment, activity monitoring of daily living and diets, and for numerous other health issues [10,11].

One of the critical issues in HAR is the classification of the different activities performed by the users. The studies conducted in past show that machine learning algorithms such as Naïve Bayes tree (NBTREE), Decision Trees (C4.5), Neural Network (NN), k-Nearest Neighbour (k-NN) and Support Vector Machine (SVM) have been used for classification employing smartphone data [12–15]. Recently, Ensemble learning with bagging and boosting techniques have been found to enhance the accuracy of classifiers. Ensemble learning has been effectively tested and validated adequately on various datasets [16,17].

The study described in this paper highlights the classification of six different daily activities based on the data acquired from inertial sensors of a smartphone using ensemble methods such as Bagging, Adaboost, Rotation forest, Ensemble nested dichotomies (END) and Random subspace together with two base learner techniques such as SVM and Random forest (RF). Two datasets from the UCI Machine Learning Repository [18] were utilised in the study, where each dataset involved 30 subjects performing activities such as walking, walking upstairs, walking downstairs, sitting, standing and laying. The classifications have been evaluated in terms of accuracy, precision, recall, F-measure, and receiver operating characteristic (ROC) curve. This paper is structured as follows: Section 2 discusses the literature review in terms of related work and general HAR system. Methods are presented in Section 3, followed by Section 4 which describes the results and discussion, and finally, conclusions are provided in Section 5.

2. Literature Review

2.1. Related Work

Over the last few years, from 2013 to 2018, human daily activities such as sitting on chairs and on the floor, lying right, lying left, slow walk, brisk walk, walking upstairs, walking downstairs, standing, laying, etc. have been extensively studied. Many researchers have worked towards evaluation of HAR using the smartphone with different daily activities. These activities are classified based on feature extraction schemes that are broadly categorized as time and frequency domains. In [19–25] researchers have implemented the time domain and frequency domain feature extraction as a combined approach. Other researchers in [26–31] have used feature extraction in the time domain only, whilst, in [32] researchers applied the frequency domain and time-frequency domain. The authors in [33] have chosen the time domain, frequency domain and time-frequency domain for feature extraction. Out of these studies, Saha et al. [19] found that the ensemble classifier performs best, with an overall accuracy rate of 94% using accelerometer and gyroscope sensor data. In the research carried out by Mohamed et al. [20], a combination of accelerometer data from the arm, belt and pocket analysed using

rotation forest with the base learner C4.5, was found to provide the best overall classification accuracy rate of 98.9% [20]. Researchers in [21,32], and [23–25] have analyzed the same dataset. Ronao and Cho [21] found that classification using two stages of continuous hidden Markov model (TS-CHMM) achieved the highest overall accuracy rate of 93.18%. Jiang, Yin et al. [32] have obtained the best overall accuracy rate of 97.59% using deep convolution neural network (DCNN). Research reported by Kastner et al. [23] provided the best results with an overall classification accuracy rate of 96.23% with generalized learning vector quantization (GLVQ). Romero et al. [24] have found that One vs. one (OVO), OVO-SVM gives the best overall classification accuracy rate as 96.4%; whilst, Anguita et al. [25] managed to gain a little improvement and reported best overall accuracy rate as 96.5% for classification using One vs. all (OVA), OVA-SVM. Researchers in [26–28], have analyzed the same dataset for walking, jogging, walking downstairs, walking, upstairs, sitting, and standing activities. A study conducted by Sufyan et al. [26] found that classification on voting Multilayer perceptron (MLP) and NBtree give the best accuracy rate for classification based on each activity. This study found that Voting MLP-NBtree gives the best accuracy rate of classification on walking at 99.23%, jogging at 98.86%, walking upstairs at 93.35%, walking downstairs at 90.15%, sitting at 98.37% and standing at 98.37%. Research by Daghistani and Alshammari [27] found the best overall classification accuracy rate using Adaboost (J48) at 94.034%, whilst, Catal et al. [28] reported the highest overall classification accuracy rate based on voting (J48, logistic regression and MLP) as 94.06%. Gupta and Kumar [29] stated the study on human activities on sitting, standing, walking and running produced the best overall accuracy rate of 98.83% using an Adaboost classifier. Research by Gao et al. [33] showed that C4.5 was the best classifier with a 96.4% overall accuracy rate for lying, sitting, standing, walking and transition activity. Bayat et al. [30] studied HAR comparison between different classifiers such as MLP, SVM, RF, simple logistic, logitboost, Logistic model tree (LMT) and voting classifier with the triaxial accelerometer data in a smartphone that placed in pocket and hand. The results of this study showed that data in hand produces best overall classification accuracy rate using voting combination of MLP, logitboost and SVM classifier with 91.15%, whilst, data in pocket gives best accuracy of classification rate as 90.34% using voting combination of MLP, RF and simple logistic. Research studies from Ha and Ryu [31] have reported that the best overall classification accuracy rate of 97.8% was obtained with an ensemble method known as Error correcting output coding (ECOC) that was combined with the random forest as the base learner. Overall, it can be concluded that Ensemble methods produce better results compared with other algorithms.

2.2. HAR System

HAR is a way toward recognizing common human activities in daily living. It is turning into an attractive research field because of numerous areas of application. Physiological signals, environmental signals, location data and acceleration signals are the categories of input data that are acquired from wearable sensors in an HAR setting. Physiological signals data such as skin conductivity, heart rate, skin temperature, respiration rate and electrocardiography signals have also been considered in a few research studies to improve the recognition accuracy [34]. Environmental signals, for example, audio level, temperature, humidity are proposed to provide information exposing the individual's environment [35,36]. The Global Positioning System (GPS) empowers all kinds of area-based information. Current smartphones are equipped with integrated GPS, making it exceptionally advantageous for context-aware applications such as the recognition of the individual's transportation mode [37]. Accelerometers can be categorized into two types: either as body wearable sensors or incorporated with new models of mobile devices such as smartphones and smart watches which record the body movement [38,39]. The majority of research work in HAR applications is performed using wearable sensors. Figure 1 shows the HAR with different wearable sensors.
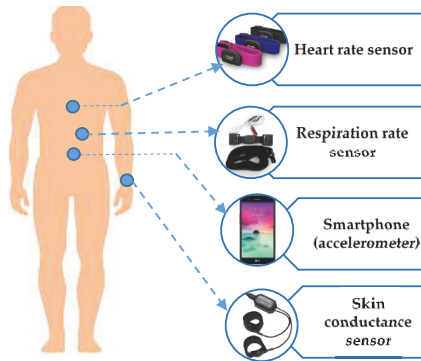
Figure 1. Wearable sensors for human daily activities.

The various types of wearable sensors are used to identify different human activities in three categories: gesture, low-level activities and high-level activities as shown in Figure 2.

Figure 2. Categories of human daily activities.

Gestures activities are classified as extremely short activities such as bending the arm, or opening and closing the hands. Human daily activities such as sitting, standing and running which typically last between seconds and a few minutes are known as low-level activities. Activities like working at the office, cleaning the house and sightseeing normally last for more than a few minutes up to a few hours are known as high-level activities [40]. Detailed information about the general process for training and testing the data of HAR systems based on wearable sensors can be found in [41].

3. Methods

3.1. Signal Processing Method for HAR

The signal processing method of HAR system that was carried out in this research is shown in Figure 3.
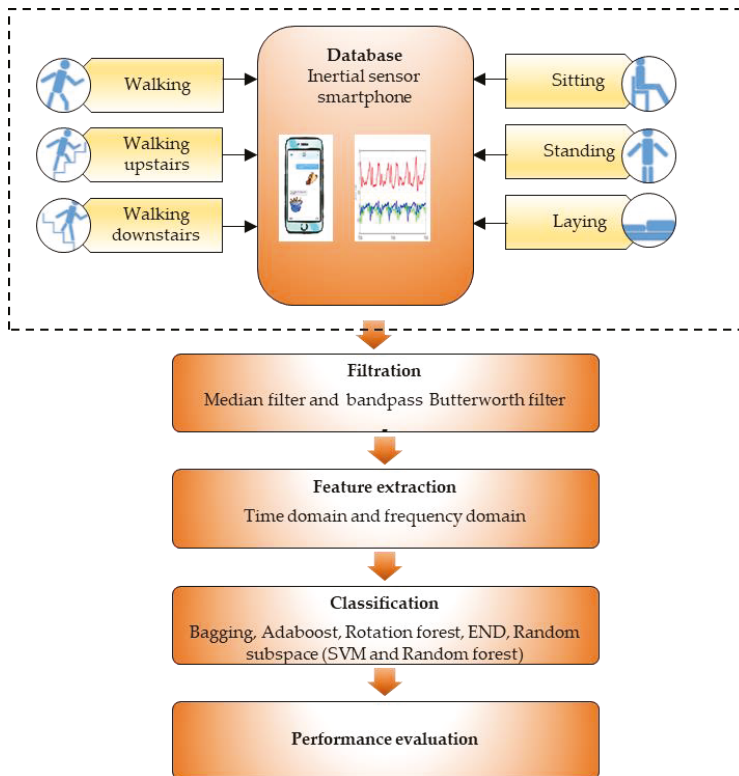
Figure 3. Signal processing method of HAR system.

According to Figure 3, the raw signals of human daily activities such as walking, walking upstairs, walking downstairs, sitting, standing and laying are acquired from the inertial sensors of smartphones. Then, the raw signals are segmented using sliding windows and filtered using median and band pass Butterworth filters to remove the irrelevant information or noise. Next, feature extraction of time and frequency domain are implemented. Thereafter, all the data of extracted features are classified with ensemble classifiers using base learners of SVM and RF. The performance of classification of is measured using performance evaluation metrics such as precision, recall, F-measure, accuracy and ROC.

3.2. Database

For this study, two datasets were obtained from UCI Machine Learning Repository [18]. The datasets consist of dataset 1 and dataset 2, containing 5447 data samples and 10,299 data samples, respectively. Dataset 1 was collected from a smartphone worn around the waist of 31 participants within the age group of 22–79 years. Each activity was performed for 60 s. For dataset 2, a group of 30 volunteers with ages ranging 19–48 years were selected. Each person was instructed to follow

a protocol of activities while wearing a waist mounted smartphone. For both datasets, the data collections were performed using Samsung Galaxy S2 smartphones which comprise an accelerometer and gyroscope to acquire tri-axial linear acceleration and angular velocity signals at a sampling rate of 50 Hz. Although the tasks were carried out in laboratory conditions, volunteers were asked to perform freely the sequence of activities aiming to simulate a more naturalistic dataset. Table 1 shows the activities that were carried out for both the datasets.

Table 1. List of daily living activities.

Activity Reference	Description of Activity
A1	Walking
A2	Walking upstairs
A3	Walking downstairs
A4	Sitting
A5	Standing
A6	Laying

3.3. Data Pre-Processing Filtering and Feature Extraction

The raw signals were processed by applying a median filter and 3rd order low pass Butterworth filter with 20 Hz cut-off. These signals were then segmented with fixed width sliding windows of 2.56 s with 50% overlap. A total of 17 features extracted from the time and frequency domains for each window were used in this work, where only about six features are shown in Table 2. The full 17 features can be found in [42].

Table 2. 6 Features extraction of time and frequency domain for each window.

Feature	Description
Min	Smallest value in the array
Max	Largest value in the array
Std	Standard deviation
Entropy	Signal entropy
Kurtosis	Kurtosis of the frequency domain signal
Skewness	Skewness of the frequency domain signal

3.4. Classification Techniques—Ensemble Methods

The principle thought behind the ensemble method is to evaluate a few single classifiers and combine them to acquire a classifier that surpasses each individual one of them. The motivation behind supervised learning is to classify patterns or instances that are specified into a group denoted as labels or classes. Often, the classification is determined by a classification models (classifiers) that are inferred from a pre-classified design model. Nevertheless, the classification employs knowledge that is provided by a specialist in the application area which is referred to as training data. The training set is a standardized supervised learning set that has a set of instances. The labels of the instances in the training set are identified and the objective is to develop a model with a specific end goal to label new instances. An algorithm which builds the model is called inducer and a case of an inducer for a specific training set is known as a classifier [43]. An ensemble comprises several inducers which are commonly referred to as the base classifier or base learner. A base learner is an algorithm that receives a set of labelled instances as input and produces a model that simplifies these instances. Predictions are determined for new unclassified instances by utilizing the created model. The generalization capability of an ensemble is usually more robust than that of base learners. As a matter of fact, ensemble method is likable because it can boost weak learners, referred to base classifiers which are marginally better than random estimate to strong learners which can make very accurate predictions. In this research work, ensemble methods such as Bagging, Adaboost M1, Rotation forest, Ensembles of

Nested Dichotomies (END) and random subspace were used for classification. An ensemble inducer can be of any type of base classifiers such as decision tree, neural network, k-NN and others type of base learner algorithm [43]. In this research work, the base learners applied were SVM and RF. The detail information about the ensemble methods can be found in [44] for bagging, [45,46] for Adaboost.M1, [47] for Rotation Forest, [48] for Ensembles Nested Dichotomies (END), [49] for Random subspace, [50] for Random Forest (RF) and [51] for SVM.

The above stated five Ensemble methods and two base learner algorithms were used to classify six human daily activities based on a classifier tool known as WEKA 3.8 Version [52] with model evaluation of the holdout method (contains 70% of the training set and 30% of the test set) and 10-fold cross-validation method. A Wilcoxon test was performed on the results to discover if the accuracy rate of classification instances was significantly different for SVM as a baseline compared to the RF as base learners in five different ensemble methods. This statistical test was conducted on IBM SPSS version 20 [53]. A value of p less than 0.05 is considered as statistically significant when the confidence level is set to 95%.

3.5. Performance Evaluation

The performance evaluation that was implemented in this study is based on the following expressions:

$$\text{Precison} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F} - \text{measure} = \frac{2\,(TP)}{2\,((TP) + (FN))} \tag{3}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Receiver operating characteristic (ROC) curve is a plot graph of the true positive rate against the false positive rate at different classification threshold setting. The true positive rate is also known as Recall in Equation (2) and false positive is defined as following expression:

$$\text{False positive rate} = \frac{FP}{FP + TN} \tag{5}$$

where TP—True Positive, TN—True Negative, FP—False Positive, and FN—False Negative.

4. Results and Discussion

4.1. Performance Evaluation of Dataset 1

4.1.1. Holdout and 10 Cross-Validations for Precision, Recall, F-measure, and ROC Evaluation

Table 3 presents the performance evaluation of END classifier with SVM and RF as base learner including precision, recall, F-measure and receiver operating characteristic (ROC) as the best classifier of the holdout method.

As shown in Table 3, the results of END classifier evaluation on holdout method obtained the best precision in activity A6 with 100% using SVM as a base learner compared to RF at 99.1%. For activity A5, both SVM and RF base learners achieved the precision results at 88.8%. But in activity A4, better precision was obtained by RF with 96.6% followed by SVM with 92.8%. For A3, A2 and A1, precision results show that SVM has given higher result range of 95.3% to 96.1% compared to RF with 89.6% to 92.7%. Recall results show that SVM obtained 99.10% and RF obtained 97.30% for activity A6. However, for activity A5, RF produced 94.6% compared to SVM with 91.5%. For A4, A3, A2, and A1, recall results show that SVM produced results ranging from 90.4% to 97.6% compared to RF with 90.1%

to 94.9%. Results of F-measure evaluation for activity A6 is 99.6% for SVM and 98.2% for RF. However, the F-measure results for activity A4 and A5 for RF are 93.2% and 91.6% respectively compared to SVM with 91.60% and 90.20%. Activities A1, A2, and A3 give SVM better F-measure results ranging from 94.7% to 96.9%, higher than RF range from 90.2% to 93.8%. RF gained greater results for ROC evaluation with results ranging from 99.4% to 100% compared to SVM which produced 95.2% to 99.8% for all activities. Table 4 presents the performance evaluation of Random subspace classifier with SVM and RF as base learner for 10-fold cross-validation method.

Table 3. Performance evaluation of each activity with random subspace classifier on the holdout method.

	END (Holdout)							
	SVM				**RF**			
Activity	**Precision**	**Recall**	**F-measure**	**ROC**	**Precision**	**Recall**	**F-measure**	**ROC**
A1	96.10%	97.60%	96.90%	99.20%	92.70%	94.90%	93.80%	99.70%
A2	95.30%	94.00%	94.70%	98.80%	89.60%	90.80%	90.20%	99.40%
A3	95.50%	96.00%	95.70%	98.80%	92.30%	91.90%	92.10%	99.50%
A4	92.80%	90.40%	91.60%	95.20%	96.60%	90.10%	93.20%	99.60%
A5	88.80%	91.50%	90.20%	96.20%	88.80%	94.60%	91.60%	99.40%
A6	100.00%	99.10%	99.60%	99.80%	99.10%	97.30%	98.20%	100.00%

Table 4. Performance evaluation for each activity of a random subspace classifier on 10-fold cross-validation method.

	Random Subspace (10-fold Cross-Validation)							
	SVM				**Random Forest**			
Activity	**Precision**	**Recall**	**F-measure**	**ROC**	**Precision**	**Recall**	**F-measure**	**ROC**
A1	95.60%	97.70%	96.70%	99.30%	90.30%	95.90%	93.10%	99.60%
A2	95.40%	94.80%	95.10%	98.70%	92.50%	90.00%	91.20%	99.30%
A3	96.60%	94.70%	95.70%	98.30%	93.80%	90.10%	91.90%	99.40%
A4	93.00%	93.90%	93.40%	98.20%	96.40%	94.20%	95.30%	99.80%
A5	93.50%	92.50%	93.00%	98.60%	94.10%	95.90%	95.00%	99.60%
A6	99.00%	99.40%	99.20%	99.80%	98.00%	98.20%	98.10%	100.00%

As shown in Table 4, the results of Random subspace classifier using 10-fold cross-validation indicates SVM as base learner produces better precision range of results from 95.6% to 99% compared to RF with 90.3% to 98% for activities A1, A2, A3 and A6. However, RF has obtained better precision results of 94.2% and 95.9% compared to SVM with 93% and 93.5% for activities A4 and A5. For recall, activities A1, A2, A3, and A6 with SVM produced superior results from 94.7% to 99.4% compared to RF with 90% to 98.2%. However, RF has obtained better recall results of 94.2% and 95.9% compared to SVM with 93.9% and 92.5% for activities A4 and A5. The results of F-measure for activities A1, A2, A3 and A6 ranged from 95.1% to 99.2% for SVM and 91.2% to 98.1% for RF. The ROC results for RF were from 99.3% to 100% and 98.2% to 99.8% for SVM for all the activities. In dataset 1, overall results of 10-fold cross-validation model evaluation give better results compared to holdout for both base learners.

4.1.2. Holdout and 10-Fold Cross-Validations for Overall Accuracy Rate

Figures 4 and 5 show the accuracy rate classification for each activity for ensemble method with base learners SVM and RF using holdout and 10-fold cross validation methods, respectively, for dataset 1.

As shown in Figure 4, the accuracy rate of activity A1, A2, A3 and A6 of Random subspace classifier with SVM for holdout method produced superior accuracy rate with the range from 98.4% to 99.9% compared to RF which produced an accuracy rate of 97.0% to 99.4%. On the other hand, RF produced better results with an accuracy rate of 97.1% and 96.9% compared to SVM with 97.1% and 96.9% activities A4 and A5.

Accuracy rate holdout method
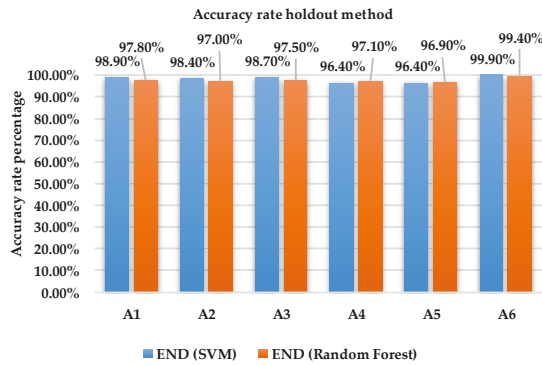
Figure 4. Accuracy rate of each activity holdout method dataset 1.

Accuracy rate 10 fold cross validation method
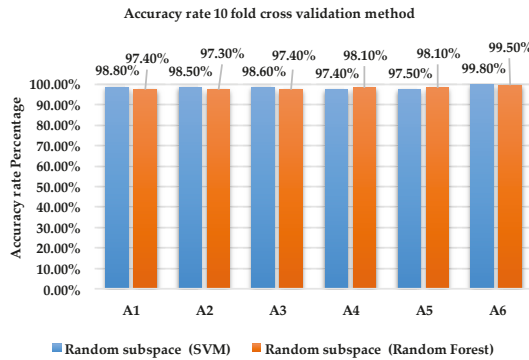
Figure 5. Accuracy rate of each activity 10 fold cross validation method dataset 1.

Results from Figure 5 indicate that the accuracy rate for activities A1, A2, A3 and A6 of Random subspace classifier with SVM using 10-fold cross validation is higher with 98.5% to 99.8% compared to RF with 97.3% to 99.5%. The accuracy rate for activities A4 and A5 is 97.4% 97.5% for SVM and 98.1% for RF.

Tables 5 and 6 show the overall accuracy rate of all other ensemble methods with base learners SVM and RF using holdout method and 10-fold cross validation method for dataset 1.

Table 5. Overall performance evaluation ensemble methods on the holdout method.

Overall Accuracy Rate			
	Holdout		
Ensemble Method	SVM	RF	*p*-Value
Bagging	93.83%	91.62%	0.028
Adaboost	94.24%	94.24%	0.917
Rotation forest	89.95%	92.23%	0.344
END	94.50%	93.16%	0.172
Random subspace	94.24%	92.76%	0.116

Results in Table 5 demonstrate that SVM has a significantly greater accuracy with Bagging (93.83%, $p = 0.028$). There were no significant differences for other ensemble methods using the holdout method and 10-fold cross-validation method as shown in Table 6 for dataset 1.

Table 6. Overall performance evaluation ensemble classifiers for 10-fold cross-validation method.

Overall Accuracy Rate			
	10-Fold Cross-Validation		
Ensemble Method	**SVM**	**Random Forest**	***p*-Value**
Bagging	94.57%	92.88%	0.173
Adaboost	94.84%	94.74%	0.917
Rotation forest	90.65%	93.65%	0.075
END	95.14%	94.48%	0.249
Random subspace	95.33%	94.08%	0.249

4.2. Performance Evaluation of Dataset 2

4.2.1. Holdout Method for Precision, Recall, F-measure, and ROC Evaluation

Table 7 presents the performance evaluation of random subspace classifier with SVM and RF as base leaners as the best classifier of holdout method.

Table 7. Performance evaluation of each activity with random subspace classifier on the holdout method.

Random Subspace (Holdout)						
	SVM			**RF**		
Activity	**Precision**	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
A1	99.80%	100.00%	99.90%	98.20%	99.00%	98.60%
A2	98.90%	99.50%	99.20%	98.20%	98.60%	98.40%
A3	99.80%	99.00%	99.40%	98.80%	97.30%	98.00%
A4	96.70%	97.20%	97.00%	96.60%	95.30%	95.90%
A5	97.70%	97.00%	97.30%	95.80%	97.20%	96.50%
A6	100.00%	100.00%	100.00%	100.00%	99.80%	99.90%

As shown in Table 7, the results of random subspace classifier evaluation on holdout method obtained the best precision in activity A6 with 100% for SVM and RF. For other activities, SVM shows better precision results between 96.7% and 99.8% compared to RF which obtained 95.8% to 98.8%. For recall, SVM achieved 100% compared to RF 99.8% for activity A6. SVM results for activities A1, A2, A3 and A4 ranges from 97% to 100% compared RF which obtained 95.8% to 99%. However, RF produced better recall results with 97.2% compared to SVM which recorded 97% for activity A5. The F-measure evaluation shows that SVM obtained 100% compared to RF that produced 99.9% for activity A6. For activity A4, SVM and RF obtained 97% and 95.9% respectively which are the lowest performance compared to activities A1, A2, A3 and A5. Table 8 shows the ROC curve of each activity of Random subspace with SVM classifier and Random subspace with RF using hold out method for dataset 2.

As shown in Table 8, the results of ROC evaluation gained 1.000 in activities A1, A2, A3 and A6 in both base learners. For activities A4 and A5, the ROC results for RF was 0.999 compared to 0.995 and 0.998 for SVM.

Table 8. ROC for each activity of the random subspace classifier on holdout method.

Random Subspace (Holdout)		
	SVM	**RF**
Activity	**ROC**	**ROC**
A1	1.000	1.000
A2	1.000	1.000
A3	1.000	1.000
A4	0.995	0.999
A5	0.998	0.999
A6	1.000	1.000

4.2.2. 10-Fold Cross Validation for Precision, Recall, F-measure, and ROC Evaluation

The cross-validation method evaluates fold (k) = 10 for all activities of dataset 2. Table 9 presents the performance evaluation of Random subspace as the best classifier of 10-fold cross-validation method.

Table 9. Performance evaluation for each activity of the random subspace classifier on 10-fold cross-validation method. Random subspace (10-fold cross-validation method).

	Random Subspace (10-Fold Cross-Validation Method)					
	SVM			**RF**		
Activity	**Precision**	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
A1	99.90%	100.00%	100.00%	99.90%	98.40%	98.70%
A2	99.70%	99.70%	99.70%	97.50%	99.20%	98.30%
A3	99.70%	99.80%	99.80%	98.50%	97.60%	98.00%
A4	97.90%	98.00%	98.00%	97.00%	95.20%	96.10%
A5	98.20%	98.10%	98.10%	95.60%	97.30%	96.40%
A6	100.00%	100.00%	100.00%	100.00%	99.80%	99.90%

From Table 9, the precision results for random subspace classifier on 10-fold cross-validation method shows that SVM and RF obtained 99.9% for activity A1 and 100% for both the SVM and RF for activity A6. For activities A2, A3, A4, and A5, SVM obtained between 97.9% and 99.9% and RF 95.6% to 98.5%. Recall results achieved 100% for activities A1 and A6 using SVM compared to RF which received 98.4% and 99.8%. For activities A2, A3, A4, A5, SVM gained between 98% and 99.8%, whereas, RF obtained between 95.2% and 99.2%. The results of F-measure for activities A1 and A6 for SVM was 100% compared to RF which obtained between 98.7% and 99.9%. SVM obtained between 98% and 99.7% compared to RF which produced results between 96.1% and 98.3% for the results of F-measure in activities A2, A3, A4, and A5.

Table 10 shows the ROC of each activity of Random subspace with SVM and Random subspace with RF classifier using 10-fold cross validation method dataset 2.

Table 10. ROC for each activity of the random subspace classifier on 10-fold cross validation method.

	Random Subspace (10-Fold Cross Validation)	
	SVM	**RF**
Activity	**ROC**	**ROC**
A1	1.000	0.999
A2	1.000	0.999
A3	1.000	0.999
A4	0.999	0.998
A5	0.999	0.999
A6	1.000	1.000

As shown in Table 10, the results of ROC evaluation for activities A1, A2, A3, and A6 for SVM is 1.000, whereas, RF produced 1.000 only for A6. For activities A4 and A5, SVM recorded 0.999. RF achieved 0.998 for activity A4 and 0.999 for A1, A2, A3 and A5.

As can be seen in Tables 7–10, overall performance evaluation in each activity on Random subspace classifier with base learner SVM is better than RF for holdout method and 10-fold cross-validation methods. The 10-fold cross-validation model evaluation gives better results compared to holdout for both base learners.

Representative ROC curves for SVM and RF are shown in Figures 6 and 7 based on Table 10 for activities A1, A2 and A6.

ROC curve for activity A1 10 fold cross validation method

Plot area=1.000

True positive rate

1

0.5

0

0   0.5

False positive rate

(a)

ROC curve for activity A2 10 fold cross validation method

Plot area=1.000

True positive rate

1

0.5

0

0   0.5   1

False positive rate

(b)

ROC curve for activity A6 10 fold cross validation method

Plot area=1.000

True positive rate

1

0.5

0

0   0.5   1

False positive rate
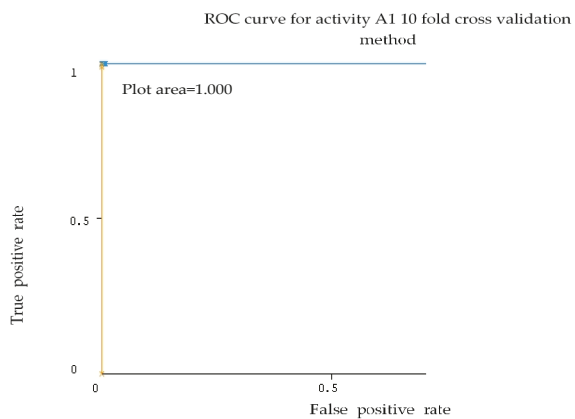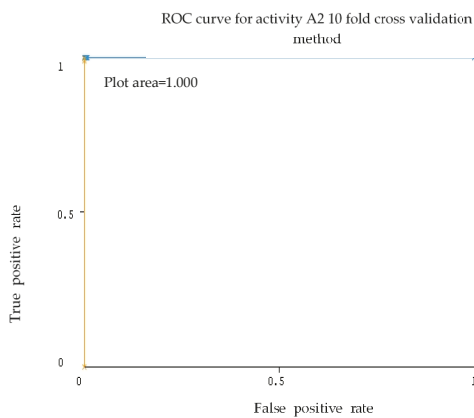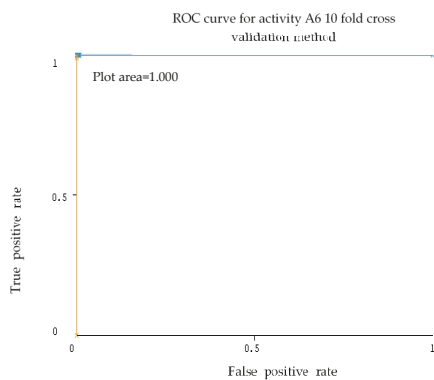
(c)

Figure 6. ROC graph activity A1 (**a**), A2 (**b**) and A6 (**c**) of Random subspace with SVM classifier using 10 fold cross validation method dataset 2.

ROC curve for activity A1 10 fold cross validation method

(a)

ROC curve for activity A2 10 fold cross validation method

(b)

ROC curve for activity A6 10 fold cross validation method
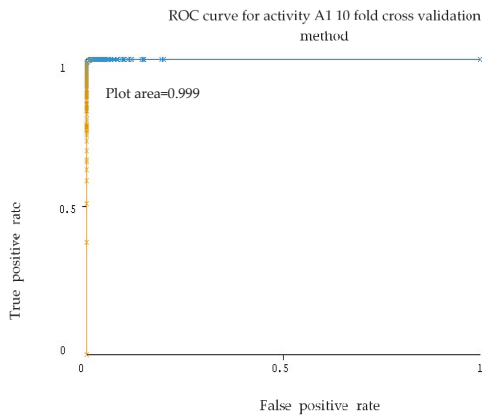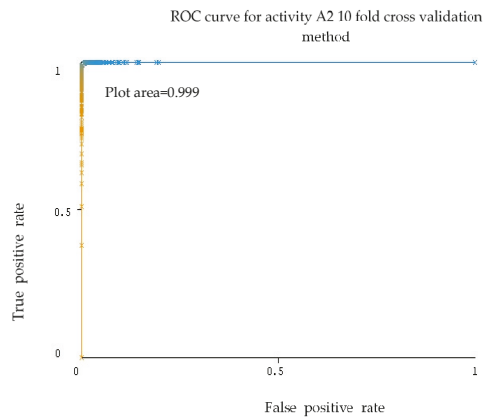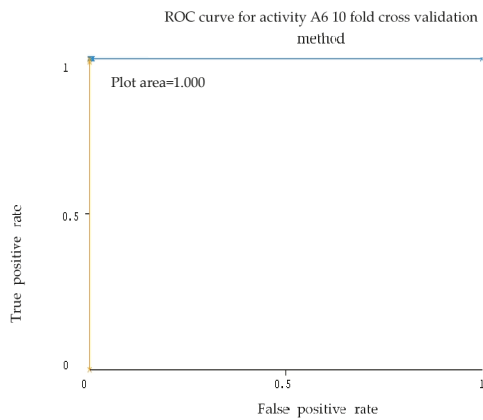
(c)

Figure 7. ROC graph activity A1 (**a**), A2 (**b**) and A6 (**c**) of Random subspace with RF classifier using 10-fold cross validation method dataset 2.

As can be seen in Figures 6 and 7, the activities A1, A2 and A6 for SVM and RF, the ROC results produced 1.000. For activities A1 and A2, the ROC results are 0.999 for RF.

4.2.3. Holdout and 10-Fold CROSS-validations for Overall Accuracy Rate Classification

Figures 8 and 9 show the accuracy rate of classification for each activity for ensemble method with base learners SVM and RF using holdout method and 10-fold cross validation method for dataset 2.
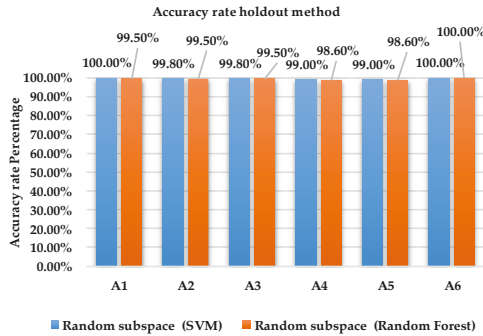
Figure 8. Accuracy rate of each activity for holdout method using dataset 2.
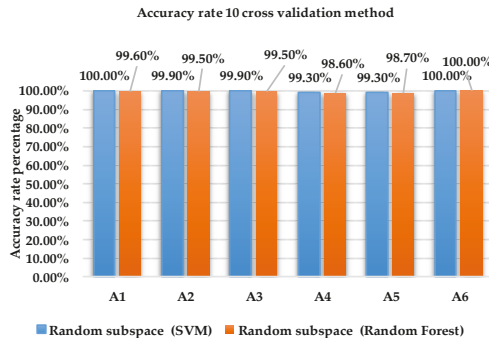
Figure 9. Accuracy rate of each activity of 10-fold cross validation method using dataset 2.

As shown in Figure 8, the accuracy rate of activity A6 is 100% for Random subspace classifier with SVM and RF using the holdout method. For activities A1, A2, A3, A4 and A5, the accuracy rate for SVM is between 99% and 100% compared to RF which obtained results between 98.6% and 99.5%.

As shown in Figure 9, the accuracy rate of activity A6 is 100% for Random subspace classifier with SVM and RF using the 10-fold cross validation method. For activities A1, A2, A3, A4 and A5, the accuracy rate for SVM is between 99.9% and 100% compared to RF which obtained results between 98.7% and 99.6%.

Tables 11 and 12 present the overall accuracy rate classification of other ensemble methods with base learners SVM and RF on holdout method and 10-fold cross validation method for dataset 2.

Results in Table 11 show that SVM base learner has significantly greater accuracy rate than RF with the same probability value (*p*-value) when using Bagging (98.54%, *p* = 0.028), END (98.61%, *p* = 0.028) and Random subspace (98.74%, *p* = 0.028). In this case, random subspace gives the highest accuracy rate in the holdout method. There was no significant difference in accuracy between the SVM and RF when employing Adaboost and RF for holdout method because of the higher *p*-value. From Table 12, SVM base learner demonstrates significantly greater accuracy rate than RF, whilst the

random subspace gives the highest accuracy rate with (99.22%, $p = 0.028$). The results of END (99.20%, $p = 0.028$), Adaboost (99.17%, $p = 0.028$) and Bagging (99.07%, $p = 0.028$) have significantly greater overall accuracy rate than RF as a base learner in 10-fold cross-validation method.

Table 11. Overall performance evaluation of ensemble classifiers for the holdout method.

	Overall Accuracy Rate		
	Holdout		
Ensemble Method	**SVM**	**RF**	***p*-Value**
Bagging	98.54%	97.18%	0.028
Adaboost	98.43%	98.07%	0.686
Rotation forest	98.07%	98.03%	0.893
END	98.61%	98.03%	0.028
Random subspace	98.74%	97.86%	0.028

Table 12. Overall performance evaluation of ensemble methods for the 10-fold cross-validation.

	Overall Accuracy Rate		
	10-Fold Cross-Validation		
Ensemble Method	**SVM**	**RF**	***p*-Value**
Bagging	99.07%	97.43%	0.028
Adaboost	99.17%	98.82%	0.028
Rotation forest	98.43%	98.22%	0.249
END	99.20%	98.28%	0.028
Random subspace	99.22%	97.91%	0.028

4.3. Comparative Analysis

The comparison of the overall accuracy rate of classification between different methods of classification with previous research work is represented in Table 13.

Table 13. Comparison of overall accuracy of classification with previous research work.

Reference	Evaluation Method	Dataset	Classification Method	Overall Accuracy Rate
Proposed classifier	10-fold Cross-validation	10,000 samples	Random subspace-SVM	99.22%
Proposed classifier	Holdout	10,000 samples	Random subspace-SVM	98.74%
Ronao and Cho (2017) [21]	10-fold Cross-validation	10,000 samples	Two stages of continuous Hidden Markov model	93.18%
Anguita et al. (2013) [25]	Holdout	10,000 samples	OVA MC-SVM-Gaussian kernel	96.5%
Romero-Paredes et al. (2013) [24]	Holdout	10,000 samples	OVO MC-SVM-Linear Kernel majority voting	96.40%
Kastner et al. (2013) [23]	Holdout	10,000 samples	Kernel generalized learning vector quantization	96.23%

As shown in Table 13, the proposed classifier with Random subspace-SVM achieved an accuracy rate of 99.22% for 10-fold cross-validation and 98.74% for holdout method. These proposed classifiers show improvement of overall accuracy rate of classification on the same dataset (10,000 samples) compared to previous work done by Ranoa and Chao [21] that has achieved an overall accuracy rate of 93.18%, Anguita et al. [25] that obtained overall accuracy rate of 96.5%, Romero-Paredes et al. [24] that acquired an overall accuracy rate of 96.4% and Kastner et al. [23] that produced overall accuracy rate of 96.23%. This shows that the Random subspace-SVM ensemble method has the capability to produce higher accuracy due to its ability to find a hyper plane which separates positive and negative training observations, and maximizes the margin between these observations through this hyper plane compared to other methods such as Two stages of continuous Hidden Markov model, OVA MC-SVM-Gaussian kernel, OVO MC-SVM-Linear Kernel majority voting and Kernel generalized learning vector quantization.

Although RF classifiers have predictive performance comparable to that of the best performing alternatives such as SVMs for classification of HAR, nevertheless, this research shows that SVM has a slight edge over RF. From a small sample size (dataset 1) to a larger sample size (dataset 2), the accuracy significantly increased with RF. This is true for SVM as well. This shows that sample size has more impact on the classification accuracy of both the RF and SVM. This is consistent with the results reported by Shao and Lunetta [54], Thanh and Kappas [55], Hasan et al. [56], Solé et al. [57] and Sheshasaayee and Thailambal [58].

It can be established that for the proposed Random subspace-SVM ensemble classification method, the 10-fold Cross-validation evaluation method produced better results than the holdout evaluation method. This is supported by various researchers such as Bengio et al. [59], Kim [60] and Sakr et al. [61].

It can be deduced that the ensemble method gives instinctive, straightforward, well-designed, and strong resolutions for an assortment of machine learning issues. As pointed out by Polikar [16], this method was initially created to enhance classification accuracy by decreasing the modification in classifier outputs. Ensemble methods have since ended up being exceptionally powerful in various areas that are difficult to address utilizing a single model-based classifier. Generally, most ensemble methods are self-determined for the type of base classifier used to construct the ensemble. This is a significant advantage that permits developing a specific kind of classifier that might be known to be most appropriate for a certain application.

5. Conclusions

In this study, different ensemble classifiers with different base learner algorithms were implemented to classify six human daily activities based on tri-axial inertial smartphone data. Comparative studies of classification techniques were presented using Bagging, Adaboost, Rotation forest, END and Random subspace with base learner as SVM and RF. The performance measures used to evaluate the classification techniques include overall accuracy, precision, recall, F-measure and ROC. Holdout and 10-fold cross-validation evaluation methods were used in the model evaluation of classification. As seen from the obtained results, Random subspace classifier with SVM gives the best results overall accuracy rate over different ensemble classifiers. The comparison in each activity classification showed the overall results performance of precision, recall, F-measure and ROC and accuracy rate using 10-fold cross-validation method was slightly higher compared to the holdout method. It can be summarized that ensemble classifiers have produced improved performance for the HAR with six different activities such as walking, walking upstairs, walking downstairs, sitting, standing and lying. In future, other methods would be explored to improve the performance.

References

1. Abowd, G.D.; Dey, A.K.; Orr, R.; Brotherton, J. Context Tracking in Wearable and Ubiquitous Computing. In Proceedings of the 1st International Symposium on Wearable Computers, Cambridge, MA, USA, 13–14 October 1997; pp. 179–180.
2. Fikri, M.; Fahmi, A.; Negara, P.; Sayeed, S.; Choi, D. Classification Algorithms in Human Activity Recognition Using Smartphones. *Int. J. Comput. Inf. Eng.* **2012**, *6*, 77–84.

3. Wu, W.H.; Bui, A.A.T.; Batalin, M.A.; Au, L.K.; Binney, J.D.; Kaiser, W.J. MEDIC: Medical Embedded Device for Individualized Care. *Artif. Intell. Med.* **2008**, *42*, 137–152. [CrossRef] [PubMed]

4. Patel, S.; Park, H.; Bonato, P.; Chan, L.; Rodgers, M. A Review of Wearable Sensors and Systems with Application in Rehabilitation. *J. NeuroEng. Rehabil.* **2012**, *9*, 21. [CrossRef] [PubMed]

5. Abowd, G.D.; Atkeson, C.G.; Hong, J.; Long, S.; Kooper, R.; Pinkerton, M. Cyberguide: A Mobile Context Aware Tour Guide. *Wirel. Netw.* **1997**, *3*, 421–433. [CrossRef]

6. Alt, F.; Shirazi, A.S.; Pfeiffer, M.; Holleis, P.; Schmidt, A. TaxiMedia: An Interactive Context-Aware Entertainment and Advertising System. In Proceedings of the Informatik 2009: Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Lübeck, Germany, 28 September–2 October 2009.

7. Mitra, S.; Acharya, T. Gesture Recognition: A Survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2007**, *37*, 311–324. [CrossRef]

8. Turaga, P.; Chellappa, R.; Subrahmanian, V.S.; Udrea, O. Machine Recognition of Human Activities: A Survey. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1473–1488. [CrossRef]

9. Aggarwal, J.K.; Ryoo, M.S. Human Activity Analysis: A Review. *ACM Comput. Surv.* **2011**, *43*, 1–47. [CrossRef]

10. Obermayer, J.L.; Riley, W.T.; Asif, O.; Jean-Mary, J. College Smoking-Cessation Using Cell Phone Text Messaging. *J Am. Coll. Health* **2004**, *53*, 71–78. [CrossRef] [PubMed]

11. Aguiar, B.; Silva, J.; Rocha, T.; Carneiro, S.; Sousa, I. Monitoring Physical Activity and Energy Expenditure with Smartphones. In Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, 1–4 June 2014; pp. 664–667.

12. Del Rosario, M.B.; Redmond, S.J.; Lovell, N.H. Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. *Sensors* **2015**, *15*, 18901–18933. [CrossRef] [PubMed]

13. Micucci, D.; Mobilio, M.; Napoletano, P. UniMiB SHAR: Dataset for Human Activity Recognition Using Acceleration Data from Smartphones. *Appl. Sci.* **2017**, *7*, 1101. [CrossRef]

14. Pires, I.M.; Garcia, N.M.; Pombo, N.; Flórez-Revuelta, F. From Data Acquisition to Data Fusion: A Comprehensive Review and a Roadmap for the Identification of Activities of Daily Living Using Mobile Devices. *Sensors* **2016**, *16*, 184. [CrossRef] [PubMed]

15. Jain, A.; Kanhangad, V. Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors. *IEEE Sens. J.* **2018**, *18*, 1169–1177. [CrossRef]

16. Polikar, R. Ensemble Based Systems in Decision Making. *IEEE Circuits Syst. Mag.* **2006**, *6*, 21–45. [CrossRef]

17. Sagi, O.; Rokach, L. Ensemble Learning: A Survey. *Wires Data Min. Knowl. Discovery* **2018**, *8*, e1249. [CrossRef]

18. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In Proceedings of the 4th International Workshop, IWAAL 2012, Vitoria-Gasteiz, Spain, 3–5 December 2012.

19. Saha, J.; Chowdhury, C.; Chowdhury, I.R.; Biswas, S.; Aslam, N. An Ensemble of Condition Based Classifiers for Device Independent Detailed Human Activity Recognition Using Smartphones. *Information* **2018**, *9*, 94. [CrossRef]

20. Mohamed, R.; Zainudin, M.N.S.; Sulaiman, M.N.; Perumal, T.; Mustapha, N. Multi-Label Classification for Physical Activity Recognition from Various Accelerometer Sensor Positions. *J. Inf. Commun. Technol.* **2018**, *18*, 209–231.

21. Ronao, C.A.; Cho, S.-B. Recognizing Human Activities from Smartphone Sensors Using Hierarchical Continuous Hidden Markov Models. *Int. J. Distrib. Sens. Netw.* **2017**, *13*, 1–16. [CrossRef]

22. Gao, L.; Bourke, A.K.; Nelson, J. Evaluation of Accelerometer Based Multi-Sensor versus Single-Sensor Activity Recognition Systems. *Med. Eng. Phys.* **2014**, *36*, 779–785. [CrossRef] [PubMed]

23. Kastner, M.; Strickert, M.; Villmann, T. A Sparse Kernelized Matrix Learning Vector Quantization Model for Human Activity Recognition. In Proceedings of the Europearn Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), Bruges, Belgium, 24–26 April 2013.

24. Romera-Paredes, B.; Aung, M.S.H.; Bianchi-Berthouze, N. A One-vs-One Classifier Ensemble with Majority Voting for Activity Recognition. In Proceedings of the Europearn Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), Bruges, Belgium, 24–26 April 2013.

25. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A Public Domain Dataset for Human Activity Recognition Using Smartphones. In Proceedings of the Europearn Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), Bruges, Belgium, 24–26 April 2013.

26. Azmi, M.S.M.; Sulaiman, M.N. Accelerator-Based Human Activity Recognition Using Voting Technique with NBTree and MLP Classifiers. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2017**, *7*, 146–152. [CrossRef]
27. Daghistani, T.; Alshammari, R. Improving Accelerometer-Based Activity Recognition by Using Ensemble of Classifiers. *(IJACSA) Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 128–133. [CrossRef]
28. Catal, C.; Tufekci, S.; Pirmit, E.; Kocabag, G. On the Use of Ensemble of Classifiers for Accelerometer-Based Activity Recognition. *Appl. Soft Comput.* **2015**, *37*, 1018–1022. [CrossRef]
29. Gupta, S.; Kumar, A. Human Activity Recognition through Smartphone's Tri-Axial Accelerometer Using Time Domain Wave Analysis and Machine Learning. *Int. J. Compt. Appl.* **2015**, *127*, 22–26. [CrossRef]
30. Bayat, A.; Pomplun, M.; Tran, D.A. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Comput. Sci.* **2014**, *34*, 450–457. [CrossRef]
31. Ha, E.T.; Ryu, K.R. Activity Recognition by Smartphone Accelerometer Data Using Ensemble Learning Methods. *Int. J. Electr. Comput. Eng.* **2014**, *8*, 480–483.
32. Jiang, W.; Yin, Z. Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 1307–1310.
33. Erdaş, B.C.; Atasoy, I.; Açici, K.; Oğul, H. Integrating Features for Accelerometer-Based Activity Recognition. *Procedia Comput. Sci.* **2016**, *98*, 522–527. [CrossRef]
34. Lara, Ó.D.; Prez, A.J.; Labrador, M.A.; Posada, J.D. Centinela: A Human Activity Recognition System Based on Acceleration and Vital Sign Data. *Pervasive Mob. Comput.* **2012**, *8*, 717–729. [CrossRef]
35. Pärkkä, J.; Ermes, M.; Korpipää, P.; Mäntyjärvi, J.; Peltola, J.; Korhonen, I. Activity Classification Using Realistic Data from Wearable Sensors. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 119–128. [CrossRef] [PubMed]
36. Maurer, U.; Smailagic, A.; Siewiorek, D.P.; Deisher, M. Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions. In Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06), Cambridge, MA, USA, 3–5 April 2006.
37. Reddy, S.; Mun, M.; Burke, J.; Estrin, D.; Hansen, M.; Srivastava, M. Using Mobile Phones to Determine Transportation Modes. *ACM Trans. Sens. Netw.* **2010**, *6*, 1–27. [CrossRef]
38. Shoaib, M.; Bosch, S.; Incel, O.D.; Scholten, H.; Havinga, P.J.M. Complex Human Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors. *Sensors* **2016**, *16*, 426. [CrossRef] [PubMed]
39. Ali, H.M.; Muslim, A.M. Human Activity Recognition Using Smartphone and Smartwatch. *Int. J. Comput. Eng. Res. Trends* **2016**, *3*, 568–576. [CrossRef]
40. Bajrami, G.; Derawi, M.O.; Bours, P. Towards an Automatic Gait Recognition System Using Activity Recognition (Wearable Based). In Proceedings of the 2011 Third International Workshop on Security and Communication Networks (IWSCN), Gjovik, Norway, 18–20 May 2011.
41. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition Using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2012**, *15*, 1192–1209. [CrossRef]
42. Nurhanim, K.; Elamvazuthi, I.; Izhar, L.I.; Ganesan, T. Classification of Human Activity based on Smartphone Inertial Sensor using Support Vector Machine. In Proceedings of the 2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA), Kuala Lumpur, Malaysia, 19–21 September 2017.
43. Rokach, L. Ensemble-Based Classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [CrossRef]
44. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
45. Schapire, R.E.; Singer, Y. Improved Boosting Algorithms Using Confidence-Rated Predictions. *Mach. Learn.* **1999**, *37*, 297–336. [CrossRef]
46. Freund, Y.; Schapire, R.E. A Desicion-Theoretic Generalization of on-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1995**, *55*, 119–139. [CrossRef]
47. Rodríguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation Forest: A New Classifier Ensemble Method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [CrossRef] [PubMed]
48. Frank, E.; Kramer, S. Ensembles of Nested Dichotomies for Multi-Class Problems. In Proceedings of the twenty-first international conference on Machine learning, Banff, AB, Canada, 4–8 July 2004.
49. Ho, T.K. The Random Subspace Method for Constructing Decision Forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
50. Amit, Y.; Geman, D. Randomized Inquiries about Shape: An Application to Handwritten Digit Recognition. Available online: http://www.dtic.mil/dtic/tr/fulltext/u2/a290812.pdf (accessed on 27 May 2018).

51. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. A Practical Guide to Support Vector Classification. *BJU Int.* **2008**, *101*, 1396–1400.
52. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA Data Mining Software: An Update. *ACM SIGKDD Explor. Newslett.* **2009**, *11*, 10–18. [CrossRef]
53. IBM SPSS Software Ver. 20. Available online: http://www.ibm.com (accessed on 24 June 2014).
54. Shao, Y.; Lunetta, R.S. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 78–87. [CrossRef]
55. Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2018**, *18*, 18. [CrossRef] [PubMed]
56. Hasan, M.A.M.; Nasser, M.; Pal, B.; Ahmad, S. Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS). *J. Intell. Learn. Syst. Appl.* **2014**, *6*, 45–52. [CrossRef]
57. Solé, X.; Ramisa, A.; Torras, C. Evaluation of Random Forests on Large-Scale Classification Problems Using a Bag-of-Visual-Words Representation. In *Artificial Intelligence Research and Development: Recent Advances and Development*; Museros, L., Pujol, O., Agell, N., Eds.; IOS Press: Amsterdam, The Netherlands, 2014; pp. 273–276.
58. Sheshasaayee, A.; Thailambal, G. Comparison of Classification Algorithms in Text Mining. *Int. J. Pure Appl. Math.* **2017**, *116*, 425–433.
59. Bengio, Y.; Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.
60. Kim, J.H. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Comput. Stat. Data Anal.* **2009**, *53*, 3735–3745. [CrossRef]
61. Sakr, S.; Elshawi, R.; Ahmed, A.; Qureshi, W.T.; Brawner, C.; Keteyian, S.; Blaha, M.J.; AI-Mallah, M.H. Using Machine Learning on Cardiorespiratory Fitness Data for Predicting Hypertension: The Henry Ford ExercIse Testing (FIT) Project. *PLoS ONE* **2018**, *13*, e0195344. [CrossRef] [PubMed]

Comparison of Different Algorithms for Calculating Velocity and Stride Length in Running Using Inertial Measurement Units

Markus Zrenner [1,*], Stefan Gradl [1], Ulf Jensen [2], Martin Ullrich [1] and Bjoern M. Eskofier [1]

[1]   Machine Learning and Data Analytics Lab, Department of Computer Science,
     Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91052 Erlangen, Germany;
     stefan.gradl@fau.de (S.G.); martin.ullrich@fau.de (M.U.); bjoern.eskofier@fau.de (B.M.E.)
[2]   Finance & IT—IT Innovation, Adidas AG, 91074 Herzogenaurach, Germany; ulf.jensen@adidas.com
*   Correspondence: markus.zrenner@fau.de; Tel.: +49-9131-85-20162

Abstract: Running has a positive impact on human health and is an accessible sport for most people. There is high demand for tracking running performance and progress for amateurs and professionals alike. The parameters velocity and distance are thereby of main interest. In this work, we evaluate the accuracy of four algorithms, which calculate the stride velocity and stride length during running using data of an inertial measurement unit (IMU) placed in the midsole of a running shoe. The four algorithms are based on stride time, foot acceleration, foot trajectory estimation, and deep learning, respectively. They are compared using two studies: a laboratory-based study comprising 2377 strides from 27 subjects with 3D motion tracking as a reference and a field study comprising 12 subjects performing a 3.2-km run in a real-world setup. The results show that the foot trajectory estimation algorithm performs best, achieving a mean error of 0.032 ± 0.274 m/s for the velocity estimation and 0.022 ± 0.157 m for the stride length. An interesting alternative for systems with a low energy budget is the acceleration-based approach. Our results support the implementation decision for running velocity and distance tracking using IMUs embedded in the sole of a running shoe.

Keywords: wearable sensors; inertial measurement unit; gait; running; stride length; velocity; smart shoe

1. Introduction

Distance running is a very popular sport. Two main reasons for this popularity are simplicity and the health benefit, as running can be done in small and restricted time-frames and does not require a specific location. Besides sports gear, no equipment is needed. Moreover, running improves health. Studies have shown that aerobic endurance training like running can reduce blood pressure [1] and that moderate running twice a week (>51 min or six miles) reduces overall mortality risk and the occurrence of cardiovascular diseases [2]. However, overtraining can also lead to a higher risk of injury of the lower extremities for distance runners [3].

Tracking running performance over time can prevent overtraining and greatly support a healthy and effective training. A training diary helps to maintain the right training intensity and volume, which are essential for both performance and health enhancement. Training records are also motivating, as they highlight both effort and progress. However, a precise, objective, and easy measurement of relevant parameters is needed. Two common parameters that both professional and amateur runners use to track their performance is the average velocity and total distance. With these parameters, the running workout can be easily categorized, rated, and compared. In the past, runners estimated the distance of a predefined running track and took time with a stopwatch to calculate the average

velocity of a distance run. With the rise of wearable technology in recent years, easier and more precise methods have become available.

1.1. Literature Review

The predominant approach to tracking average velocity and total distance during running is the global positioning system (GPS). Smartphones or even smartwatches comprise a GPS chip, which allows a satellite-based localization of a runner. By tracking the runner's absolute position over a complete run and using a solution to the second geodetic problem [4], the distance of a run can be measured. By incorporating the sampling frequency of the GPS module, a continuous time series of velocity values for the run can be computed. Thus, GPS delivers a time series of velocity, the cumulative distance, and the localization of the running track. From these data, the average velocity and the total distance can be extracted. The drawbacks of GPS are the additional gear (smartwatch, smartphone), the high energy demand, and the restriction to outdoor use.

Integrating sensors directly into running shoes can solve these issues. One type of sensor that can be integrated into a shoe is an inertial measurement unit (IMU). It is a small, lightweight, and inexpensive sensor, which is capable of measuring triaxial accelerations and triaxial angular rates. A shoe setup with integrated IMUs overcomes the described GPS issues: runners only need a running shoe with integrated IMU; IMUs are energy efficient and work both indoors and outdoors. Using IMU data, it is possible to compute a stride length and an average velocity value per stride. The underlying assumption for the velocity computation is that the average velocity of the foot per stride matches the running velocity. By collecting stride velocity values and accumulating the stride length values over time, a distance measure and a continuous velocity recording of a complete run can be provided. The following paragraphs describe the state-of-the-art of four approaches for IMU data processing for calculating these metrics.

In biomechanics, the relationships between stride frequency, stride length, running velocity, and body height was investigated [5]. The results indicated that with increasing running velocity, stride frequency and stride length increase. Thus, increasing running velocity is an interaction of increasing stride length and stride frequency [5]. Stride length itself depends on body height and can be expressed as a relative stride length. From these relationships, a generic model relating running velocity and stride length on the basis of the stride frequency can be deduced. The general idea behind this approach is the inverse correlation between velocity and stride time (the higher the velocity, the shorter the stride time). Thus, in order to estimate the stride length, only the stride time has to be distinguished by segmenting the data into single strides. An average velocity of the stride can then be calculated using the stride length and the measured stride duration.

Recently, Gradl et al. [6] proposed an algorithm that uses quadratic regression to compute the velocity of movements. The velocity was evaluated during running, as well as other movements and showed a relative error of $6.9 \pm 5.5\%$. The proposed algorithm is solely based on foot acceleration. Single strides are segmented from the data stream. Afterwards, the acceleration signal of all axes is integrated prior to the initial ground contact. Finally, the resulting integral value is converted to a velocity value using a quadratic regression model.

Another method to compute velocity and stride length values from IMU signals is to reconstruct the trajectory of the sensor in the course of a stride. This method is heavily used for gait analysis for geriatric patients [7–9] or in inertial navigation scenarios [10,11]. For trajectory reconstruction, sensor fusion techniques must be applied to both the accelerometer and the gyroscope. Several fusion algorithms to cope with this task exist. Bailey et al. [12] and Foxlin et al. [13] used extended Kalman filters to compute the trajectory from the acceleration and angular rate signals, while Rampp et al. [7] applied a linear dedrifting technique. Both algorithms rely on a zero-velocity update during the stance phase for the initialization of the orientation. The literature shows that this approach works well while analyzing walking [7], but it was not evaluated for free running. Bailey et al. [12] applied

their approach to treadmill running and showed a good accuracy of 0.03 ± 0.2 m/s. However, they evaluated neither the velocity nor the stride length in a free running scenario.

Deep learning techniques also show good results in IMU-based classification and regression tasks [14,15]. Hannink et al. [16] showed that deep convolutional neural network regression outperforms traditional stride length estimation in geriatric gait analysis. They trained a network with two convolutional layers, which was fed with the 6D IMU raw data of a stride. The output layer had a single node and provided an estimate of stride length.

1.2. Contribution

Most of the described algorithms were evaluated either for walking or for running on a treadmill. However, both of these conditions yield different signal characteristics to those of free running. In running, different strike patterns, such as rearfoot or forefoot strike, exist and affect the performance of these algorithms. Besides, the movement is also more dynamic, which yields higher accelerations, angular rates, and impacts. Therefore, our contribution is the comparison of different algorithmic approaches for computing average velocity and stride length during overground running using an IMU embedded into the sole of a running shoe. We evaluate these algorithms on a large database including high variation of the input data. Additionally, we run a field study to assess the performance in a real-world scenario. Based on the results, we give implementation recommendations for specific use cases.

2. Methods

2.1. Data Collection

We conducted two data collection studies for algorithm comparison, a lab study and a field study. The lab study was conducted in a sports research lab to evaluate the performance of the algorithms against ground-truth stride length and velocity labels on a per stride basis. A 3D motion tracking system was used as a reference. The field study was conducted on a 400-m outdoor running track to evaluate the performance regarding the total distance on a continuous 3.2-km run in realistic free-running conditions. The track length was used as a reference.

2.1.1. Lab Study

In the lab study, data from 27 amateur runners (21 male, 6 female) were recorded. The dataset included runners with different strike types. Six of the subjects were forefoot/midfoot runners, and 21 subjects were rearfoot runners. The classification of the strike type was based on the definitions of Altman et al. [17]. Further anthropometric data can be found in Table 1. Before data acquisition, all subjects were informed about the related risks and gave written consent to participate in the study and for the collected data to be published.

Table 1. Anthropometric data of subjects participating in the lab study.

Parameter	Mean $\pm$ Standard Deviation
Age (years)	24.9 ± 2.4
Shoe size (U.S.)	9.3 ± 1.4
Height (cm)	178.6 ± 8.0

The subjects were equipped with running shoes in matching sizes (Response Cushion 21, Adidas AG, Herzogenaurach, Germany), as depicted in Figure 1a. This model had a cavity in the right shoe midsole for the placement of a sensor. We cut another cavity of the same size at the same location into the left shoe midsole to be able to acquire data from both the left and the right shoe in order to record more data for the training and evaluation of the algorithms. The specific IMU we used was

the miPod sensor [18]. The accelerometer of the sensor was configured with a range of ± 16 g and the gyroscope with a range of $\pm 2000\,\frac{\circ}{s}$, and data were sampled with a frequency of $f_s = 200$ Hz and a resolution of 16 bit. Before each data acquisition, the IMUs were calibrated using the calibration procedure introduced by Ferraris et al. [19]. Figure 1a depicts the orientation of the sensor in the sole of the running shoe: x points in the lateral direction, y in the dorsoventral direction, and z in the craniocaudal direction.

As the gold standard for velocity and stride length, we used a motion capture system (Vicon Motion Systems Inc., Oxford, UK) with 16 infrared cameras and recorded data with a sampling rate of $f_s = 200$ Hz. A submodel of the marker setup introduced by Michel et al. [20] containing six markers on each shoe (see Figure 1a) was used. The marker on the heel (for rearfoot runners) and the lateral sided toe marker (for forefoot/midfoot runners) were used to extract strides. Depending on the strike type, minima in the trajectory of the corresponding markers were used to label initial ground contacts [21]. The IMUs and the motion tracking system were synchronized using a wireless trigger [22], which was connected to light barriers (S40 Series, Datalogic, Bologna, Italy). The light barriers triggered the start and the end of the recording for each trial in both systems. Using the described synchronization technique, we were able to match strides in the motion capture gold standard data to strides in the IMU signal.

The subjects were asked to run various trials with different velocities in the range of 2–6 m/s. We defined these velocity ranges to cover a wide range of relevant running velocities. As the capture volume was restricted to 6 m, and the stride length varied depending on the running velocity, different numbers of strides were recorded for the different running velocities. We recorded five additional trials for the two high velocity ranges to increase the number of captured strides. The velocity ranges and number of trials recorded can be found in Table 2. The subjects were asked to accelerate before and keep the pace within the capture volume. We measured the velocity at the beginning of the motion capture system volume using the above-mentioned light barriers used for synchronization. The velocity measured by the light barriers was used to ensure that a sufficient number of trials were recorded within each velocity range, for each subject. If necessary, the subjects were instructed to run faster or slower in order to ensure the defined number of trials in each velocity range. The ground truth value for each stride's velocity v_{ref} was computed from the motion capture reference as:

$$v_{ref} = \frac{d_{ref}}{t_{ref}} = \frac{d_{ref} \cdot f_s}{N_{stride}} \tag{1}$$

where d_{ref} is the stride length obtained by the difference of the positional data obtained by the motion capture system between two consecutive initial ground contacts, t_{ref} the corresponding reference stride time, N_{stride} the number of samples in between two consecutive initial ground contacts, and f_s the sampling rate. Figure 1b illustrates the setup and running path of the subjects during the lab data recording. Overall, 2377 strides were recorded during the lab study for the evaluation of the algorithms.

Table 2. Number of trials and recorded strides per velocity range in the lab study.

Velocity Range	# of Trials	# of Strides
2–3 m/s	10	921
3–4 m/s	10	558
4–5 m/s	15	544
5–6 m/s	15	354

(**a**) Shoe   (**b**) Setup runway
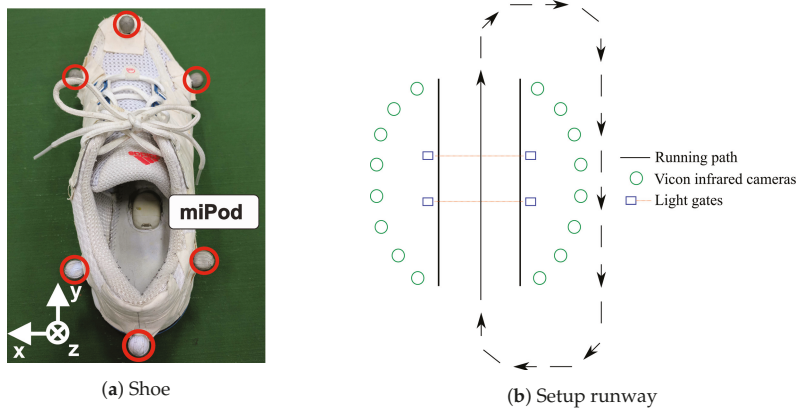
Figure 1. (**a**) Shoe equipped with a miPod sensor and the marker setup. The IMU is located within the sole of the running shoe. The marker setup allowed for a computation of velocity and stride length. (**b**) Illustration of reference system setup. The subjects ran through the capture volume of the motion tracking system, created by 16 infrared cameras, and looped back around.

2.1.2. Field Study

The goal of the field study was to evaluate the algorithm performance regarding total distance in a real-world scenario. We recorded twelve subjects who performed a self-paced 3.2-km run by completing eight rounds on a 400-m tartan track. We used this setup to be able to obtain a reference distance accurately.

The equipment (IMUs, shoes) and settings were the same as described in the lab study (Section 2.1.1) to enable a direct comparison of the results. The subjects participating in the field study were not part of the lab study. Additionally, we recorded GPS data using a smartphone (Galaxy S8, Samsung Inc., Seoul, South Korea) and the fitness application Strava (Strava, Strava Inc., San Francisco, CA, USA), which is, with 136 million uploaded runs in 2017, one of the most popular fitness apps worldwide [23]. It also has the capability to export the GPS track in the GPX-format [24] allowing for a computation of the distance of the running track. The accuracy of the implemented algorithms can be compared to the GPS data for the total distance of the run. We used the great circle distance to compute the total distance of the GPS measurements [25]. Our computed total distance from the exported GPX file matched the distance that Strava provided via its services. Thus, we could compare the accuracy of the different algorithms to state-of-the-art running platform distance measurements.

2.2. Algorithms

In this section, the algorithms will be described in detail. The section starts with the stride segmentation algorithm, which is required for all algorithms, except the acceleration-based algorithm, which includes a different approach to segment steps. Afterwards, the algorithms are described in the following order: *Stride time*, (foot) *Acceleration*, (foot) *Trajectory* estimation, and *Deep Learning*.

2.2.1. Stride Segmentation

The first step in the IMU signal processing for velocity and distance calculation was the stride segmentation. In this step, single strides were extracted from the continuous IMU data stream with a threshold-based algorithm. Common algorithms use the distinct peaks in the acceleration signal in the dorsoventral direction $a_y[n]$ during initial ground contact to mark the beginning of a stride [26]. We enhanced this idea and used the beginning of the distinctive peak to mark the beginning of a stride.

This procedure is valid, as the ground already exerts a force to the IMU at the time instance of the peak in the acceleration signal. Using the peak itself would mean to mark a point in time that is part of the ground contact. To find the sample before the acceleration peak, we first differentiated the acceleration signal in the dorsoventral direction $a_y[n]$ and consecutively squared the resulting value to obtain a signal $H[n]$ with amplified peak values.

$$H[n] = (a_y[n] - a_y[n-1])^2 \tag{2}$$

In the signal $H[n]$, the maxima were detected by comparing them to an empirical threshold (empirical threshold: $H[n] > 1000\left[\left(\frac{m}{s^2}\right)^2\right]$). For every detected maximum, the onset of the rise of $H[n]$ was determined by setting all values below the threshold to zero and looking for the index of the last non-zero value in $H[n]$ before the detected maximum. This index n_{IC} was a potential candidate for an initial ground contact. To eliminate false detections, we added a detector for the swing phase prior to the peak in the acceleration signal. The swing phase detector computed an integral of $a_y[n]$ backwards from the first detected non-zero value until the first zero-crossing. This is the point in time where the foot starts decelerating during the swing phase. The integral value $S[n_{IC}]$ for the initial ground contact candidate n_{IC} was computed as:

$$S[n_{IC}] = \sum_{n=n_{ZC}}^{n_{IC}} \frac{1}{f_s} a_y[n] \tag{3}$$

In this equation, n_{ZC} corresponds to the index of the zero crossing marking the start of the deceleration. If the integral value $S[n_{IC}]$ exceeds an empirically-set threshold (empirical threshold: $S[n_{IC}] < -3\left[\frac{m}{s}\right]$), a swing phase is detected, and thus, the index of the first non-zero value before the acceleration peak is labeled as an initial ground contact. The described stride segmentation is depicted in Figure 2.
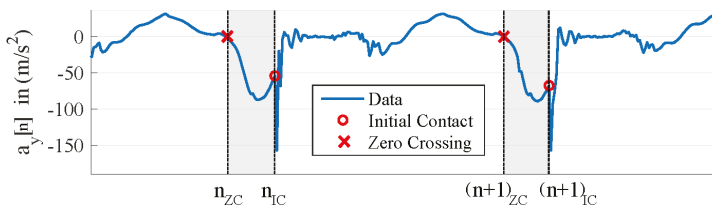
Figure 2. Example for the stride segmentation. The plot shows the acceleration signal in the dorsoventral direction $a_y[n]$, the detected initial ground contact n_{IC}, and the beginning of the swing phase (zero crossing n_{ZC}) to confirm the stride candidate. The marked area depicts the integration area for the swing phase detection.

2.2.2. Stride Time

Cavanagh et al. described the relationship between running velocity, stride length, and stride frequency [5]. Stride frequency is an inverse measure of the stride time and describes the number of strides per minute. They showed that runners can increase their running velocity either by increasing their stride length or by increasing their stride frequency, thus decreasing the stride time. For lower velocities, runners tend to increase the stride length, while for higher velocities, they tend to increase the stride frequency. Thus, both the stride time and stride length have no linear dependency on running velocity. Furthermore, it has to be noted that runners control their velocity individually. The stride length and therefore the velocity also depend on other parameters like the gender and the height of the runner. Male runners show greater stride lengths compared to female runners. The stride length increases with the body height [5].

We used these biomechanical relations to build an algorithm that estimates stride length and velocity. Cavanagh et al. [5] provided averaged values for the non-linear correlation between the stride time t_{stride} and a relative stride length $d_{stride,rel}$, which is calculated by dividing the absolute stride length d_{stride} by the runner's height h. We looked for further publications describing this relationship and came up with two step functions for males and females that discretized the underlying non-linear relationship between stride time and stride length for each gender. The definition of the functions for both male and female runners can be found in Table 3.

Table 3. Definition of step functions for the relative stride length $d_{stride,rel}[t_{stride}]$ for (a) male and (b) female runners for the *Stride time* algorithm.

(a) Male			(b) Female		
$t_{\textbf{stride}}$ **(s)**	$d_{\textbf{stride,rel}}$	**Reference**	$t_{\textbf{stride}}$ **(s)**	$d_{\textbf{stride,rel}}$	**Reference**
$0.800 < t_{stride}$	0.830	[5,27]	$0.800 < t_{stride}$	0.826	[27,28]
$0.748 < t_{stride} \le 0.800$	1.080	[27,29]	$0.735 < t_{stride} \le 0.800$	1.110	[5,27,29]
$0.720 < t_{stride} \le 0.748$	1.260	[5]	$0.720 < t_{stride} \le 0.735$	1.260	[5]
$0.713 < t_{stride} \le 0.720$	1.330	[5]	$0.704 < t_{stride} \le 0.720$	1.400	[5]
$0.706 < t_{stride} \le 0.713$	1.410	[5]	$0.667 < t_{stride} \le 0.704$	1.500	[5]
$0.698 < t_{stride} \le 0.706$	1.490	[5]	$0.607 < t_{stride} \le 0.667$	1.720	[5]
$0.694 < t_{stride} \le 0.698$	1.590	[5]	$0.578 < t_{stride} \le 0.607$	1.920	[5]
$0.687 < t_{stride} \le 0.694$	1.740	[5]	$0.500 < t_{stride} \le 0.578$	2.080	[5]
$0.678 < t_{stride} \le 0.687$	1.880	[5]	$t_{stride} \le 0.500$	2.170	[30]
$0.664 < t_{stride} \le 0.678$	1.960	[5]			
$0.649 < t_{stride} \le 0.664$	2.015	[5]			
$0.500 < t_{stride} \le 0.649$	2.060	[5]			
$t_{stride} \le 0.500$	2.170	[30]			

The stride time t_{stride} was obtained by dividing the number of samples of one stride N_{stride} by the sampling frequency f_s.

$$t_{stride} = \frac{N_{stride}}{f_s} = \frac{(n+1)_{IC} - n_{IC}}{f_s} \qquad (4)$$

N_{stride} was computed by subtracting the indices of two consecutive initial ground contacts $(n+1)_{IC}$ and n_{IC} obtained from the stride segmentation algorithm. After obtaining the relative stride length $d_{stride,rel}$ of the runner based on the gender and the stride time, the absolute stride length d_{stride} was computed by multiplying $d_{stride,rel}$ from the table and the runner's height h in meters.

$$d_{stride} = h \cdot d_{stride,rel} \qquad (5)$$

The running velocity v_{stride} was then calculated using the stride time and the stride length.

$$v_{stride} = \frac{d_{stride}}{t_{stride}} \qquad (6)$$

Thus, the *Stride time* algorithm is solely based on the stride time. Gender and body height are usually known in all applications.

2.2.3. Acceleration

The *Acceleration* method introduced in [6] uses only acceleration data for step segmentation and the computation of stride length and stride velocity. The method correlates the velocity of the foot (and thus, the subject) with the acceleration during the swing phase of the foot. It consists of three different algorithmic steps: (1) a continuous calculation of an integration value with a strong correlation to the movement velocity, (2) a stride segmentation based on initial ground contacts to determine the swing phase of the foot, and (3) a regression model to translate the continuous integration value to the velocity value.

We used the step segmentation algorithm from the cited publication due its applicability to movements other than running [6]. The inputs to the processing pipeline were the sampled triaxial acceleration signals from the foot sensor a_x, a_y, and a_z. After smoothing the input signals using a sliding window filter, the integration value ι was calculated as a multi-step absolute averaging across all directional components with:

$$\iota[n] = \frac{1}{L+1} \sum_{i=0}^{L} \sum_{d=x,y,z} |s_d[n-i]|,$$ (7)

where L is the window length, which is expected to be the same as the duration of the foot swing phase.

Individual strides were determined in the smoothed dorsoventral acceleration signal s_y using a peak detection process in combination with two knowledge-based parameter thresholds. The goal was to detect and isolate the high impact response of initial ground contact in the smoothed signal [6]. Each time a valid stride was detected by the stride segmentation algorithm, the average velocity per stride v_{stride} was determined based on a second degree polynomial regression function:

$$v_{stride} = A + B \cdot \iota[n_{IC}] + C \cdot \iota[n_{IC}]^2,$$ (8)

where the constants A, B, and C are derived during a regression model training phase where known reference velocity observations are matched to velocity integration values using parametric regression analysis. A trained regression model can be observed in Figure 3.
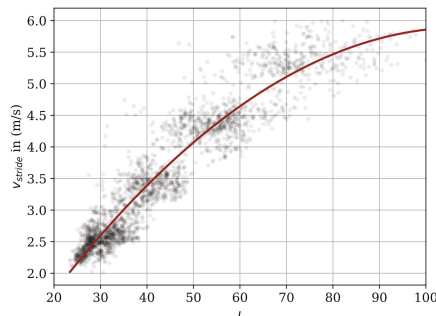
Figure 3. Polynomial function of second degree (red line) that relates the velocity integration value ι to the reference velocity values v_{stride} (grey dots).

2.2.4. Trajectory

Based on the foot trajectory, the stride length and stride velocity can be deduced. The trajectory of the sensor during running can be computed using an extended Kalman filter approach or using dedrifting techniques. We applied dedrifting techniques due to two reasons: Firstly, Bailey et al. [12] showed that the results for the mean step velocity of the two techniques did not differ significantly with respect to accuracy (extended Kalman filter: 0.03 ± 0.02 m/s, linear dedrifting: 0.0 ± 0.03 m/s). Secondly, the same authors showed in a different article that a sampling rate of more than 250 Hz is required for an extended Kalman filter approach [31]. For embedded use cases (e.g., a smart shoe scenario), low sampling rates are beneficial from an energy perspective. In gait analysis, the linear dedrifting technique showed promising results for a lower sampling rate of 200 Hz [7].

Trajectory reconstruction algorithms based on linear dedrifting consist of four steps, as depicted in Figure 4, and have both the triaxial accelerometer and the triaxial gyroscope measurements as an input. In the following paragraphs, the four algorithmic steps will be explained in detail. Orientation

is computed by integrating the gyroscope measurements, and the position is obtained by integrating the accelerometer measurements.

| Midstance detection | Orientation estimation | Gravity removal | Dedrifted integration |
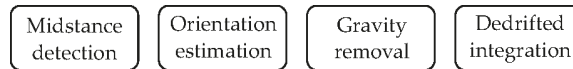
Figure 4. The four steps of the algorithm for the trajectory reconstruction based on linear dedrifting.

Midstance detection: A common problem with computing the trajectory from IMU measurements is the drift of the sensors introduced by noise in the acceleration and the angular rate measurements. This drift is limited by using zero velocity updates [32]. The idea behind these updates is to reinitialize the position and the orientation of the sensor for every stride. By applying that technique, absolute position in space is lost; however, the individual stride parameters can be computed more accurately. The reason for the higher accuracy lies in the integration of shorter durations and thus a smaller accumulated error. The point in time for the reinitialization of the stride values originates from gait analysis and is the midstance phase during a stride cycle. At this point in time, the foot has its lowest velocity, and the orientation of the foot is known, because during midstance in gait, the foot is expected to be flat on the ground. Thus, it can be assumed that the orientation of the sensor can be computed statically using the acceleration measurement. This allows the initialization of the position and velocity to zero and the orientation with respect to gravity. To find midstance, we computed the minimum gyroscopic energy after initial ground contact [32] in a 250-ms time interval. The duration of this time interval is the average time of the stance phase for velocities up to 6 m/s [33]. Hereafter, the trajectory reconstruction will be performed on strides segmented from midstance to midstance.

Orientation estimation: After initializing the orientation based on the accelerometer measurement during midstance, the orientation of the sensor was computed using the gyroscope measurements. This step is necessary to calculate the orientation of the sensor so that gravity can be removed, which is an essential step for the computation of the position in space from the acceleration signal. For the orientation computation, we used the same quaternion integration approach as described by Rampp et al. [7].

Gravity removal: After the orientation estimation, gravity was removed. Without this removal, the gravitational acceleration of 9.81 m/s^2 would be integrated additionally into the acceleration caused by running, which would lead to a large error over the duration of a stride. To remove gravity, we used the orientation of the sensor obtained by the gyroscope integration to rotate the acceleration measured in the sensor coordinate system to the world coordinate system. In the world coordinate system, we subtracted the gravitational acceleration from the measured acceleration.

Dedrifted integration: The last step to come up with the full trajectory of the stride was to compute the position of the sensor by a double integration of the gravity removed acceleration. The first integration computed the velocity of the sensor over time, followed by the second integration, which resulted in the position of the sensor over time. Despite the gravity removal, there was still noise in the acceleration signal, causing drift in the results. This drift was reduced by dedrifting the velocity signal obtained after the first integration. The core idea behind dedrifting is the fact that we assume the velocity to be zero during midstance. For every stride, we fit a linear function in the velocity signal for all three directions, which was determined by the first and last velocity value of the stride. To dedrift the velocity signal, we subtracted the linear function from the integrated velocity signal, which enforced the velocity to be zero for both the first and the second midstance. This process is depicted in Figure 5.

(**a**) Velocity before dedrifting

(**b**) Dedrifted velocity
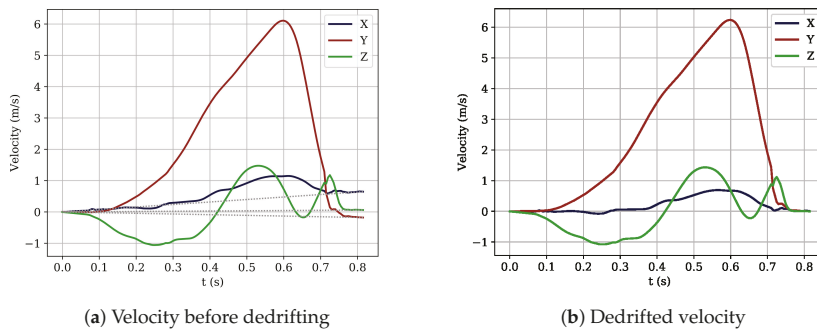
Figure 5. Visualization of the dedrifting method that ensures that the velocity during the second midstance is zero. (**a**) Velocity signal before dedrifting. The grey doted linear function is fit between the first and last point of the stride (midstance). (**b**) Velocity signal after dedrifting.

Calculation of stride length and velocity: From the position of the sensor in space obtained after integrating the dedrifted velocity signal, the stride length d_{stride} and the average stride velocity v_{stride} were computed. The stride length was calculated as the L2-norm of the position in space at the index of the second midstance. Velocity was calculated by dividing stride length by stride time.

2.2.5. Deep Learning

After outperforming conventional methods in various other fields like speech recognition, visual object recognition, and object detection [34], the methodology of deep learning started to become more and more popular for IMU data processing. Hannink et al. [16] introduced a deep convolutional regression network for calculating the stride length from raw IMU data in geriatric patients. The network learned a model for stride length regression based on raw IMU data without any domain knowledge. In this work, we used an adapted architecture for the stride length computation in running gait, which is depicted in Figure 6. It consisted of two convolutional layers, two max pooling layers, one flattening layer, and two fully-connected layers. For the implementation of the architecture, we used Keras [35] with a TensorFlow backend [36].
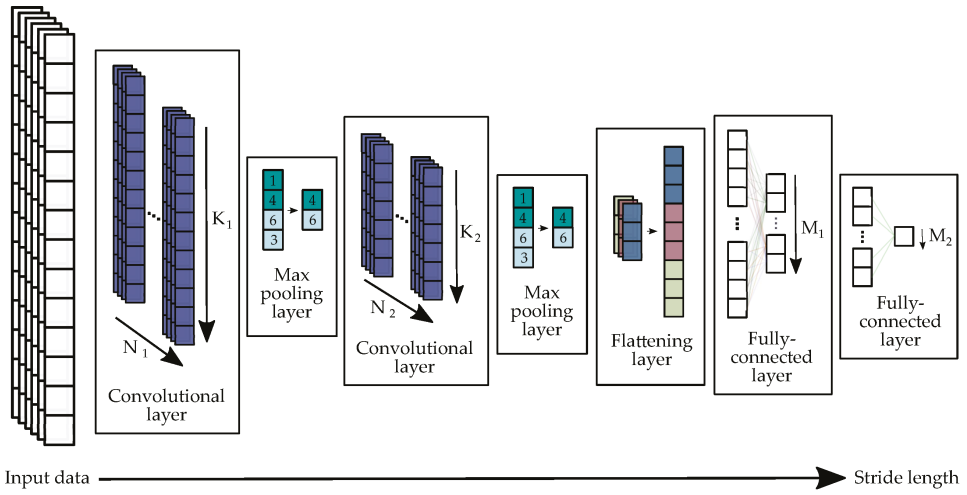
Figure 6. Architecture of the convolutional neural network for stride length regression based on the raw 6D-IMU signal. For the first convolutional layer, we used $N_1 = 32$ filter kernels of kernel length $K_1 = 30$. The second convolutional layer consisted of $N_2 = 16$ filter kernels of kernel length $K_2 = 15$. The first fully-connected layer had $M_1 = 128$ outputs that served as input to the second fully-connected layer, which had only a $M_2 = 1$ output. This output represented the computed stride length.

Before feeding data into the network, the segmented 6D-IMU data of a single stride were zero padded to 200 samples to assure a constant number of samples as an input to the network. One convolutional layer consisted of N convolution filters. The N outputs of a convolutional layer $O^{(j)}$ with $j = 1 \ldots N$ are called feature maps and were computed by the convolution of the six IMU input channels x_c with $c = 1...6$ with the filter kernel $\phi_c^{(j)}$ of length K, adding biases $b_c^{(j)}$ and finally applying a ReLU activation function:

$$O^{(j)} = \text{ReLU}\Big(\sum_{c=0}^{6} (\Phi_c^{(j)} \times x_c + b_c^{(j)}) \Big) \tag{9}$$

This formula has to be applied for all $j = 1 \ldots N$ filters to produce N feature maps $O^{(j)}$ after each convolutional layer. Thus, the two tunable parameters in the convolutional layers are the number of kernel coefficients K and the number of filters N. In the first convolutional layer, the kernel size was $K_1 = 30$ and the number of filters $N_1 = 32$. In the second convolutional layer, the kernel size was $K_2 = 15$ and the number of filters $N_2 = 16$ filters. After each convolutional layer, the resulting feature map was fed into a max pooling layer, which downsampled the resulting feature map by a downsampling factor of two by taking the maximum in non-overlapping windows of size two.

After the second max pooling layer, the feature map was flattened to produce a one-dimensional feature list that can be fed into the fully-connected layers. Thus, the flattening layer appended the N_2-dimensional output of the second max pooling layer after each other into one feature list. The two fully-connected layers at the end of the architecture computed a weighted sum of all $k = 1 \ldots N_f$ input features φ_k of the one-dimensional feature vector with weights $w_{k,j}$ and added biases b_k. A ReLU function again activated the positive features.

$$F_j = \text{ReLU}\Big(\sum_{k=0}^{N_f-1} (w_{k,j} \cdot \varphi_k + b_{k,j}) \Big) \tag{10}$$

The outputs of the fully-connected layers were feature lists F_j with $j = 1 \ldots M$, where M describes the number features. In our architecture, the first fully-connected layer had $M_1 = 128$ output features. The second fully-connected layer had only $M_2 = 1$ output feature, which was the resulting target value. In our implementation, the regressed target value was the stride length.

To prevent overfitting, we also added a dropout layer to our network [37]. The dropout layer was stacked between the two fully-connected layers and dropped 30% of the neurons. During training, we fed the data into the network in five epochs with a batch size of 16. We trained the network both for the stride length and for the velocity. The network with the stride length as the output outperformed the velocity approach and was therefore used for the evaluation in this publication. Thus, the velocity v_{stride} for the *Deep Learning* approach was computed by dividing the stride length d_{stride} obtained from the neural network by the stride time t_{stride} obtained from the stride segmentation.

2.3. Evaluation

2.3.1. Lab Study

The results of the lab study dataset will be evaluated using the mean error (ME) and standard deviation (Std), the mean absolute percentage error (MAPE), and the mean absolute error (MAE). We provide all these measures to make our results comparable to prior studies.

For the evaluation of the *Acceleration* and *Deep Learning* algorithms, we used leave-one-subject-out cross-validation to prevent overfitted results. We also show Bland–Altman plots [38] to visualize the results.

2.3.2. Field Study

For the evaluation of the 3.2-km field study dataset, we used the MAE to evaluate the total distance of the runs. After segmenting the strides and calculating the stride lengths for each stride, we accumulated the single stride lengths and compared them to the ground truth value of 3200 m. The reason for choosing the MAE for this evaluation was the fact that the absolute deviation of the ground truth value is of great importance to runners. For the *Acceleration* and *Deep Learning* algorithms, we computed the regression models based on the lab study dataset. Due to having different subjects participating in the lab and the field study, the results were not overfitted. The GPS measurements of the total distance of the individual runs were also evaluated by comparing them to the gold standard value of 3.2 km.

3. Results

3.1. Lab Study

Table 4 depicts the mean errors and standard deviations for both stride velocity and stride length of the four different algorithms for the lab study dataset. The results were averaged over all strides in the lab study dataset. The results show that the *Trajectory* algorithm performed best considering both the ME $\pm$ Std and the MAE.

Table 4. Mean error (ME) and standard deviations (Std), mean percentage error (MAPE), and mean absolute error (MAE) of stride length and average velocity per stride of the four algorithms for the lab study dataset.

Parameter	Error Measure	Stride Time	Acceleration	Trajectory	Deep Learning
Velocity	ME ± Std (m/s)	0.209 ± 0.782	0.005 ± 0.350	0.028 ± 0.252	0.055 ± 0.285
	MAPE (%)	17.2	7.7	3.5	5.9
	MAE (m/s)	0.622	0.272	0.133	0.216
Stride length	ME ± Std (cm)	17.7 ± 57.3	−0.5 ± 25.6	2.00 ± 14.1	2.5 ± 20.1
	MAPE (%)	17.1	7.9	2.8	5.9
	MAE (cm)	45.2	19.9	7.6	15.3

Figure 7 shows the results of the stride length for the different velocity ranges. The MEs of the *Deep Learning* algorithm increased with higher velocities. The *Trajectory* showed lower MEs for the three slower velocity ranges than for the highest velocity range. The *Acceleration* algorithm showed small errors from 3–5 m/s. Its performance dropped for the outer velocity ranges from 2–3 m/s and from 5–6 m/s. The *Stride time* algorithm worked well for the velocity range of 2–3 m/s and 5–6 m/s; however, it showed large errors of more than 40 cm for the other velocity ranges.
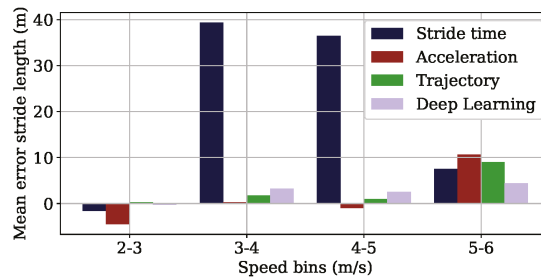
Figure 7. Mean error of the stride length of the four different algorithms for the different velocity ranges the subjects ran in the lab study.

Figure 8 shows the Bland–Altman plots for both the stride length and the average velocity per stride for the lab study dataset. The results are color coded into the velocity ranges presented in Table 2. The *Trajectory* algorithm performed well for velocities up to 5 m/s. For the high velocity range, larger errors could be observed. The *Stride time* algorithm performed worst and showed a linear error distribution in the Bland–Altman plots. In the *Acceleration*, *Trajectory*, and *Deep Learning* plots for stride length, we see the samples of the different velocity ranges overlapping. This overlap is not visible in the velocity plots.

(**a**) Algorithm: *Stride time*; metric: stride length

(**b**) Algorithm: *Stride time*; metric: velocity

(**c**) Algorithm: *Acceleration*; metric: stride length

(**d**) Algorithm: *Acceleration*; metric: velocity

(**e**) Algorithm: *Trajectory*; metric: stride length

(**f**) Algorithm: *Trajectory*; metric: velocity

(**g**) Algorithm: *Deep Learning*; metric: stride length

(**h**) Algorithm: *Deep Learning*; metric: velocity
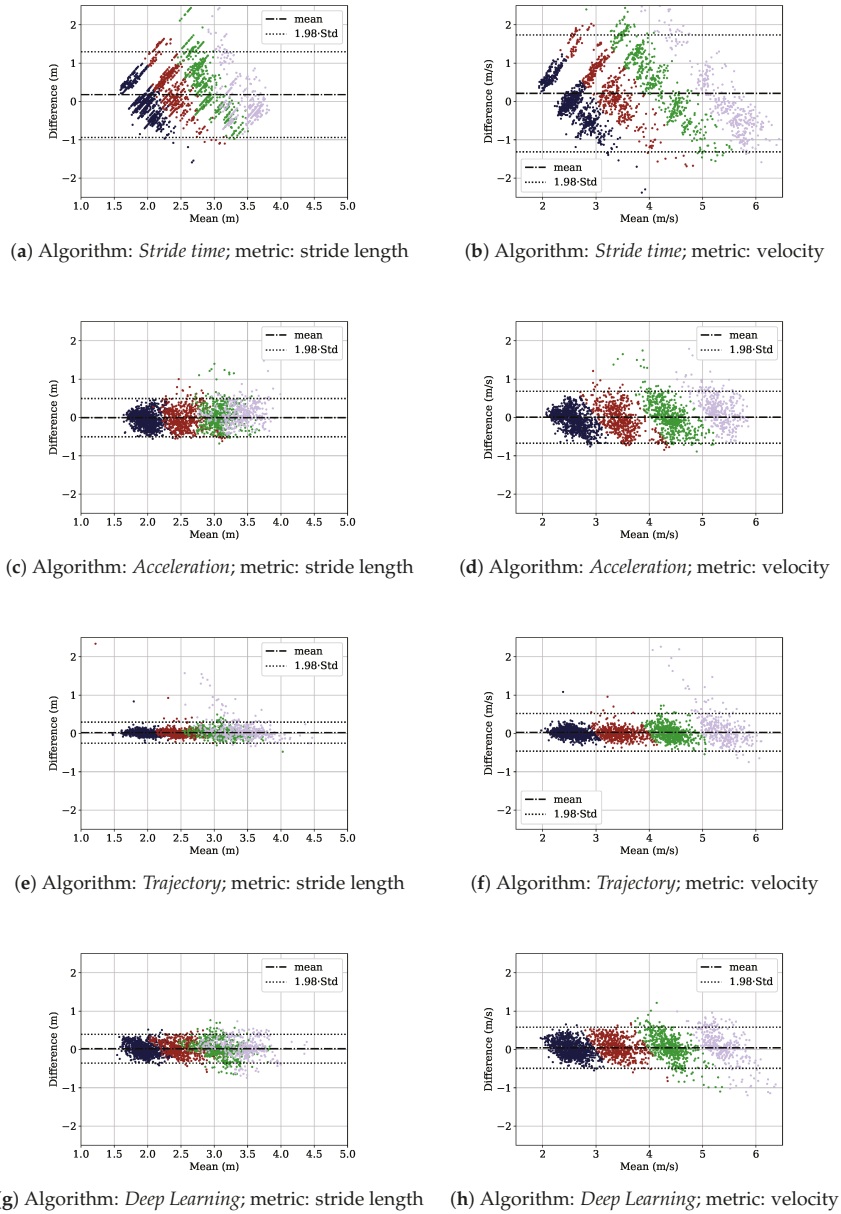
Figure 8. Bland–Altman plots for stride length and velocity for the four algorithms. Each row contains the metrics for one algorithm. The individual samples are color coded depending on the velocity bin of the sample: 2–3 m/s blue, 3–4 m/s red, 4–5 m/s green, 5–6 m/s purple. The dotted-dashed horizontal lines depict the mean error and the dotted horizontal line the 95% confidence interval.

3.2. Field Study

Figure 9 shows the MAE of the total running distance for the field study dataset, both for the algorithms and GPS-based estimation. The figure indicates that the *Trajectory* algorithm performed best with a MAE of 94.0 m. The error was comparable to that of the GPS-based estimate (82.1 m).
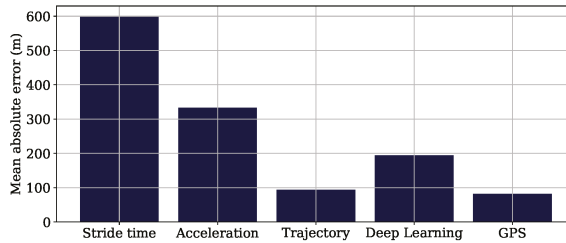
Figure 9. Mean absolute error of the 3.2-km run for the four different algorithms and GPS.

4. Discussion

Firstly, we will compare our results to existing literature. Afterwards, we will discuss the results of the lab study including a detailed evaluation of the individual algorithms with respect to their accuracy and their advantages and disadvantages in a smart shoe scenario. Special emphasis will be placed on the number of sensors that are needed to run the algorithms and the underlying power consumption of these sensors. Finally, the results of the field study on the tartan track will be discussed.

4.1. Comparison to Existing Literature

Different papers already evaluated the stride length or the velocity of single strides. Three of these papers are listed in Table 5. These three publications used similar approaches to ours: Bailey et al. [12] used a trajectory approach using a linear dedrifting technique; Gradl et al. [6] used the described acceleration approach; and Hannink et al. [16] a DCNN approach.

Table 5. Results of other publications related to stride length and velocity calculation.

	Gait Type	# Subjects	# Strides	Parameter	Error Measure	Result
Bailey et al. [12]	Running	5	1800	Velocity	ME	0.04 ± 0.03 m/s
Gradl et al. [6]	Running	9	795	Velocity	MAPE	$6.9 \pm 5.5\%$
Hannink et al. [16]	Walking	101	~1392	Stride length	ME	0.01 ± 5.37 cm

With respect to the standard deviation, the results of the trajectory implementation of Bailey et al. [12] are better than our results (Table 4). They also evaluated 1800 strides; however, these strides only originated from running velocities ranging from 2.3–3.4 m/s. We also evaluated our results for this velocity range and obtained an error of 0.004 ± 0.107 m/s. We observe that our standard deviation is still higher than the standard deviation reported from Bailey et al. One reason for that might be the higher number of different runners with different running styles who participated in our study. Furthermore, their study was conducted on a treadmill. On a treadmill, the variability of different strides at a given velocity is lower and does not reproduce overground running kinematics [39].

The errors reported by Gradl et al. [6] were obtained on a smaller database than the one presented in this paper. Thus, our worse results are due to the higher variability in our dataset, which the second degree polynomial could not appropriately approximate.

The results of Hannink et al. [16] were evaluated for gait in geriatric patients. Hence, there is a general difference in the stride patterns, causing differences in the results. Further differences between the setup of our network architecture and study population are listed and discussed in the following section.

4.2. Lab Study

In this section, we will discuss the results of the lab study for each algorithm in detail with respect to their advantages/disadvantages and the number of sensors needed for their implementation.

Stride time: The *Stride time* algorithm leads to the lowest accuracy for the lab study dataset. Even though stride time and stride length relative to the subject's height correlate non-linearly, the correlation does not seem to be high enough to compute velocity and stride length accurately. The low correlation is also visible in Figure 10. The gray dots are the relative stride length values obtained from the lab study dataset, and the red line is the step function for male subjects defined in Table 3a. We see that the step function does not approximate the underlying data accurately. The standard deviation of the relative stride length within a certain stride time range (e.g., $0.748 < t_{stride} \leq 0.800$) of the step function is high. This is due to the fact that velocity is controlled by stride frequency and stride length. The *Stride time* algorithm cannot handle that fact, as it only depends on stride frequency.
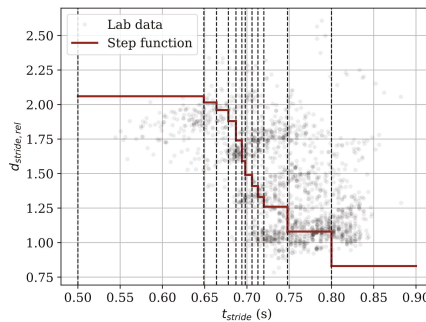
Figure 10. Visualization of the correlation between the stride time t_{stride} and the relative stride length $d_{stride,rel}$ for male subjects. The light gray dots depict the data obtained from the field study, whereas the red curve and the black dashed lines visualize the step function obtained from literature and implemented in the *Stride time* algorithm.

In the Bland–Altman plots for the stride length metric (Figure 8), the other three algorithms showed overlapping sample clouds. This indicates that people increased their velocity both by increasing their stride length and by decreasing their stride time in higher velocities. The other algorithms are capable of dealing with this effect due to the fact that the sample clouds are separated in the Bland–Altman plots of the velocity metric. This is not observable in the plots for the *Stride time* algorithm. Thus, the other algorithms can deal better with the velocity control via stride frequency and stride length.

Furthermore, we want to discuss the shape of the *Stride time* algorithm's Bland–Altman plots briefly. The long diagonal lines in the plots (Figure 8b) originate from the steps in the step function introduced in Table 3. One line belongs to one stride time range. The small deviations within the diagonals originate from the different body heights. We observed that for some stride time ranges, the gold standard velocity ranged from 2–6 m/s (color coded within one diagonal), showing that the stride time ranges of the step function obtained from the literature do not generalize well. Furthermore,

the relative stride lengths presented in Table 3 are averaged over specific study populations. Even if a subject controls its stride frequency in the exact same manner as encoded by the stride time ranges of the step function, the resulting stride length could be incorrect due to an incorrect relative stride length.

Despite the algorithm's low accuracy, an advantage of the stride time algorithm is that it can be implemented very energy efficiently. In the case of an IMU scenario, only a stride segmentation is necessary to compute the stride time. The stride segmentation presented in this paper only needs the sampling of the acceleration in the dorsoventral direction; thus, a 1D-accelerometer would be sufficient. In fact, strides could be segmented without an IMU using sensors such as piezo-electric switches to detect the ground contact [40].

Acceleration: The plot with the ME for the different velocity ranges in Figure 7 shows that the *Acceleration* algorithm works better for the the two velocity ranges from 3–4 m/s and 4–5 m/s. In addition, the Bland–Altman plots in Figure 8 show outliers especially for the highest velocity range for both the stride length and the average velocity. The reason for that can be observed in Figure 3, where we see that the second degree polynomial used to map the velocity integration value ι to the velocity value approximates the reference data better for the velocity range from 3–5 m/s and especially not well for the highest velocity range. This can be explained by the spread of the underlying data being too large to be represented by the polynomial.

However, the *Acceleration* algorithm outperforms the *Stride time* algorithm and shows comparable performance to the *Deep Learning* algorithm for the velocity range of 3–4 m/s. The advantage of the *Acceleration* algorithm over the better performing *Trajectory* and slightly better performing *Deep Learning* algorithm is its energy efficiency. For the computation of the stride length and the velocity, only a triaxial accelerometer needs to be sampled. Sampling only an accelerometer consumes less energy than sampling the gyroscope or sampling both sensors. For example, for the MPU9250 from InvenSense, the supply current needed for sampling only the accelerometer is less than 15% of the current needed for sampling both the accelerometer and the gyroscope [41]. Furthermore, the sampling rate can be further reduced for the *Acceleration* algorithm [6]. We also tested the reduction of the sampling rate for the lab study dataset and observed that a reduction to 60 Hz does not affect the accuracy of the algorithm. With such a low sampling rate, the energy consumption can be further reduced. Another advantage of the algorithm is its generalizability and its applicability to other movements like side stepping [6].

Foot trajectory: The *Trajectory* algorithm performs best for the lab study dataset. Especially for velocities up to 5 m/s, the algorithm achieves a ME of less than 0.012 m for the stride length and 0.014 m/s for the average velocity. For velocities higher than 5 m/s, the accuracy drops. In the Bland–Altman plots (Figure 8e,f), outliers for this velocity range are visible. The zero-velocity update based on the detection of the minimum energy in the gyroscope signal is error prone for such high velocities. The foot has no real zero-velocity phase and is always in motion. Thus, the underlying zero-velocity assumption does not hold. One way how to improve this algorithm is to propose a better solution for the initial condition when applying it to higher running velocities. Future work could evaluate whether a regression model based on the velocity during the swing phase would be a better initial condition.

For the *Trajectory* algorithm, we were also interested in the applicability of the zero velocity update for the different strike types due to the foot never being flat on the ground for forefoot runners. Hence, we also evaluated the accuracy of the *Trajectory* algorithm for the different strike types. The violin plots for forefoot and rearfoot runners are depicted in Figure 11. The plots show that the MEs do not differ significantly for the two strike types. However, the standard deviation is higher for forefoot runners. The low ME both for the forefoot and the rearfoot strike type can be explained by the fact that we align the foot during the zero velocity phase with gravity. The higher standard deviation originates in the more dynamic nature of the forefoot running style. Thus, the zero velocity phase cannot be detected accurately, which results in higher errors.
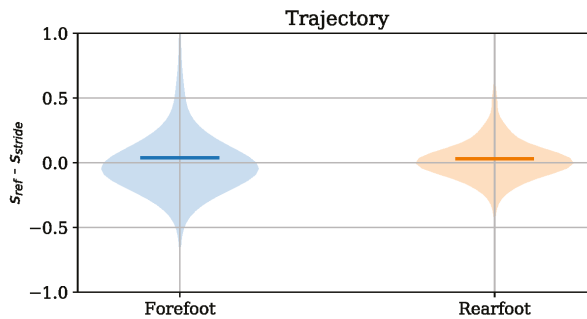
Figure 11. Violin plots of the error ($s_{ref} - s_{stride}$) in the velocity computation for forefoot strikers and rearfoot strikers.

An advantage of the *Trajectory* algorithm is that it provides more information about the stride than the velocity and the stride length. During the computation of these parameters, the orientation of the shoe in space is also calculated, which allows for a determination of other parameters like the sole angle, which defines the strike pattern or the range of motion in the frontal plane that is associated with pronation [42]. Furthermore, the algorithm uses solely signal processing and has no training phase, which makes it well applicable to unseen data. This holds for lower velocities and the transition to walking.

In terms of an embedded implementation and energy efficiency, the *Trajectory* algorithm needs both accelerometer and gyroscope data. Thus, it needs more energy than the *Stride time* and the *Acceleration* algorithm for acquiring 6D-IMU data.

Deep learning: The *Deep Learning* algorithm produced an ME of less than 0.095 m/s for the velocity and 0.104 m for the stride length for all velocity ranges in the lab study dataset. Compared to Hannink et al. [16], we reduced both the number of filters in the second convolutional layer and the number of outputs in the fully-connected layer, because the results using the identical structure yielded worse results for our use case. The differences in the architecture are listed in Table 6. Generally, the performance of the DCNN network is worse compared to the results reported in [16].

Table 6. Differences in the study setup and architecture presented in [16] from our DCNN implementation.

	N_2	M_1	# Parameters Trained	ME $\pm$ Std	Range Stride Length Data Set	# Training Samples
Hannink et al. [16]	64	1024	2,332,385	0.01 $\pm$ 5.37 cm	0.14–1.30 m	~1392
Our approach	16	128	85,425	1.3 $\pm$ 19.4 cm	1.22–4.84 cm	2377

We see that our approach needs less parameters due to the reduction of filters in the second convolutional layer and the smaller output number of the fully-connected layer. However, our results show a larger error. The reason for that might be a larger variation in our training data and the different strike pattern in running. The range of the target parameter of stride length is 3.62 m in the lab study of this work and 1.16 m in the dataset for geriatric patients of Hannink et al. [16]. The strike patterns in running differ significantly for forefoot and rearfoot runners, which also introduces more variation in the input data.

Besides, we observed that during training, the training errors and validation errors still varied after the five training epochs, even though we had more training samples than Hannink et al. [16]. Increasing the number of epochs or batches did not change the varying validation errors. This indicates that the DCNN does not generalize well. Thus, the results might be further improved by incorporating more data samples in the training process of the network.

The embedded implementation of the presented method is a challenge as the DCNN model comprises 85,425 parameters. However, it is still in a range where it can be implemented on a microcontroller. For this method, the acceleration and the gyroscope have to be sampled. This further increases the energy demand compared to the *Acceleration* approach. Taking computational effort and performance into account, the *Acceleration* method would be a better trade-off for an embedded implementation.

4.3. Field Study

The aim of the field study dataset was the evaluation of the estimation of the overall distance of a run in an outside and real-world scenario. The *Trajectory* algorithm also worked best for this dataset. With an MAE of 94.0 m, it is comparable to the results of GPS, which also produced an MAE of 82.1 m, and is used in state-of-the-art running platforms tracking athlete performances. Besides, the IMU technology has the advantage that it allows velocity and distance computations indoors or in scenarios where no satellite connection for GPS is available. Based on the presented results, we argue that although the *Trajectory* algorithm has high standard deviations in the lab study for the stride length calculation, these have no major impact on the computation for longer distances based on stride length. We believe this is due to errors canceling out over time. As the subjects' average velocity was 3.48 m/s during the data acquisition, the high velocity range of 5–6 m/s was not reached for the amateur runners that participated in this study. We expect the results to be worse for the high velocity range, which can be reached by professional athletes.

The *Stride time* algorithm showed the worst performance for the field study dataset (MAE of 599.7 m). Despite its best energy efficiency, our results indicate that its accuracy is too low to use for tracking velocity and distance. The *Deep Learning* approach (MAE 194.5 m) performs better than the *Acceleration* approach (MAE 333.1 m). Due to the fact that the the neural network also needs the 6D-IMU data as an input, it has no benefit compared to the *Trajectory* approach, which performs better. The *Acceleration* approach only requires the sampling of the triaxial accelerometer, which makes it more energy efficient. Despite its decreased accuracy, we propose to use this algorithm in use cases where very strict energy limitations occur.

5. Conclusions and Future Work

In this study, we compared four different algorithms with respect to their performance on stride length and mean average velocity per stride calculation for running. We conducted two studies to evaluate the accuracy of the algorithms: one study in a laboratory environment with a motion capture system as the ground truth, in which we acquired 2377 strides of 27 subjects, and one field study in a real-world scenario. We showed that the *Trajectory* algorithm performs best and especially well for velocities up to 5 m/s. The results of the field study showed that this algorithm does not only work on single strides, but also on longer outdoor runs in a real-world scenario. The MAEs for this scenario showed that the trajectory is comparable to GPS measurements, which is the common method for total distance tracking in amateur running. However, the *Trajectory* algorithm is more costly energy wise due to the fact that both the acceleration and the gyroscope have to be acquired with a sampling rate of 200 Hz. When it comes to an energy-efficient use case, the *Acceleration* algorithm is a good choice, as it only needs to sample the accelerometer, and the sampling rate can be decreased to 60 Hz.

We therefore propose the implementation of the *Trajectory* algorithm for use cases with no energy limitations and the implementation of the *Acceleration* algorithm for use cases with energy restrictions.

In future work, we want to address further parameters that can be computed using inertial measurement units and other sensors located in the sole of a running shoe. Using data acquired by sensors on both feet, it is possible to perform bi-lateral analysis by combining the information of both sensors. Thus, the contribution of the individual lower limbs to the running movement can be further evaluated. Using only data from IMUs within the sole of a running shoe and the *Trajectory* algorithm, analysis regarding imbalances in stride length, stride time or orientation of the two feet

can be conducted. Furthermore, other temporal parameters like flight time or stance time could be computed by adding a toe-off detection. Due to inaccuracies with the toe-off detection in running using only one IMU per foot [43], we plan to also incorporate pressure sensors for toe-off detection into the soles of a running shoe.

Author Contributions: M.Z. conceived of and designed the experiments; M.Z. and M.U. performed the experiments; M.Z., S.G., and U.J. analyzed the data; B.M.E. contributed reagents/materials/analysis tools; M.Z. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cornelissen, V.A.; Fagard, R.H. Effects of endurance training on blood pressure, blood pressure—Regulating mechanisms, and cardiovascular risk factors. *Hypertension* **2005**, *46*, 667–675. [CrossRef] [PubMed]
2. Lee, D.C.; Pate, R.R.; Lavie, C.J.; Sui, X.; Church, T.S.; Blair, S.N. Leisure-time running reduces all-cause and cardiovascular mortality risk. *J. Am. Coll. Cardiol.* **2014**, *64*, 472–481. [CrossRef] [PubMed]
3. Gallo, R.A.; Plakke, M.; Silvis, M.L. Common leg injuries of long-distance runners: Anatomical and biomechanical approach. *Sports Health* **2012**, *4*, 485–495. [CrossRef] [PubMed]
4. Mitchell, J.S.; Mount, D.M.; Papadimitriou, C.H. The discrete geodesic problem. *SIAM J. Comput.* **1987**, *16*, 647–668. [CrossRef]
5. Cavanagh, P.R. *Biomechanics of Distance Running*; Human Kinetics: Champaign, IL, USA, 1990.
6. Gradl, S.; Zrenner, M.; Schuldhaus, D.; Wirth, M.; Cegielny, T.; Zwick, C. Movement Speed Estimation Based on Foot Acceleration Patterns. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, HI, USA, 17–21 July 2018; pp. 3506–3508.
7. Rampp, A.; Barth, J.; Schuelein, S.; Gassmann, K.G.; Klucken, J.; Eskofier, B.M. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1089–1097. [CrossRef] [PubMed]
8. Mariani, B.; Hoskovec, C.; Rochat, S.; Büla, C.; Penders, J.; Aminian, K. 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *J. Biomech.* **2010**, *43*, 2999–3006. [CrossRef] [PubMed]
9. Ferrari, A.; Ginis, P.; Hardegger, M.; Casamassima, F.; Rocchi, L.; Chiari, L. A mobile Kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 764–773. [CrossRef] [PubMed]
10. Bird, J.; Arden, D. Indoor navigation with foot-mounted strapdown inertial navigation and magnetic sensors [emerging opportunities for localization and tracking]. *IEEE Wirel. Commun.* **2011**, *18*, 28–35. [CrossRef]
11. Stroembaeck, P.; Rantakokko, J.; Wirkander, S.L.; Alexandersson, M.; Fors, K.; Skog, I.; Händel, P. Foot-mounted inertial navigation and cooperative sensor fusion for indoor positioning. In Proceedings of the ION International Technical Meeting (ITM), San Diego, CA, USA, 25–27 January 2010; pp. 89–98.
12. Bailey, G.; Harle, R. Assessment of foot kinematics during steady state running using a foot-mounted IMU. *Procedia Eng.* **2014**, *72*, 32–37. [CrossRef]
13. Foxlin, E. Pedestrian tracking with shoe-mounted inertial sensors. *IEEE Comput. Gr. Appl.* **2005**, *25*, 38–46. [CrossRef]
14. Kautz, T.; Groh, B.H.; Hannink, J.; Jensen, U.; Strubberg, H.; Eskofier, B.M. Activity recognition in beach volleyball using a Deep Convolutional Neural Network. *Data Min. Knowl. Discov.* **2017**, *31*, 1678–1705. [CrossRef]
15. Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]
16. Hannink, J.; Kautz, T.; Pasluosta, C.; Barth, J.; Schulein, S.; Gassmann, K.G.; Klucken, J.; Eskofier, B. Mobile Stride Length Estimation with Deep Convolutional Neural Networks. *IEEE J. Biomed. Health Inf.* **2017**, *21*, 85–93. [CrossRef] [PubMed]

17. Altman, A.R.; Davis, I.S. A kinematic method for footstrike pattern detection in barefoot and shod runners. *Gait Posture* **2012**, *35*, 298–300. [CrossRef] [PubMed]

18. Blank, P.; Kugler, P.; Schlarb, H.; Eskofier, B.M. A Wearable Sensor System for Sports and Fitness Applications. In Proceedings of the 19th Annual Congress of the European College of Sport Science, Amsterdam, The Netherlands, 2–5 July 2014; p. 703

19. Ferraris, F.; Grimaldi, U.; Parvis, M. Procedure for effortless in-field calibration of three-axial rate gyro and accelerometers. *Sens. Mater.* **1995**, *7*, 311–330.

20. Michel, K.J.; Kleindienst, F.I.; Krabbe, B. Development of a lower extremity model for sport shoe research. In Proceedings of the 13rd Annual Meeting of ESMAC, Warsaw, Poland, 23–25 March 2004; p. 80.

21. Maiwald, C.; Sterzing, T.; Mayer, T.; Milani, T. Detecting foot-to-ground contact from kinematic data in running. *Footwear Sci.* **2009**, *1*, 111–118. [CrossRef]

22. Kugler, P.; Schlarb, H.; Blinn, J.; Picard, A.; Eskofier, B. A wireless trigger for synchronization of wearable sensors to external systems during recording of human gait. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 4537–4540.

23. Strava Stories—2017 in Stats. 2018. Available online: https://blog.strava.com/2017-in-stats/ (accessed on 22 August 2018).

24. GPX—The GPS Exchange Format. Available online: http://www.topografix.com/gpx.asp (accessed on 28 August 2018).

25. Karney, C.F. Algorithms for geodesics. *J. Geod.* **2013**, *87*, 43–55. [CrossRef]

26. Strohrmann, C.; Harms, H.; Tröster, G.; Hensler, S.; Müller, R. Out of the lab and into the woods: kinematic analysis in running using wearable sensors. In Proceedings of the 13rd International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 119–122.

27. Danion, F.; Varraine, E.; Bonnard, M.; Pailhous, J. Stride variability in human gait: The effect of stride frequency and stride length. *Gait Posture* **2003**, *18*, 69–77. [CrossRef]

28. Perry, J.; Davids, J.R. Gait analysis: Normal and pathological function. *J. Pediat. Orthop.* **1992**, *12*, 815. [CrossRef]

29. Elliott, B.; Blanksby, B. Optimal stride length considerations for male and female recreational runners. *Br. J. Sports Med.* **1979**, *13*, 15. [CrossRef] [PubMed]

30. Hunter, J.P.; Marshall, R.N.; McNair, P.J. Interaction of step length and step rate during sprint running. *Med. Sci. Sports Exerc.* **2004**, *36*, 261–271. [CrossRef] [PubMed]

31. Bailey, G.; Harle, R. Sampling Rates and Sensor Requirements for Kinematic Assessment During Running Using Foot Mounted IMUs. In *International Congress on Sports Science Research and Technology Support*; Springer: Berlin, Germany, 2014; pp. 42–56.

32. Skog, I.; Handel, P.; Nilsson, J.O.; Rantakokko, J. Zero-velocity detection—An algorithm evaluation. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2657–2666. [CrossRef] [PubMed]

33. De Wit, B.; De Clercq, D.; Aerts, P. Biomechanical analysis of the stance phase during barefoot and shod running. *J. Biomech.* **2000**, *33*, 269–278. [CrossRef]

34. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

35. Keras. 2015. Available online: https://keras.io (accessed on 22 August 2018).

36. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Available online: tensorflow.org (accessed on 29 November 2018)

37. Srivastava, N.; Hinton, G.E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

38. Bland, J.M.; Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *327*, 307–310. [CrossRef]

39. Lindsay, T.R.; Noakes, T.D.; McGregor, S.J. Effect of treadmill versus overground running on the structure of variability of stride timing. *Percept. Mot. Skills* **2014**, *118*, 331–346. [CrossRef] [PubMed]

40. Abdul Razak, A.H.; Zayegh, A.; Begg, R.K.; Wahab, Y. Foot plantar pressure measurement system: A review. *Sensors* **2012**, *12*, 9884–9912. [CrossRef] [PubMed]

41. MPU-9250 Datasheet. Available online: https://www.invensense.com/download-pdf/mpu-9250-datasheet/ (accessed on 22 August 2018).

42. Zrenner, M.; Ullrich, M.; Zobel, P.; Jensen, U.; Laser, F.; Groh, B.; Dümler, B.; Eskofier, B. Kinematic parameter evaluation for the purpose of a wearable running shoe recommendation. In Proceedings of the 15th IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN), Las Vegas, NV, USA, 4–7 March 2018; pp. 106–109.
43. Mo, S.; Chow, D.H. Accuracy of three methods in gait event detection during overground running. *Gait Posture* **2018**, *59*, 93–98. [CrossRef] [PubMed]

sensors

Article

Validation of an Inertial Sensor Algorithm to Quantify Head and Trunk Movement in Healthy Young Adults and Individuals with Mild Traumatic Brain Injury

Lucy Parrington [1,2,*], **Deborah A. Jehu** [1,2], **Peter C. Fino** [1,2,3], **Sean Pearson** [4], **Mahmoud El-Gohary** [4] and **Laurie A. King** [1,2,5]

1   Department of Neurology, Oregon Health & Science University, 3181 S.W. Sam Jackson Park Rd., Portland, OR 97239, USA; deb.jehu@ubc.ca (D.A.J.); peter.fino@utah.edu (P.C.F.); kingla@ohsu.edu (L.A.K.)
2   VA Portland Health Care System, 3710 SW US Veterans Hospital Road, Portland, OR 97239, USA
3   Department of Health, Kinesiology and Recreation, University of Utah, 250 S 1850 E, Salt Lake City, UT 84112, USA
4   APDM Wearable Technologies, Portland, OR 97201, USA; seanp@apdm.com (S.P.); mahmoud@apdm.com (M.E.-G.)
5   National Center for Rehabilitative Auditory Research (NCRAR), VA Portland Health Care System, 3710 SW US Veterans Hospital Road/P5, Portland, OR 97239, USA
*   Correspondence: parringt@ohsu.edu

Abstract: Wearable inertial measurement units (IMUs) may provide useful, objective information to clinicians interested in quantifying head movements as patients' progress through vestibular rehabilitation. The purpose of this study was to validate an IMU-based algorithm against criterion data (motion capture) to estimate average head and trunk range of motion (ROM) and average peak velocity. Ten participants completed two trials of standing and walking tasks while moving the head with and without moving the trunk. Validity was assessed using a combination of Intra-class Correlation Coefficients (ICC), root mean square error (RMSE), and percent error. Bland-Altman plots were used to assess bias. Excellent agreement was found between the IMU and criterion data for head ROM and peak rotational velocity (average ICC > 0.9). The trunk showed good agreement for most conditions (average ICC > 0.8). Average RMSE for both ROM (head = 2.64°; trunk = 2.48°) and peak rotational velocity (head = 11.76 °/s; trunk = 7.37 °/s) was low. The average percent error was below 5% for head and trunk ROM and peak rotational velocity. No clear pattern of bias was found for any measure across conditions. Findings suggest IMUs may provide a promising solution for estimating head and trunk movement, and a practical solution for tracking progression throughout rehabilitation or home exercise monitoring.

Keywords: concussion; inertial motion units (IMUs); vestibular exercises; validation; motion capture

1. Introduction

Individuals with mild traumatic brain injury (mTBI) suffering from impaired vestibular and ocular-motor impairments may be prescribed vestibular rehabilitation consisting of head and trunk movements during physical therapy. Vestibular rehabilitation typically includes gradual increases in the range of motion (ROM) and velocity of head movements and has shown promising improvements in the reduction of symptoms and greater overall function in mTBI patients [1–3]. While a physical therapist trains the patient in these exercises, a home exercise program is a vital part of vestibular rehabilitation therapy. Head and trunk movements can be compromised in this population, as the

perceived head position relative to the trunk is generally impaired in individuals with vestibular pathologies [4]. Consequently, avoidance behavior and maladaptive strategies, such as limiting the head ROM and rotational velocity, may be used in an effort to minimize symptoms [4]. These subtle impaired movements, such as head and trunk velocity, are often not detected visually [5], and have the potential to interfere with successful rehabilitation. Thus, the ability to quantify and track these movements both within the clinic and during a home exercise program may be highly beneficial.

Optical motion capture is commonly considered the gold-standard for tracking human movement. However, these motion capture systems are costly, can be limited by optical occlusion, and require a dedicated motion laboratory with specialists to collect, process and interpret the data. These factors make motion capture a suboptimal choice for clinical assessment in practice. Conversely, inertial measurement units (IMUs) provide a promising alternative due to the lower associated cost, simplified experimental set-up, and the ability to acquire data during everyday life, such as home monitoring [6].

IMUs have been shown to be a viable tool for measuring human motion across a wide range of activities and environments (for review across multiple domains see References [7–11]). While Duc and colleagues [12] have reported very good agreement with motion capture, and good retest reliability when measuring cervical ROM [12], it is unclear if IMUs can accurately estimate ROM and angular velocity during more complex movements. Additionally, the accuracy of IMU output can be subject to sensor placement [13], the developed algorithm, and how well it compensates for factors such as magnetic distortion and gyroscopic drift [14,15]. Further, it is important to validate an IMU algorithm across the range of measures of interest.

Our lab is currently evaluating the use of IMUs within a standard vestibular focused home-exercise program involving head and trunk movements (ROM and peak rotational velocity) during standing and walking in individuals with mTBI. Prior to conducting the larger intervention, we wanted to determine the validity of the sensors that will be used. Therefore, the aim of this study was to determine the validity of IMUs to estimate head and trunk ROM and peak rotational velocity during head movements made when standing and walking, in comparison with three-dimensional motion capture data.

2. Methods

2.1. Participants
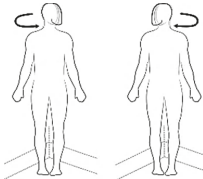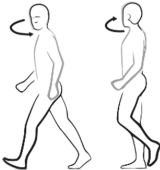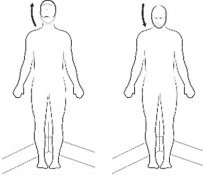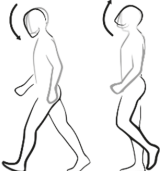
Five healthy young adults (3 females, mean (SD), age = 24.0 (2.3) years, height = 1.77 (0.12) m, mass = 77.4 (16.4) kg) and 5 individuals with mTBI (3 females, age = 36.4 (11.8) years, height = 1.74 (0.07) m, mass = 72.7 (11.2) kg, median (range), time since injury = 1.09 (0.43–4.38) years) participated in this study. Participants were included in the study if they were: (1) between 18 and 60 years old and (2) had zero to minimal cognitive impairment. In addition, the mTBI participants were included if they had a diagnosis of mTBI based upon Department of Defense criteria [16] and were still self-reporting symptoms related to their mTBI. Exclusion criteria included: (1) any musculoskeletal, neurological, or sensory problems that could explain balance deficits (not including the mTBI); (2) moderate to severe substance-use disorder within the past month (Diagnostic and Statistical Manual-5); (3) being in pain during the evaluation ($\geq$7/10 by patient subjective report); (4) pregnancy; and (5) inability to abstain from medications that could impair balance for 24 h prior to testing.

The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Oregon Health & Science University Institutional Review Board (IRB #17206). All participants provided written informed consent prior to commencing testing.

2.2. Experimental Protocol

The testing protocol included four standing and four walking conditions (Table 1). Each of the conditions required either left and right (L/R) or up and down (U/D) continuous head movement. Two of the standing conditions were visual motion sensitivity (VMS) tasks and involved movement of the head and trunk en bloc. The walkway for the walking trials was 3.9 m long to stay within the optical field of the camera-based motion system. Participants completed two 30-second trials of each condition. All testing was completed at the Balance Disorders Laboratory at Oregon Health & Science University.

Table 1. Description of standing and walking conditions.

Standing	Standing VMS	Walk	Tandem Walk
Standing with feet together while turning the head L/R.	Standing with feet together while moving head and trunk together. Participants held their dominant arm straight out in front of the body, fixed their eyes on their thumb, and concurrently moved the head, trunk, and arms en bloc L/R.	Participants walked continuously at a comfortable pace while turning the head L/R.	Participants walked in tandem at a comfortable pace while turning the head L/R.
Standing with feet together while moving the head U/D.	Standing with feet together while moving head and trunk together. Participants held their dominant arm straight out in front of the body, fixed their eyes on their thumb, and concurrently moved the head, trunk, and arms en bloc U/D.	Participants walked continuously at a comfortable pace while moving the head U/D.	Participants walked in tandem at a comfortable pace while moving the head U/D.

Note: VMS = visual motion sensitivity; L/R = Left and right; U/D = Up and down.

2.3. Equipment and Data Analysis

To collect inertial sensor data, two wearable IMUs; (APDM, Inc., Portland, OR, USA) were attached to the sternum and forehead of the participant. Each IMU included a tri-axial accelerometer (±6 g), gyroscope ($\pm2000\,^\circ$/s) and magnetometer (±6 gauss) that measured at a sampling frequency of 128 Hz. Moveo application (APDM, Inc.) was used to record the IMU data. The IMUs use wireless synchronization to ensure multiple units collect data with a precision of better than ±1 ms. Participants were also fitted with six reflective markers to collect simultaneous motion capture data. Markers were fixed to the forehead, the bilateral mandibular condyle of the head, the sternum, and the bilateral acromion process of the trunk. Motion capture data were collected using a 12-camera Motion Analysis system (120 Hz, Raptor-E, Motion Analysis Co., Santa Rosa, CA, USA) and processed using Cortex

v6.2.3 (Motion Analysis, Co.). Motion capture was synchronized with the IMU recording using an APDM synchronization box.

For each participant, a static trial was captured to define head and trunk segment position and orientation. This process allowed for each segment coordinate system to be rotated about the mediolateral axis, such that the anterior-posterior axis lies in the horizontal plane for the head and trunk in the static pose. A state space model and Kalman filter were used for sensor fusion between accelerometer, gyroscope, and magnetometer sensor data of the IMU [17]. Angular velocities were extracted from the head and trunk IMUs corresponding to rotations in the transverse plane (for L/R) and sagittal plane (for U/D). Angular displacement was calculated by integrating the angular velocity of the head in the intended direction (L/R or U/D).

Optical data were filtered using a dual-pass second order Butterworth filter (6 Hz cut-off) and up-sampled to match the sampling rate of the IMU data. The head and trunk segments (defined in Table 2) used a right-hand coordinate system. Flexion, abduction, and axial rotation were decomposed using Euler angles. Segment angles for the rotations of interest were calculated and differentiated to estimate rotational velocities.

Table 2. Head and trunk segment definitions.

Segment	Origin/Axes	Definition
Head	Origin	Midpoint between right and left mandibular condyle markers
	Mediolateral axis	Projected from the origin to the right mandibular condyle
	Anterior posterior axis	Projected from the origin to the forehead marker
	Vertical axis	Projected from the origin, orthogonal to the AP and ML axes
Trunk	Origin	Midpoint between right and left acromion process
	Mediolateral axis	Projected from the origin to the right acromion process
	Anterior posterior axis	Projected from the origin to the sternum marker
	Vertical axis	Projected from the origin, orthogonal to the AP and ML axes

For both IMU and optical datasets, time series data were segmented into individual head turns, allowing the calculation of ROM and peak rotational velocity. For the walking and tandem walking trials, only the straight walking segments were included. Portions of the trial when participants were turning at the ends of the walking path were removed. Turns were detected using a threshold turn angle greater than 45° and a peak turn velocity greater than 15 degrees per second [18].

2.4. Statistical Analysis

Validity of the lMU-based algorithm was assessed using a combination of Intra-class Correlation Coefficients (ICC(A,1)) for assessing absolute agreement, root mean square error (RMSE), as well as percent error between the IMU and criterion value of the motion capture data. Interpretation of ICC values were based on Koo and Li [19], with the following cut-offs: poor <0.5; moderate = 0.5–0.75; good = 0.75–0.9; and excellent >0.9. Bland-Altman plots [20] were also examined to gain an understanding of any patterns relating to bias.

3. Results

Representative time series data for walking with head turns L/R and walking with head turns U/D are displayed for one healthy participant (Figure 1) and one participant with mTBI (Figure 2).
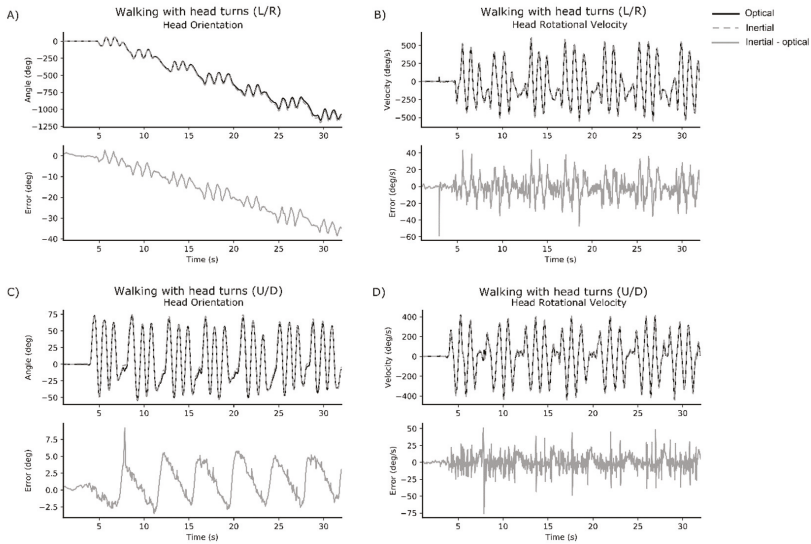
Figure 1. Representative time series data from one healthy participant. Each subplot provides the optical signal (black solid line) and IMU signal (grey dashed line) overlaid in the upper figure, and the error between optical and IMU signal below (grey solid line). Subplots represent: (**A**) head orientation for walking with head turns (L/R); (**B**) peak rotational velocity of the head for walking with head turns (L/R); (**C**) head orientation for walking with head turns (U/D); and (**D**) peak rotational velocity of the head for walking with head turns (U/D).
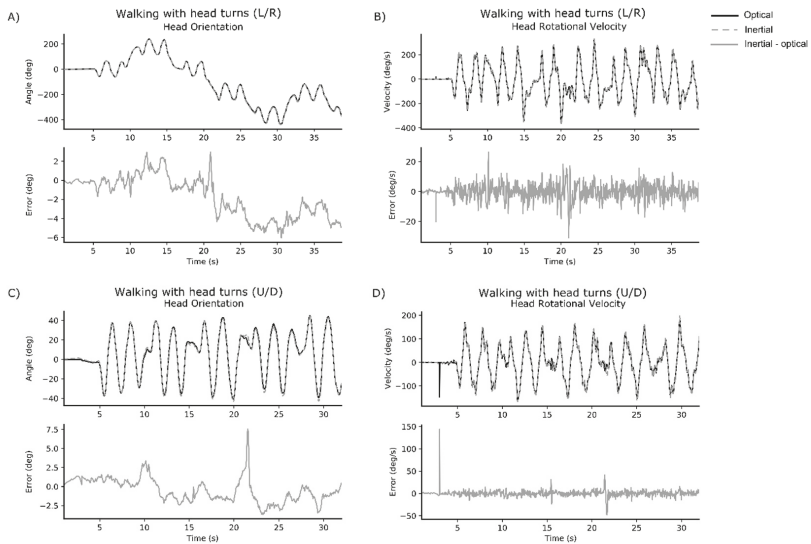
Figure 2. Representative time series data from one participant with mTBI. Each subplot provides the optical signal (black solid line) and IMU signal (grey dashed line) overlaid in the upper figure, and the error between optical and IMU signal below (grey solid line). Subplots represent: (**A**) head orientation for walking with head turns (L/R); (**B**) peak rotational velocity of the head for walking with head turns (L/R); (**C**) head orientation for walking with head turns (U/D); and (**D**) peak rotational velocity of the head for walking with head turns (U/D).

Correlation, RMSE, and percent error results for the head and trunk are presented in Tables 3 and 4, respectively. The IMU data strongly represented the criterion motion capture data for head ROM and peak rotational velocity across all conditions (ICC(A,1) > 0.9). RMSE across conditions remained low for head ROM but increased in the walking L/R and tandem walking L/R conditions for peak rotational head velocity. Despite the higher RMSE, the percent error for the head remained low across all conditions (<5%).

Table 3. Validity results comparing IMU to motion capture for head range of motion (ROM) and peak rotational velocity (ω_p).

	ICC(A,1)		RMSE		% Error	
Condition	ROM	ω_p	ROM	ω_p	ROM	ω_p
Standing L/R	0.993	0.994	2.39	11.58	−1.5	−1.6
Standing U/D	0.997	0.991	1.36	9.59	−0.4	−2.3
Standing VOR L/R	0.992	0.994	3.78	6.73	−1.8	−0.3
Standing VOR U/D	0.998	0.992	1.29	5.55	0.02	−2.0
Walking L/R	0.991	0.986	3.55	20.26	−2.0	−4.0
Walking U/D	0.985	0.991	2.71	12.52	−1.2	−3.2
Tandem walking L/R	0.994	0.987	2.93	17.04	−1.9	−4.4
Tandem walking U/D	0.985	0.988	3.11	10.84	−0.4	−2.3
Mean across conditions	0.992	0.990	2.64	11.76	−1.1	−2.5
SD across conditions	0.005	0.003	0.86	4.61	0.8	1.3

Intra-class Correlations Coefficients between the IMU and motion capture data for the trunk were stronger in the L/R direction (ROM ICC(A,1) > 0.9; peak rotational velocity ICC(A,1) > 0.9), than in the U/D direction (ROM ICC(A,1) = 0.580 to 0.907; peak rotational velocity ICC(A,1) = 0.436 to 0.787) across conditions. The reduced strength of the relationship is also mirrored in the RMSE and percent error scores for the U/D motions.

Table 4. Validity results comparing IMU to motion capture for trunk ROM and peak rotational velocity.

	ICC(A,1)		RMSE		% Error	
Condition	ROM	ω_p	ROM	ω_p	ROM	ω_p
Standing L/R	0.986	0.960	0.59	2.31	−6.0	−6.9
Standing U/D	0.580	0.815	1.82	5.41	−29.3	−11.7
Standing VOR L/R	0.985	0.997	3.55	4.16	1.9	0.6
Standing VOR U/D	0.907	0.787	6.49	12.71	19.2	5.3
Walking L/R	0.997	0.976	1.17	5.47	0.8	−6.3
Walking U/D	0.843	0.746	2.60	12.08	−13.2	−17.1
Tandem walking L/R	0.998	0.976	0.96	3.69	−1.5	−3.5
Tandem walking U/D	0.639	0.436	2.66	13.14	−2.9	0.2
Mean across conditions	0.867	0.837	2.48	7.37	−3.9	−4.9
SD across conditions	0.169	0.189	1.78	4.20	13.8	7.2

Bland-Altman plots did not indicate any clear patterns of bias. Examples are provided below for the head ROM and peak rotational velocity during walking with head turns (Figure 3).
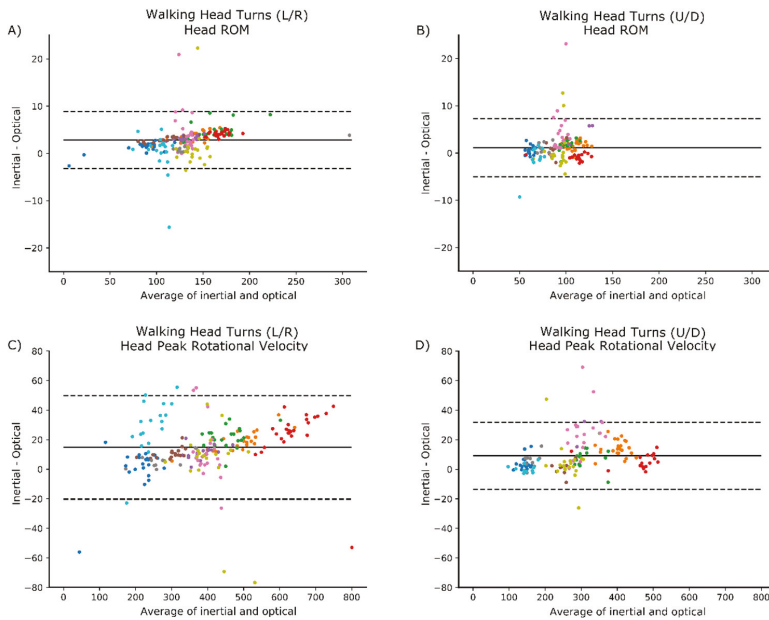
Figure 3. Bland-Altman plots for walking conditions. Each participant is represented by a different color within the plot. (**A**) ROM, walking with head turns (L/R); (**B**) ROM, walking with head turns (U/D); (**C**) peak rotational velocity, walking with head turns (L/R); and (**D**) peak rotational velocity, walking with head turns (U/D). Solid lines represent the mean difference, and dashed lines represent $\pm 1.96 \times$ SD.

4. Discussion

In this study we investigated the validity of IMUs to detect and measure head and trunk ROM and peak rotational velocity during a set of commonly prescribed vestibular rehabilitation tasks. Our findings suggest excellent validity for the IMU system when capturing head movements in both the L/R and U/D conditions and excellent validity capturing trunk movements in the L/R conditions. Inertial sensors showed moderate to excellent ability to estimate trunk ROM in the U/D conditions and was moderate to good at capturing peak rotational velocity in U/D conditions; except during tandem walking which showed poor agreement. The excellent agreement found here for head motion is consistent with previous findings [12], who showed excellent agreement for cervical angles collected using inertial sensors on the head and neck. This work also extends prior research, by identifying that IMUs can accurately estimate the ROM and peak turning velocity during both standing and locomotor tasks.

Automatically characterizing head and trunk movements during routinely prescribed vestibular exercises using IMUs is an innovative approach that will allow a more sensitive and objective analysis of progression during vestibular rehabilitation. In people with mTBI, smaller and slower head movements during performance tasks have been reported [5] but such movements are not easily quantified with the naked eye and may not be perceived by the patient performing the exercise. Quantifying such information with IMUs could inform both the treating physical therapist and, with time, the patient themselves by providing immediate feedback on velocity and quality of performance.

Despite the good agreement between IMU and motion capture systems, we believe some of the estimation errors might be attributed to a misalignment of the IMU frame relative to the anatomical axes of rotation. When the IMUs are attached to different body segments, they are not perfectly aligned with the segments' main axes of rotation. To estimate this misalignment, we asked the study

participants to remain stationary in a neutral pose for about three seconds at the beginning of the recording. This information was then used to realign the sensors' data for analysis, using matrix rotation, before calculating the joint metrics. While this addressed the misalignment of the sensors relative to the anatomical axes, it assumed the participant could both remain stationary and adopt a truly neutral initial pose. This was not always true for every participant, and we hypothesize that this contributed to the larger errors observed in a few of the subjects.

Although 3D motion capture is commonly classified as a gold-standard measurement, it is possible that reduced agreement in some cases could partially be a function of the motion capture methods implemented in this study. Firstly, the accuracy of motion capture systems can decrease as the capture volume increases [21]. Despite using a 12-camera system it is possible that the size of our capture area, elongated to collect the tandem and walking trials, played a role in the reduced agreement between the two systems. Another source of disagreement between the inertial and optical systems could be attributed to skin motion artifact and muscle movements—a known issue [22] with systems that use markers attached to the body. Similarly, skin artifacts can also influence the inertial sensor measurements resulting in potential orientation changes. These orientation changes may produce joint metric estimates that are biomechanically unlikely and lead to a disagreement between the systems.

4.1. Clinical Implications

Inertial measurement technologies enable clinicians to capture patients' movement during unconstrained activities—whether in the clinic or during daily living within their home environment. As these tools become more accessible to therapists and patients, the information they provide has the potential to help improve diagnosis and recommendations for therapeutic interventions or rehabilitation strategies. Additionally, they can facilitate the collection of clinical trial outcome measures in most outpatient clinics, rather than specialized laboratory settings. Nonetheless, we note that IMU-based systems are not a replacement for clinical experience, but rather a tool that can complement clinical judgment. Furthermore, future studies are needed to determine the feasibly of using these systems during rehabilitation programs to monitor compliance and the progression of exercise intensity.

4.2. Limitations

There are some limitations to our study that should be noted. First, validation data are limited to the exercises of interest to our group. These vestibular exercises were chosen based on their common use and prescription to persons with ongoing balance problems after mTBI. Nonetheless, validation data in this study may not be generalizable to more dynamic tasks involving multiple planes of motion. Any estimation of trunk and head ROM and peak rotational velocities in more complex movements may, therefore, require further validation. Second, as accuracy of IMU output can be subject to sensor placement [13], it is possible that a different placement or a combined placement of sensors (i.e., an IMU on both the sternum and the lumbar spine), could provide a more accurate estimation of trunk motion. Third, our capture volume for optical motion data was limited, resulting in occasional marker occlusions when participants turned around the ends of our capture volume. Gap filling of missing optical markers and the resulting underestimation of head angles may be responsible for the fluctuations in error coincident with the turns. Fourth, errors in the trunk ROM and peak velocity may have been minimized had the trunk been defined differently. For U/D head motions, in particular, it is plausible that movement artifact occurred with neck movement pulling on the skin at the top of the sternum where one marker was placed. Finally, the IMU algorithm required the participants to stand in a static position so that the orientation of the sensors could be defined. This assumes that subjects can maintain a neutral stationary position, and further investigation into the effect of this potential issue is warranted.

5. Conclusions

Our findings suggest that it is feasible to use IMUs to measure head and trunk movement and provide metrics that are clinically important. The system is portable, unobtrusive, and easy to use. These features make such systems well suited for use in the clinic to detect and characterize head and trunk movement during routine and standard vestibular rehabilitation. This study provides an initial step towards the implementation of IMUs to provide clinically meaningful information to physical therapists treating patients with imbalance after mTBI.

References

1. Alsalaheen, B.A.; Mucha, A.; Morris, L.O.; Whitney, S.L.; Furman, J.M.; Camiolo-Reddy, C.E.; Collins, M.W.; Lovell, M.R.; Sparto, P.J. Vestibular rehabilitation for dizziness and balance disorders after concussion. *J. Neurol. Phys. Ther.* **2010**, *34*, 87–93. [CrossRef] [PubMed]

2. Alsalaheen, B.A.; Whitney, S.L.; Mucha, A.; Morris, L.O.; Furman, J.M.; Sparto, P.J. Exercise prescription patterns in patients treated with vestibular rehabilitation after concussion. *Physiother. Res. Int.* **2013**, *18*, 100–108. [CrossRef] [PubMed]

3. Schneider, K.J.; Meeuwisse, W.H.; Nettel-Aguirre, A.; Barlow, K.; Boyd, L.; Kang, J.; Emery, C.A. Cervicovestibular rehabilitation in sport-related concussion: A randomised controlled trial. *Br. J. Sports Med.* **2014**, *48*, 1294–1298. [CrossRef] [PubMed]

4. Borel, L.; Harlay, F.; Magnan, J.; Chays, A.; Lacour, M. Deficits and recovery of head and trunk orientation and stabilization after unilateral vestibular loss. *Brain* **2002**, *125*, 880–894. [CrossRef] [PubMed]

5. Fino, P.C.; Wilhelm, J.; Parrington, L.; Stuart, S.; Chesnutt, J.C.; King, L.A. Inertial sensors reveal subtle motor deficits when walking with horizontal head turns after concussion. *J. Head Trauma Rehabil.* **2018**. [CrossRef] [PubMed]

6. El-Gohary, M.; Pearson, S.; McNames, J.; Mancini, M.; Horak, F.; Mellone, S.; Chiari, L. Continuous monitoring of turning in patients with movement disability. *Sensors* **2014**, *14*, 356–369. [CrossRef] [PubMed]

7. O'Reilly, M.; Caulfield, B.; Ward, T.; Johnston, W.; Doherty, C. Wearable inertial sensor systems for lower limb exercise detection and evaluation: A systematic review. *Sports Med.* **2018**, *48*, 1–26. [CrossRef] [PubMed]

8. Cuesta-Vargas, A.I.; Galán-Mercant, A.; Williams, J.M. The use of inertial sensors system for human motion analysis. *Phys. Ther. Rev.* **2010**, *15*, 462–473. [CrossRef] [PubMed]

9. Magalhaes, F.A.d.; Vannozzi, G.; Gatta, G.; Fantozzi, S. Wearable inertial sensors in swimming motion analysis: A systematic review. *J. Sports Sci.* **2015**, *33*, 732–745. [CrossRef] [PubMed]

10. Sprager, S.; Juric, M. Inertial sensor-based gait recognition: A review. *Sensors* **2015**, *15*, 22089–22127. [CrossRef] [PubMed]

11. Chambers, R.; Gabbett, T.J.; Cole, M.H.; Beard, A. The use of wearable microsensors to quantify sport-specific movements. *Sports Med.* **2015**, *45*, 1065–1081. [CrossRef] [PubMed]

12. Duc, C.; Salvia, P.; Lubansu, A.; Feipel, V.; Aminian, K. A wearable inertial system to assess the cervical spine mobility: Comparison with an optoelectronic-based motion capture evaluation. *Med. Eng. Phys.* **2014**, *36*, 49–56. [CrossRef] [PubMed]

13. Theobald, P.S.; Jones, M.D.; Williams, J.M. Do inertial sensors represent a viable method to reliably measure cervical spine range of motion? *Man. Ther.* **2012**, *17*, 92–96. [CrossRef] [PubMed]

14. De Vries, W.; Veeger, H.; Baten, C.; Van Der Helm, F. Magnetic distortion in motion labs, implications for validating inertial magnetic sensors. *Gait Posture* **2009**, *29*, 535–541. [CrossRef] [PubMed]

15. Damgrave, R.G.J.; Lutters, D. The drift of the Xsens Moven motion capturing suit during common movements in a working environment. In Proceedings of the 19th CIRP Design Conference–Competitive Design, Cranfield, UK, 30–31 March 2009.

16. The Management of Concussion-mild Traumatic Brain Injury Working Group. *VA/DoD Clinical Practice Guidelines for the Management of Concussion/mild Traumatic Brain Injury*; Department of Veteran's Affairs: Washington, DC, USA, 2016; pp. 1–133.

17. Kok, M.; Hol, J.D.; Schön, T.B. Using inertial sensors for position and orientation estimation. *Found. Trends Signal Process.* **2017**, *11*, 1–89. [CrossRef]

18. Pearson, S.; Mancini, M.; El-Gohary, M.; McNames, J.; Horak, F. Turn detection and characterization with inertial sensors. In Proceedings of the 1st International Congress on Sports Science Research and Technology Support, Lisbon, Portugal, 15–17 November 2015; pp. 19–22.

19. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]

20. Bland, J.M.; Altman, D.G. Agreement between methods of measurement with multiple observations per individual. *J. Biopharm. Stat.* **2007**, *17*, 571–582. [CrossRef] [PubMed]

21. van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**. [CrossRef] [PubMed]

22. Cutti, A.G.; Paolini, G.; Troncossi, M.; Cappello, A.; Davalli, A. Soft tissue artefact assessment in humeral axial rotation. *Gait Posture* **2005**, *21*, 341–349. [CrossRef] [PubMed]

sensors

Article

Gradient-Based Multi-Objective Feature Selection for Gait Mode Recognition of Transfemoral Amputees

Gholamreza Khademi *,†, Hanieh Mohammadi † and Dan Simon

Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH 44115, USA; h.mohammadi17@csuohio.edu (H.M.); d.j.simon@csuohio.edu (D.S.)

* Correspondence: g.khademi17@csuohio.edu; Tel.: +1-216-801-9346

† These authors contributed equally to this work.

Abstract: One control challenge in prosthetic legs is seamless transition from one gait mode to another. User intent recognition (UIR) is a high-level controller that tells a low-level controller to switch to the identified activity mode, depending on the user's intent and environment. We propose a new framework to design an optimal UIR system with simultaneous maximum performance and minimum complexity for gait mode recognition. We use multi-objective optimization (MOO) to find an optimal feature subset that creates a trade-off between these two conflicting objectives. The main contribution of this paper is two-fold: (1) a new gradient-based multi-objective feature selection (GMOFS) method for optimal UIR design; and (2) the application of advanced evolutionary MOO methods for UIR. GMOFS is an embedded method that simultaneously performs feature selection and classification by incorporating an elastic net in multilayer perceptron neural network training. Experimental data are collected from six subjects, including three able-bodied subjects and three transfemoral amputees. We implement GMOFS and four variants of multi-objective biogeography-based optimization (MOBBO) for optimal feature subset selection, and we compare their performances using normalized hypervolume and relative coverage. GMOFS demonstrates competitive performance compared to the four MOBBO methods. We achieve a mean classification accuracy of 97.14% ± 1.51% and 98.45% ± 1.22% with the optimal selected subset for able-bodied and amputee subjects, respectively, while using only 23% of the available features. Results thus indicate the potential of advanced optimization methods to simultaneously achieve accurate, reliable, and compact UIR for locomotion mode detection of lower-limb amputees with prostheses.

Keywords: user intent recognition; transfemoral prosthesis; multi-objective optimization; biogeography-based optimization

1. Introduction

Prosthetic legs have significantly enhanced the lifestyle of individuals with a transfemoral amputation. Prostheses help lower-limb amputees regain their walking mobility for activities such as level walking, stair ascent and descent, incline walking, sitting and standing, etc. One active research area is the development of a functional control system for each walking task [1–3]. The main design objective is to enable amputees to achieve walking that is similar to that of able-bodied persons, while minimizing metabolic energy expenditure. Challenges include recognizing gait modes automatically, selecting the appropriate control system corresponding to the identified gait mode, and achieving a smooth transition in real time. Activity mode recognition must be achieved in parallel with control system development to address these problems. Activity mode recognition is referred to as high-level control, while control system design for each walking activity is referred to as low-level control [4]. The focus of this paper is the development of a high-level control system.

In the design of an intent recognition system, several questions arise, including which input signals and machine learning algorithms will provide a UIR system with fast and reliable prediction performance. Previous research has addressed these questions in different ways. For instance, surface electromyography (sEMG) signals were used to train UIR [5,6]. Although sEMG resulted in high classification accuracy, Ref. [7] reported uncertain performance due to sEMG signal variability in real-world conditions. Variation could be because of electrode shift [8], skin temperature change [9], or muscle volume change [10]. Therefore, external sensors on the prosthesis have received significant attention. For instance, classifiers have been trained with data collected from mechanical sensors [11], optical distance sensors [12], and inertial measurement units [7]. In addition, Refs. [13,14] showed that the fusion of sensory measurements could enhance learning, although the amputee subject could be inconvenienced by wearing additional sensors. Various supervised machine learning algorithms have been implemented to build UIR systems, including linear discriminant analysis (LDA) [15], quadratic discriminant analysis (QDA) [16], Gaussian mixture models (GMMs) [11], support vector machines (SVMs) [14], and artificial neural networks (ANNs) [5]. To avoid the need for user-specific classifier training, Ref. [17] proposed a user-independent UIR system in which classifier performance is robust to user-specific characteristics.

Current UIR systems have been designed with one goal in mind: highest possible prediction accuracy. In clinical applications, it is extremely important that UIR can accurately predict activity modes with substantially different characteristics because misclassification can cause a loss of balance [7,18]. However, there remains a gap in the design of UIR with low complexity. UIR has low complexity if it can be implemented with only significant features extracted from minimal sensing hardware. UIR with low complexity is important because such systems: (1) eliminate unneeded body-worn sensors that may be irritating and cumbersome; (2) avoid numerical instability and overfitting during training; (3) are robust to noisy measurement signals and sensor failures; and (4) decrease computational effort, which is important for real-time operation. These reasons have motivated previous research to apply feature selection to design UIR with low complexity [7,19]. They used sequential forward selection to obtain a subset with only the most informative features. In contrast, in this paper, we develop a new framework for UIR that *simultaneously* achieves maximum accuracy and minimum complexity. Complexity and accuracy are two conflicting objectives. To the best of our knowledge, this paper is the first attempt to find a compromise solution for this problem using multi-objective optimization.

The main contributions of this paper are two-fold: (1) a new MOO method called GMOFS for optimal feature subset selection; and (2) the application of four evolutionary MOBBO methods for the UIR problem, including vector evaluated BBO (VEBBO), non-dominated sorting BBO (NSBBO), niched Pareto BBO (NPBBO), and strength Pareto BBO (SPBBO). We have chosen to use BBO in this paper as the evolutionary algorithm (EA) because of its demonstrated effectiveness and recent popularity for optimizing real-world problems [20,21]. MOBBO methods have the potential to find the global optimum [22,23]; however, they are computationally expensive due to the many required fitness function evaluations. To avoid this drawback, we propose GMOFS for feature selection.

Several different types of feature selection methods have been proposed. *Filter* methods are feature selection methods that assess the quality of a subset of features independently or with respect to the output class [24]. *Wrapper* methods are feature selection methods that assess the quality of a subset of features by measuring the prediction accuracy of a classifier that is trained with that subset [25]. *Embedded* methods are feature selection methods that overcome the disadvantages of filter and wrapper methods. Unlike filter methods, embedded methods account for the bias of the classifier, and, unlike wrapper methods, they are computationally efficient [26,27]. Various embedded feature selection algorithms have been proposed, mostly for linear problems with a single objective [27,28]. Embedded methods also incorporate regularization algorithms, such as ridge regression [29], least absolute shrinkage and selection [30], and elastic nets [31].

GMOFS is our newly proposed embedded method that simultaneously performs feature selection and classification, and that accounts for multiple objectives in nonlinear systems such as UIR. GMOFS incorporates an elastic net in multilayer perceptron (MLP) neural network training. The elastic net uses a Lagrange multiplier with a complexity parameter to reduce the feature set to an optimal subset, and the MLP classifier is trained with the optimal subset. We investigate the influence of the complexity parameter on the solution of the constrained MLP optimization problem. We then use the optimization solutions to obtain a GMOFS Pareto front, which is a set of non-dominated solutions that are equally important apart from the designer's subjective preference of objectives.

Section 2 presents a general framework for UIR. In Section 2.1, an informative set of signals reflecting various walking tasks are collected experimentally from three able-bodied and three amputee subjects. In addition, the data are filtered and processed to eliminate noise and missing data points. In Section 2.2, we use both disjoint windowing and overlapped windowing to extract data frames. The length of the data frame and the increment of the moving window are chosen to compromise the informativeness of the data and the computational effort, while taking real-time computational constraints into account. In Section 2.3, various time-domain (TD) and frequency-domain (FD) features are extracted from each data frame for each measurement signal. A training data set is obtained in which all features are normalized to have a zero mean and unity variance. In Section 2.4, we use a pre-selection approach to exclude insignificant features, and then apply MOO for final feature selection. We implement GMOFS and four variants of MOBBO to minimize the size of the selected feature subset and maximize the prediction accuracy. In Section 2.5, the performance of several classifiers, including LDA, QDA, SVMs with both linear and radial basis function (RBF) kernels, MLPs, and decision trees (DTs), are compared, and the best one is selected for UIR. In Section 2.6, majority voting filter (MVF) is implemented to avoid sudden jumps between identified classes and enhance UIR performance. Section 4 discusses the experimental setup and classification results for the optimally designed UIR system. Finally, Section 5 discusses conclusions and future work.

2. Materials and Methods

In this section, we present the methodology used to design the user intent recognition (UIR) system. The architecture of the UIR system is illustrated in Figure 1. Our new contribution is a novel feature selection method based on multi-objective optimization (MOO), as illustrated in Figure 1 in the double-lined box (Section 2.4). In this box, the application of four multi-objective biogeography-based optimization (MOBBO) methods for gait mode classification is new, and a novel MOO-based feature selection method called gradient-based multi-objective feature selection (GMOFS) is new. The remaining parts of the UIR system are implemented based on the existing literature. The role of each subsystem is explained in more detail in the following subsections.
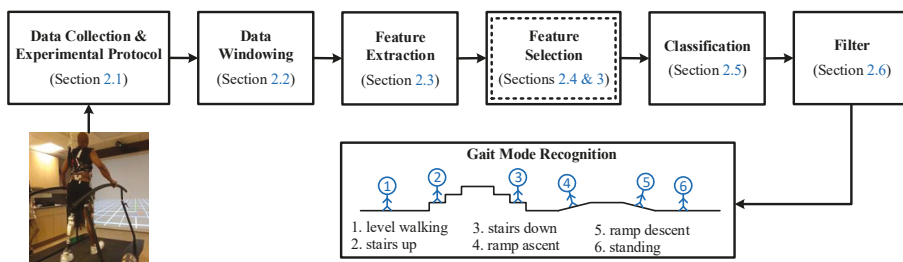
Figure 1. Architecture of user intent recognition system. The double-lined box indicates that an evolutionary algorithm is used for optimization.

2.1. Data Collection and Experimental Protocol

Data collection can significantly affect the accuracy of UIR. Input signals must be informative enough to accurately discriminate between various human gait modes. In this paper, we collect vertical hip position and thigh angle to indicate the state of the residual limb, and thigh moment to indicate the user–prosthesis interaction. These signals are like an implicit communication link between the user and the prosthesis and can be used to infer user intent.

To design and evaluate the performance of the UIR system, we collect these signals from three able-bodied subjects (AB01, AB02, and AB03) and three transfemoral amputee subjects (AM01, AM02, and AM03). All the experiments were approved by the Department of Veterans Affairs Institutional Review Board. The above-knee amputees wore an Ottobock prosthesis on the right leg. Data were collected for able-bodied subjects during four different activity modes: (1) standing (ST), (2) normal walking (NW) at user-preferred speed (PS), (3) slow walking (SW), and (4) fast walking (FW). We asked subjects to walk slower and faster than their normal walking speed for SW and FW modes, respectively, allowing them to choose comfortable velocities for these two modes. Due to physical limitations, we collect data during only three activity modes for the amputee subjects: ST, NW, and SW. Table 1 shows the physical characteristics of the subjects.

Table 1. Physical characteristics of the six human test subjects. AB and AM represent able-bodied and amputee subject, respectively.

	Gender	Age (years)	Weight (kg)	Height (cm)	Walking Speed (m/s)		
					SW	PS	FW
AB01	Male	37	79.5	188	0.98	1.30	1.63
AB02	Male	20	73.9	172	0.86	1.15	1.44
AB03	Male	28	80.9	179	0.75	1.00	1.25
AM01	Male	32	79.1	174	0.60	1.00	-
AM02	Male	64	99.2	177	0.56	0.94	-
AM03	Male	35	81.7	176	0.60	0.90	-

The data were collected at the Motion Study Laboratory of the Cleveland Department of Veterans Affairs Medical Center with 47 reflective markers on each subject's body. Subjects were asked to walk on a treadmill with built-in force sensors. A 16-camera Vicon system (Denver, CO, USA) recorded kinematic data at 100 Hz. Ground reaction force along three axes were collected from the force sensors at 1000 Hz. Data were filtered with a second-order low-pass filter with a cutoff frequency of 6 Hz. A 3D biomechanical rigid body model was constructed from the marker data, and segmental and joint kinematics (joint displacements) and kinetics (joint moments) were computed as inputs for the UIR system. Detailed methods and sample results can be found in [32]. The experimental setup is illustrated in Figure 2. Note that the lower-limb amputee demographic at the Veterans Affairs Medical Center, where data collection was performed, is dominated by males. Over 98% of veterans who underwent amputation in 2011 were male [33]. Our future work will need to include more subjects and wider demographics (for example, ages and genders).

Note that, in the real-world, non-laboratory settings, we would measure the required input signals directly rather than with cameras. For example, we could use piezo-electric sensors or multi-axis load cells for force sensing, and optical encoders or inertial measurement units for accurate position and angle sensing during stance and swing phases [7,11].

The able-bodied subjects were asked to perform four sequences of walking trials, each lasting approximately 60 s. Each sequence consists of four different gait modes (ST, SW, NW, and FW) and each mode was maintained for several seconds. Figure 3 illustrates a sample walking trial for AB01. The amputee subjects performed six sequences of three different walking modes, each lasting approximately 30 s.

In summary, we note a few important points. Firstly, in this paper, the type of walking activities used for recognition is not our main focus, but rather the assessment of the proposed methodology to eliminate irrelevant/redundant features for UIR is our main goal. Secondly, in human activity mode recognition applications, an entire stride is typically used for non-real-time classification [34]. However, in UIR, we use a small window of measurement signals, mostly within a few milliseconds, to identify user's intent for real-time prosthesis control.

Figure 2. Experimental setup for data collection. The **left** figure shows an able-bodied subject and the **right** figure shows an amputee subject with an Ottobock prosthesis on the right leg.
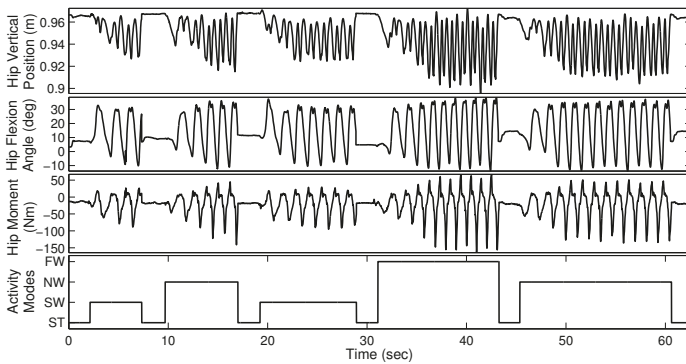
Figure 3. Sample walking trial with four different gait modes for able-bodied subject AB01. Although data is available for both legs, we require only one side for gait mode recognition. The data from the two legs look similar because of gait symmetry.

2.2. Data Windowing

To effectively classify human gait modes, we extract appropriate features from a frame (window) of measurement signals. L_f is the length of a frame in milliseconds. The number of samples in the frame depends on the frame length L_f and the sampling rate. A short frame fails to provide an informative data set and may lead to significant classification bias and variance. On the other hand,

a long frame is a computational burden for real-time implementation. In this paper, L_f is chosen to trade off feature informativeness and computational load.

We apply two different methods for data windowing: disjoint windowing and overlapped windowing [35]. Figure 4 illustrates the two windowing approaches. In disjoint windowing, the class outcome O_i corresponding to frame S_i is output every L_f ms. τ is the time required for feature extraction, classification, commanding the appropriate low-level controller, and prosthesis response time. In overlapped windowing, we use a sliding frame with length L_f and increment I, and the class outcome is output every I ms. Disjoint windowing is a special case of overlapped windowing when $I = L_f$. To achieve real-time operation, the parameters of the windowing approaches should satisfy

$$
\begin{aligned}
\tau \le L_f \qquad &\text{disjoint windowing,} \\
\tau \le I \le L_f \quad &\text{overlapped windowing.}
\end{aligned}
\tag{1}
$$

In this paper, we apply disjoint and overlapped windowing with various frame and increment lengths. We consider two important characteristics to determine L_f [35]: (1) the minimum interval between two distinct muscle contractions is 200 ms [36], and (2) the delay between user intent and the resultant prosthesis motion should be no more than 300 ms [37,38]. The first property implies that a 200 ms frame of data should have the potential to provide informative features for gait mode classification. The second property, which is known as the real-time constraint, ensures that the amputee will experience the prosthesis as responsive to his or her intent. The real-time constraint requires $\tau \le L_f \le 300$ ms for disjoint windowing, and $I \le 300$ ms for overlapped windowing. Therefore, we use overlapped windowing when the frame length is larger than 300 ms, noting that a larger frame will require a higher computational load.

(**a**) Disjoint windowing   (**b**) Overlapped windowing
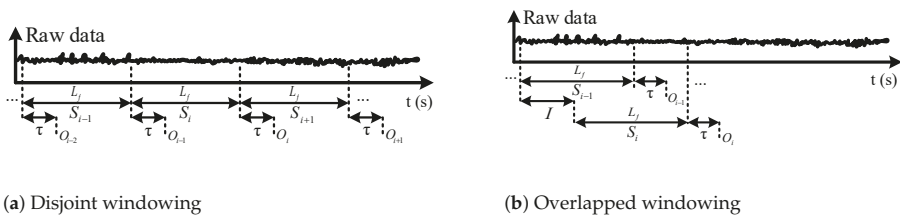
Figure 4. Data windowing. S_i represents the i-th data segment, L_f is the frame length, τ is the required processing time, I is the increment length for overlapped windowing, and O_i is the detected gait mode corresponding to frame S_i.

2.3. Feature Extraction

Various features can be extracted from a frame of measurement data and used for classification. Features should be informative enough to discriminate between various gait modes. In addition, feature extraction needs to be computationally fast for real-time implementation. In general, both time-domain (TD) and frequency-domain (FD) features are frequently used for classification [35,39,40]. We compare TD and FD features in this paper, and select the optimal subset of features for UIR.

TD features are computationally fast, and include information about the data waveform and frequency. We extract the following TD features from each frame of data: slope sign change (SSC), zero crossing (ZC), waveform length (WL), variance (VAR), mean absolute value (MAV), modified MAV (MAV1 and MAV2), root mean square (RMS), Willison amplitude (WAMP), skewness (SK), kurtosis (KU), and correlation coefficient (COR) and angle (ANG) between two frames of data from different measurement signals. The mathematical definitions of these TD features are given in [39].

In addition, multiple FD features have been extracted. FD features are computationally slower than TD features, but include information about the frame's frequency content. We use periodograms

to measure the power spectrum density (PSD) of a frame, and calculate the following FD features: mean frequency (MNF), median frequency (MDF), maximum frequency (MAXF), and fourth-order auto-regressive coefficients (AR4). The mathematical definitions of these FD features are given in [39].

Previous research has shown the applicability of these TD and FD features for prosthetic limb pattern recognition [39–41]. Therefore, we are motivated to investigate the performance of these features for gait mode recognition.

After extracting TD and FD features from a frame of measurement data, the features are concatenated and labeled to create a single training pattern. For instance, extraction of VAR, MAV + RMS, and AR4 features from a frame of three measurement signals (e.g., vertical hip position, thigh angle, and thigh moment) would produce a training vector with 3, 6, and 12 elements, respectively. We perform the above procedure for all features and all frames of measurement data to create the training data set. The training set is then normalized to equalize the relative magnitude of each feature.

2.4. Feature Selection

The objective of feature selection is to find a subset of the features that were obtained with the feature extraction method. The feature selection method attempts to find a parsimonious feature subset that results in accurate classification. However, subset size and classification accuracy are conflicting objectives. A small feature subset will probably result in high classification error, whereas a large feature subset will probably result in lower classification error. Therefore, feature selection can be viewed as a multi-objective optimization (MOO) problem. In MOO problems, no single solution can simultaneously optimize all objectives. The solutions comprise a set of possible alternative solutions known as the optimal Pareto set [21].

We seek the most informative but parsimonious subset of features for gait mode classification. Note that exhaustive search is not practical in cases with a high-dimensional set of features. A set of n features has $2^n - 1$ different subsets (excluding the null subset). Many heuristic search strategies, such as sequential forward selection, sequential backward elimination, and evolutionary search, have been suggested for this type of combinatorial problem [42]. Evolutionary algorithms (EAs) have been demonstrated as an efficient global search strategy for feature selection [43]. They generally outperform sequential forward selection and sequential backward elimination [22]. However, EA-based search strategies are computationally expensive due to the need for many cost function evaluations. To reduce computational complexity, we propose a new method called gradient-based multi-objective feature selection (GMOFS) for UIR. In addition, we propose the application of four EA-based search strategies, using multi-objective biogeography-based optimization (MOBBO), for feature selection. We then use two systematic approaches to compare the performance of the GMOFS search strategy with four variants of MOBBO.

2.5. Classification

Accurate classification of gait patterns is the ultimate goal of the user intent recognition (UIR) system. For this purpose, we assess various well-known linear and nonlinear classification techniques, including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM), decision tree (DT), and multi-layer perceptron (MLP) classifiers.

LDA and QDA classifiers do not require time-consuming iterations for training. In fact, the parameters of these classifiers are directly obtained from the training data. Although these classifiers are fast in terms of training, they are not as flexible as nonlinear classifiers such as SVM, DT, and MLP. These classifiers solve an optimization problem that minimize the classification error. In most cases, it is difficult to find optimization solutions in closed form, so we use either gradient-based optimization algorithms such as steepest descent, or evolutionary algorithms (EAs).

We use one-against-one approach to implement multi-class SVM, and we also evaluate the performance of different kernels, such as linear and RBF. We tune the parameters of the SVM kernels to achieve the best classification performance. To increase the accuracy of the MLP network, we perform

a grid search of the number of hidden nodes p from the set $\{3, 4, 5, 6, 8, 10, 15, 20\}$, and we measure the mean classification error using five-fold cross validation (CV). Then, we choose p to obtain a trade-off between classification accuracy and classifier complexity. An MLP with small p may not result in the desired accuracy, but an MLP with large p may tend to memorize the noise in the training set and lead to overfitting and poor generalization. In addition, we use Wilcoxon signed-rank tests to statistically compare the classification methods.

2.6. Filter

To enhance the prediction performance of the UIR system, we incorporate a majority voting filter (MVF). MVF alleviates transient jumps between classifier output classes and leads to smooth transitions from one classified gait mode to another. We implement the MVF using $2q + 1$ classified modes [41]: the current, q previous, and q subsequent values. The MVF output is the most frequently classified mode among those $2q + 1$ values.

To obtain q subsequent classified gait modes, MVF enforces a time delay to allow the classifier to access the required data. The real-time constraint discussed in Section 2.2 requires a time delay less than 300 ms. Therefore, an appropriate value for q should be chosen to avoid violation of the real-time constraint. The constraint requires

$$
\begin{aligned}
q \times L_f &\leq 300 \text{ ms} \qquad \text{Disjoint windowing,} \\
q \times I \; &\leq 300 \text{ ms} \qquad \text{Overlapped windowing.}
\end{aligned}
\tag{2}
$$

An MVF with very small q may not significantly improve classification performance, whereas an MVF with very large q may cause misclassification because of time delay. In this paper, we will choose a trade-off value for q, and will investigate the effect of the MVF on classification performance.

3. Feature Selection Algorithm Development

In this section, we propose a search strategy based on biogeography-based optimization (BBO), in addition to a new gradient-based algorithm, to find an optimal subset of features for designing UIR with high accuracy and low complexity. We then use two systematic approaches to compare the performance of the different search strategies.

3.1. Biogeography-Based Multi-Objective Optimization

We propose the application of multi-objective biogeography-based optimization (MOBBO) for feature selection of the UIR system. The dimension of the optimization problem is equal to the number of available features (independent variables). Each feature is represented by a binary value where 1 indicates that the feature is used for classification, and 0 indicates otherwise. Therefore, each individual in the MOO algorithm is a binary sequence with length equal to the problem dimension. We evaluate the following two objective functions for all individuals in the population:

$$
\begin{aligned}
f_1^i &= \text{number of selected features in the } i\text{-th individual,} \\
f_2^i &= \text{average prediction error using } c\text{-fold cross validation,}
\end{aligned}
\tag{3}
$$

where $i = 1, \cdots, N$, and N is the population size. We combine BBO with four MOO algorithms [21] to obtain vector evaluated BBO (VEBBO), non-dominated sorting BBO (NSBBO), niched Pareto BBO (NPBBO), and strength Pareto BBO (SPBBO). We apply these MOBBO variants to find the optimal Pareto set for the optimization problem. We investigate the performance of each method in a later section.

We use linear discriminant analysis (LDA) to compute f_2^i. LDA is widely used with evolutionary algorithms to evaluate the quality of a candidate subset for feature selection [44]. LDA does not require time-consuming iterations to build a model. This point is important because EAs require

many objective function evaluations to find the solution. In Section 4.2, we will demonstrate that all feature selection approaches presented in this paper are able to find the most significant features, even though they use different selection criteria and machine learning methods. It is possible to use either classification accuracy or error as the quality measure for the second objective. We use average classification error of *c*-fold cross validation (CV). In *c*-fold CV, we randomly divide the training set into *c* distinct folds. Then, we repeat training *c* times; each time, the model is trained using $c - 1$ folds and is tested with the remaining fold. The average of the *c* classification errors is used as the quality measure.

3.2. Gradient-Based Multi-Objective Feature Selection

Although MOBBO and other gradient-free MOO methods have the potential to find the globally optimal solution [22,23], they are computationally expensive due to the need for many iterations of the classifier training process (multiple individuals in the population, and multiple generations). To reduce computational complexity, we propose a novel algorithm called gradient-based multi-objective feature selection (GMOFS). In GMOFS, feature selection and data classification are performed simultaneously.

GMOFS incorporates a regularization penalty term to the optimization problem of its learning algorithm. The penalty term, which is handled by a Lagrange multiplier, directs the trained model toward a parsimonious as well as accurate model. We use an MLP network as the classifier, and include an elastic net to penalize the size of the selected feature subset. The first step of GMOFS is to train a constrained MLP network with the cost function

$$J = \frac{1}{2} \sum_{l=1}^{m} \sum_{j=1}^{K} \left(t_j^{(l)} - o_j^{(l)} \right)^2 + \lambda \sum_{i=1}^{n} \left(\alpha \beta_i^2 + (1 - \alpha) |\beta_i| \right), \tag{4}$$

where β_i is the multiplier of the *i*-th input feature before input to the MLP network; $t_j^{(l)}$ and $o_j^{(l)}$ are the target and actual value, respectively, of the *j*-th output neuron associated with the *l*-th training pattern; K is the number of output neurons (classes); m is the number of training patterns; and n is the number of input features. We use an MLP network with one hidden layer and p hidden neurons (including the bias node). v_{ih} denotes the weight that connects the *i*-th input neuron to the *h*-th hidden neuron, and w_{hj} denotes the weight that connects the *h*-th hidden neuron to the *j*-th output neuron. The first term of the cost function penalizes classification error while the second term, which is the elastic net, penalizes the number of selected features. The elastic net is a convex combination of ridge regression ($\alpha = 1$) and least absolute shrinkage and selection operator ($\alpha = 0$). $\lambda \geq 0$ is a complexity parameter that controls the shrinkage of the input features. Large λ leads to a shrinkage of β_i toward zero, which implies that the input feature corresponding to β_i is not significant. However, as shrinkage increases, classification error tends to increase. Therefore, λ provides a trade-off between classification error and the number of selected features. In summary, the construction of the MLP network with the elastic net is formulated as the following optimization problem:

$$\min_{\beta, w, v} J \quad \text{subject to} \quad \begin{cases} 0 \leq \beta_i \leq 1, \\ |w_{hj}| \leq a \text{ and } |v_{ih}| \leq b, \end{cases} \tag{5}$$

for all *i*, *j*, *h*, where $\beta_i = 0$ or 1 implies that the associated feature is the least or most significant input variable, respectively. a and b are the bounds for neuron weights w_{hj} and v_{ih}, respectively. However, due to the direct relationship between β_i and neuron weights v_{ih} and w_{hj}, we cannot conclude that an input feature with small β_i and large neuron weights is insignificant. To avoid optimization solutions with large weights, neuron weights are constrained. Backpropagation is used to update β_i, v_{ih}, and

w_{hj}. The derivative of J with respect to output weights w_{hj}, hidden weights v_{ih}, and input weights β_i is obtained by the chain rule as

$$\frac{\partial J}{\partial w_{hj}} = \sum_{l=1}^{m} \delta_j^{(l)} y_h^{(l)}$$

$$\frac{\partial J}{\partial v_{ih}} = \sum_{l=1}^{m} \left[\sum_{k \in D_2(h)} \left[\delta_k^{(l)} w_{hk} \right] y_h^{(l)} (1 - y_h^{(l)}) \beta_i x_i^{(l)} \right] \tag{6}$$

$$\frac{\partial J}{\partial \beta_i} = \sum_{l=1}^{m} \left\{ \sum_{s \in D_1(i)} \left[\sum_{k \in D_2(h)} \left[\delta_k^{(l)} w_{sk} \right] y_s^{(l)} (1 - y_s^{(l)}) v_{is} \right] x_i^{(l)} \right\} + \lambda \left[2\alpha\beta_i + (1 - \alpha) \frac{\beta_i}{|\beta_i|} \right],$$

where $D_1(i)$ is the set of middle layer neurons whose inputs come from the i-th input neuron, $D_2(h)$ is the set of output neurons whose inputs come from the h-th middle layer neuron, and $\delta_j^{(l)} = -(o_j^{(l)} - t_j^{(l)})(1 - t_j^{(l)}) t_j^{(l)}$. Detailed derivation of the derivative of J with respect to w_{hj} and v_{ih} is available in [45]. The derivative of J with respect to input weights β_i is straightforward to obtain using the chain rule. We use the derivatives in Equation (6) and constraints in Equation (5) along with the trust region reflective algorithm to train the constrained MLP network.

Once MLP training phase is completed, input weights β_i are sorted in descending order. The input variable with the largest β_i is the most significant feature. The second step of GMOFS is to select the most r significant features, which are associated with the r largest input weights β_i, and which satisfy

$$\frac{\sum_{i=1}^{r} \beta_i}{\sum_{j=1}^{n} \beta_j} \geq 95\%,$$

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n. \tag{7}$$

The 95% threshold value in Equation (7) determines the trade-off between the number of selected features and the accuracy of the model. A low threshold value will decrease the number of selected features and will be more likely to remove informative features that can significantly contribute to the accuracy of the UIR model. On the other hand, a high threshold value will be more likely to include irrelevant features that cannot contribute to the accuracy of the model. To tune the threshold value, we gradually increase the threshold from zero, and, for each value, we compute the accuracies of the trained MLP once with the original β_i values and once with $\beta_i = 0$ for all unselected input features. We increase the threshold value until there is no significant difference between the performances of MLP for these two cases. That point was reached with the threshold value of 95%. We then repeat the first two steps of GMOFS for different λ in the range $[\lambda_l, \lambda_u]$ with a predefined increment $\triangle\lambda$. The selected subsets associated with each λ comprise a population. The population size depends on $\triangle\lambda$. To assess the performance of the selected feature subset, we train a classifier with each selected subset and find classification error. In this population, the subset associated with $\lambda \to \infty$ has minimum size and maximum classification error, whereas the subset with $\lambda = 0$ has maximum size and probably has the lowest classification error. Thus, the size of the selected subset and the classification error, defined in Equation (3), are two conflicting objectives. To find the GMOFS Pareto front, we first obtain the Pareto set as

$$P_s = \left\{ x^* : \left[\nexists x : f_i(x) \leq f_i(x^*) \text{ for all } i \in [1,2], \text{ and } f_j(x) < f_j(x^*) \text{ for some } j \in [1,2] \right] \right\}. \tag{8}$$

x^* denotes the set of non-dominated solutions in the population and $f_i(x)$ is the *i*-th objective function. The Pareto front P_f is obtained from all function vectors $f(x)$ that correspond to the Pareto set:

$$P_f = \{f(x^*) : x^* \in P_s\}. \tag{9}$$

Note that all Pareto points are equally preferable apart from subjective prioritization. The outline of GMOFS is given in Algorithm 1.

Algorithm 1: The outline of gradient-based multi-objective feature selection (GMOFS), where x_i is the *i*-th feature in the training set X, and Y is the corresponding set of output classes.

Initialization: $\lambda = \lambda_l \leq \lambda_u$, Population $= \emptyset$, $k = 1$
While $\lambda \leq \lambda_u$
    Step 1:
    Use the training data $\{X, Y\}$ to train the constrained MLP network in Equation (4)
    by solving Equation (5)
    Step 2:
    Sort the input weights $\{\beta_i\}$ in descending order
    Use Equation (7) to select subset $S_k \subset X$ where $\text{size}(S_k) \leq \text{size}(X)$
    Step 3:
    Population $\leftarrow$ Population $+ S_k$
    $k \leftarrow k + 1$
Next $\lambda \leftarrow \lambda + \triangle\lambda$
Step 4:
For each subset S_k in Population
    Use cross-validation to train and test a classifier with dataset $\{S_k, Y\}$
    Calculate objective functions f_1^k and f_2^k using Equation (3)
Next subset S_k
Step 5:
Find the Pareto set using Equation (8)

3.3. Evaluation of Multi-Objective Optimization Pareto Fronts

We will compare the Pareto fronts obtained by each MOO algorithm using normalized hypervolume and relative coverage. These methods are popular for evaluating the quality of a Pareto front. The Pareto front normalized hypervolume is computed as follows:

$$\text{Normalized Hypervolume} = \frac{\sum_{j=1}^{N_p} \prod_{i=1}^{M} f_{ji}}{N_p}, \tag{10}$$

where M is the number of objectives, f_{ji} is the value of the *i*-th objective function of the *j*-th Pareto point, and N_p is the number of Pareto points.

Another way to compare Pareto sets is to compute the coverage of one Pareto set relative to a second Pareto set. This metric is determined by the number of solutions in the first Pareto set that are weakly dominated by at least one solution in the second Pareto set [21]. A smaller number for normalized hypervolume and relative coverage indicates better performance.

4. Results and Discussion

This section evaluates the performance of the user intent recognition (UIR) system and its subsystems as discussed in Sections 2 and 3.

4.1. Effect of Frame Length on Classification Performance

The objective of this section is to choose the appropriate data windowing method and frame length. We investigate the influence of disjoint and overlapped windowing with different frame lengths on the classification accuracy of the UIR system. We use disjoint windowing with frame lengths $L_f = \{100, 150, 200, 250\}$ ms, and overlapped windowing with frame lengths $L_f = \{200, 250, 300\}$ ms and increments $I = \{50, 150, 200\}$ ms. We extract time-domain (TD) and frequency-domain (FD) features from the data frames generated by the three measurement signals collected from able-bodied subjects. Note that the size of the data set depends on the parameters of the windowing methods. For instance, a 10-second walking sequence with disjoint windows of length 100 ms provides 100 frames and consequently 100 training patterns.

In this experiment, LDA, QDA, MLP, and SVM with RBF kernel are separately trained and tested using 10-fold cross validation (CV) for each subject with only one feature type for each of the measurements: WL, VAR, MAV, RMS, WAMP, ANG, and AR4. These features are used because they are considered the most representative TD and FD features. Figure 5 illustrates the mean classification accuracy of LDA with different frame lengths for able-bodied subjects using 10-fold CV. A single value on the horizontal axis of the figure indicates the frame length of disjoint windowing. A pair of values indicates the frame length and increment length of overlapped windowing; for instance, 200–50 denotes $L_f = 200$ ms and $I = 50$ ms. Figure 5 shows that the classification accuracy typically improves as the frame length increases. We observed similar trends with other classifiers. In fact, classifier type does not influence our choice of frame length. Therefore, we only provide classification results for LDA as a representative result.

Figure 5 shows that a larger frame is more likely to include more information, and consequently lower bias and variance in classification performance. For instance, the increase in accuracy with WL is approximately 18% when the frame length increases from 100 ms to 200 ms. The increases are 11.6% and 9.8% when using the VAR and AR4 features, respectively. However, the accuracy does not vary significantly with frame length for the remaining features in Figure 5. The figure illustrates that all representative TD features except ANG provide better classification performance than AR4, which is the representative FD feature.
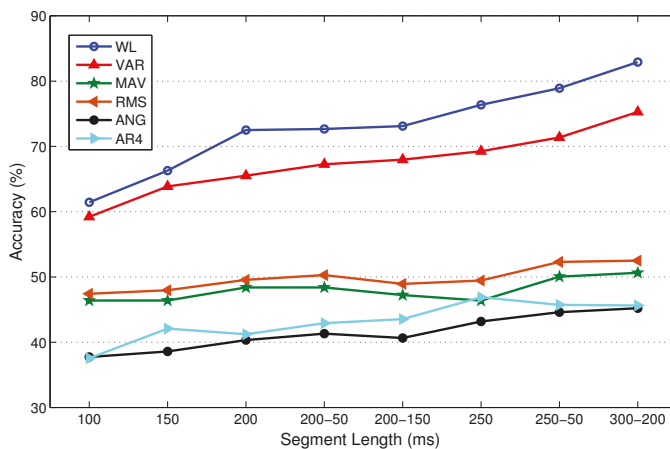
Figure 5. Mean LDA performance for the able-bodied subjects for different data windowing methods and frame lengths. On the horizontal axis, a single value indicates the frame length of disjoint windowing, and a pair of values indicates the frame length and increment length of overlapped windowing. For instance, 200–50 indicates $L_f = 200$ ms and $I = 50$ ms for overlapped windowing.

In this experiment, very small frame length is not used because it would result in poor prediction accuracy. Conversely, large frame length is not used because it would result in a violation of the real-time constraint. To find the best frame length, we statistically compare performance using Wilcoxon signed-rank tests. For this purpose, LDA is trained multiple times, each time using one TD or FD feature type. We perform LDA training individually for each feature type rather than the full feature set to provide a sufficient number of samples for statistical comparison. The null hypothesis of the test is that the differences between mean classification accuracy (average accuracy of all LDAs trained individually with every single TD and FD feature type) corresponding to two different frame lengths are from a distribution with zero mean at the specified level of significance. If the null hypothesis cannot be rejected, then we conclude that the two compared frame lengths are not statistically significantly different, as indicated by an $\approx$ sign and a T (tie). If we can reject the null hypothesis, then the two frame lengths are statistically significantly different, and this is indicated by a $+$ sign. The better frame length is the one with better mean classification accuracy and is shown by B (better) while the worse one is shown by W (worse). Table 2 provides the results of the statistical tests at a 10% significance level.

Table 2 shows that frames with length larger than 200 ms perform better than frames with length 150 or 100 ms. Table 2 shows that the two overlapped frame windows with $L_f = 250$ ms, $I = 50$ ms and $L_f = 300$ ms, $I = 200$ ms tie for similar performance, and perform better than the other frame lengths.

Taking into account the real-time constraint, the length of the MVF filter, and processing time, we choose overlapped windowing with $L_f = 250$ ms, $I = 50$ ms throughout the remainder of the paper as the best trade-off, except where specifically mentioned otherwise.

Table 2. Comparison of mean classification performance for different frame lengths (row values versus column values) using Wilcoxon signed-rank tests at a 10% significance level. Mean classification performance is considered as the average of all linear discriminant analysis (LDA) classifiers trained individually with every single time-domain (TD) and frequency-domain (FD) feature type. $\approx$ indicates that the two compared frame lengths tie (T) with similar performance and are not statistically significantly different. $+$ indicates that the two frame lengths are statistically significantly different, and B or W indicates that the row frame length performs better or worse than the column frame length, respectively. $*$ indicates that the lower triangular half of the table is equal to its upper triangular half.

Frame Length (ms)		150	200	200–50	200–150	250	250–50	300–200
100	vs.	W (+)	W (+)	W (+)	W (+)	W (+)	W (+)	W (+)
150	vs.	—	W (+)	W (+)	W (+)	W (+)	W (+)	W (+)
200	vs.	$*$	—	W (+)	T ($\approx$)	W (+)	W (+)	W (+)
200–50	vs.	$*$	$*$	—	T ($\approx$)	T ($\approx$)	W (+)	W (+)
200–150	vs.	$*$	$*$	$*$	—	W (+)	W (+)	W (+)
250	vs.	$*$	$*$	$*$	$*$	—	W (+)	W (+)
250–50	vs.	$*$	$*$	$*$	$*$	$*$	—	T ($\approx$)

We use principal component analysis (PCA) [46] and Fisher linear discriminant analysis (FLDA) [47] to visualize the training set. A training pattern is a vector of all TD and FD features extracted from a frame of raw data. We performed 2D dimension reduction for three different frame lengths. Figure 6 illustrates the 2D scatter plot for able-bodied subject AB01. To save space, we do not provide the same figures for other subjects because we obtained similar results for those subjects.

Figure 6 shows that FLDA provides better visualization than PCA in terms of gait mode separability. We verify that longer frame length leads to better gait mode separation, and eventually better classification performance. Most importantly, Figure 6 verifies the effectiveness of the TD and FD features.
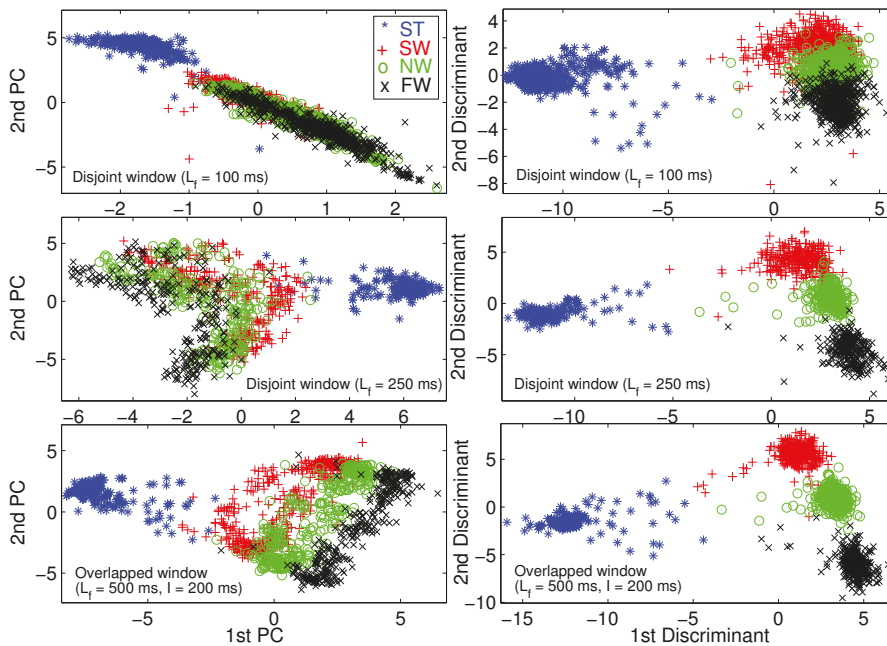
Figure 6. Two-dimensional scatter plot for visualization using principal component analysis (PCA) (**left column**) and Fisher linear discriminant analysis (FLDA) (**right column**) for able-bodied subject AB01.

4.2. Multi-Objective Feature Selection

We perform feature selection in two steps. In the first step, we exclude non-informative time-domain (TD) and frequency-domain (FD) features, and, in the second step, we use multi-objective optimization (MOO) to further refine the selected feature set. We implement five different MOO methods, including our newly proposed method, to find the optimal set of features. The complete set of features includes the following TD and FD features: F_1: slope sign change (SSC); F_2: zero crossing (ZC); F_3: waveform length (WL); F_4: variance (VAR); F_5: mean absolute value (MAV); F_6 and F_7: modified MAV (MAV1 and MAV2); F_8: root mean square (RMS); F_9: Willison amplitude (WAMP); F_{10}: skewness (SK); F_{11}: kurtosis (KU); F_{12}: median frequency (MDF); F_{13}: mean frequency (MNF); F_{14}: maximum frequency (MAXF); F_{15} and F_{16}: correlation coefficient (COR) and angle (ANG) between two frames of data, respectively; and F_{17}: fourth-order auto-regressive coefficients (AR4). We perform preliminary feature selection using able-bodied subjects. We use MOO for final feature selection using only one able-bodied subject to reduce computational effort. In Section 4.4, we will investigate the performance of the selected features with other able-bodied subjects and with the amputee subjects. We note that the optimal feature subset may vary depending on subjects. However, in this paper, we are particularly interested in obtaining an optimal feature subset from able-bodied subjects, and assessing its performance on amputee subjects. The reason for this approach is that able-bodied subjects' data are more accessible for UIR training in real-world applications. Our results in Section 4.4 will show no significant difference between UIR performances for amputee subjects when trained with the optimal feature subset and when trained with the full feature set. Future work could compare optimal subsets of features obtained from different individuals.

In the first step, we train LDA, QDA, SVM-Linear, SVM-RBF, and MLP for each of three able-bodied subjects, and separately for each individual feature type listed in the previous paragraph,

using 10-fold cross-validation (CV) for each training procedure. The mean classification accuracy over the three subjects and the ten folds are used to assess the importance of each feature type.

Figure 7 shows the mean classification accuracy and processing ratio for each feature type. *Processing ratio* indicates the relative computational load to compute each feature type—for instance, the percentage of computational load required to compute F_{17} over the computational load required to compute all feature types is 4.22%. To reduce clutter in the figure, we show LDA results as a representative of QDA and SVM-Linear since they had similar performance. Similarly, we show SVM-RBF as a representative of MLP. Figure 7 shows consistent performance of different classifiers in terms of prioritizing various feature types. Figure 7 indicates that TD features require less computational effort than FD features. For instance, F_{12}, F_{13}, and F_{14} require high computational effort compared to other features. We exclude F_{12}, F_{13}, and F_{14} from the candidate feature set due to their relatively high computational expense and poor classification accuracy. In addition, F_6 and F_7, two variants of MAV, are excluded due to their poor classification accuracy and because they provide information that is similar to MAV. Therefore, we exclude a total of five weak feature types, and pre-select the remaining 12 feature types. This results in a training vector with $11 \times 3 + 4 \times 3 = 45$ elements. Note that the AR4 feature type includes four components and thus contributes a total of 12 elements from the three measurement signals. Finally, vertical hip position does not cross zero (see Figure 3), thus the number of zero crossing (ZC) feature of this signal is excluded. To verify the lack of information in the eliminated features, we found that the combination of the excluded features with the pre-selected features did not significantly enhance classification performance. In summary, we have a training data set with 44 features.
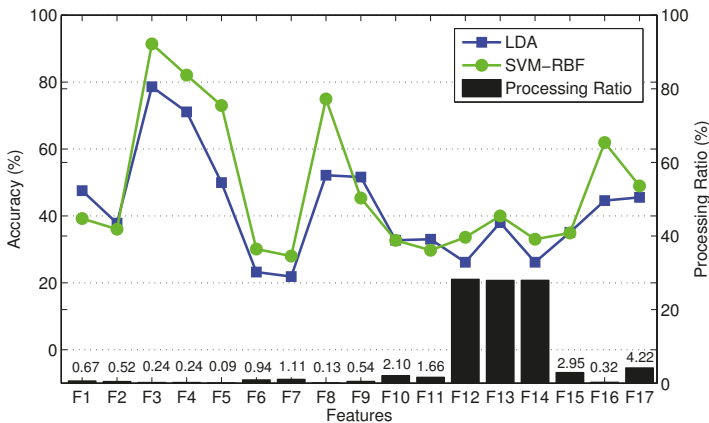
Figure 7. Mean classification accuracy of three able-bodied subjects, and processing ratio of 17 feature types trained by LDA using 10-fold cross validation.

Now, we are ready to proceed to the feature selection step. In this step, we use vector evaluated BBO (VEBBO), non-dominated sorting BBO (NSBBO), niched Pareto BBO (NPBBO), strength Pareto BBO (SPBBO), and gradient-based multi-objective feature selection (GMOFS) to select an optimal subset from the 44 pre-selected features. To reduce the computational expense, we use only the AB01 training data set in this step. We then verify that the selected subset results in a satisfactory UIR system when trained for other subjects. Table 3 shows the tuning parameters used in this paper. To tune the parameters, we performed a sensitivity analysis of multi-objective optimization (MOO) performance to each parameter, one at a time, to find a local optimum of MOO performance with respect to each parameter. For instance, GMOFS is implemented with different elastic net parameter values $\alpha = \{0, 0.5, 1\}$, and we found that the Pareto front with $\alpha = 0$ dominates Pareto fronts that are found

with other values of α. For training the neural network in GMOFS, we used the MATLAB function fmincon from the Optimization Toolbox (R2014a, MathWorks, Natick, MA, USA) to implement a trust region reflective algorithm. We mostly used default values for the fmincon parameters, but we found that the performance of GMOFS is not very sensitive to these parameters.

We run each multi-objective method for 10 independent trials, and the best Pareto front of each method is selected for MOO comparison. Results show that the GMOFS Pareto front statistically significantly dominates all four multi-objective biogeography-based optimization (MOBBO) Pareto fronts. We note that GMOFS and the MOBBO variants use different classifiers for feature selection, namely, MLP and LDA. To obtain a fair comparison of the new components of GMOFS with the MOBBO variants, we decouple the search strategy from classification performance. Note that LDA, which is used in MOBBO, is one of the most popular classification algorithms and has been widely used with evolutionary algorithms for feature selection due to its good performance and simplicity [44].

Table 3. Tuning parameters for multi-objective feature selection.

	Symbol	Value
MOBBO		
Mutation rate	μ	0.04
Number of elites	E	2
Population size	N	100
Number of generations	Gen	1000
Problem dimension	d	44
Migration model	m_{flag}	sinusoidal
GMOFS		
Number of hidden nodes	p	5
Elastic net parameter	α	0
Bound for shrinkage parameter	$[\lambda_l, \lambda_u]$	[0, 150]
Bound for neuron weights	a, b	5
Increment of shrinkage parameter	$\triangle \lambda$	1 if $0 \leq \lambda \leq 30$; and 10 if $30 < \lambda \leq 150$
Trust region reflective		
Maximum allowable iterations	MaxIter	100
Termination tolerance on the independent variable	TolX	0.001
Termination tolerance on the cost function	TolFun	0.001
Typical values for the independent variable	TypicalX	0.1
Finite difference method	FinDiffType	central

To conduct the fair comparison, we apply SVM with linear kernels to all of the optimal feature subsets found by the MOO methods. Figure 8a illustrates the Pareto fronts obtained by the five MOOs with SVM with linear kernels. Figure 8a shows that the Pareto fronts of VEBBO, SPBBO, NSBBO, and GMOFS are close, and clearly dominate the NPBBO Pareto points. Figure 8b indicates the combined Pareto front obtained from all of the non-dominated points in Figure 8a. GMOFS provides the maximum contribution to the combined Pareto front, while NPBBO does not contribute any Pareto points. All of the points in Figure 8b are labeled for easy referencing.

To systematically compare the Pareto fronts in Figure 8a, we use relative coverage and normalized hypervolume as discussed in Section 3.3. Tables 4 and 5 provide the comparison results using these two approaches. In Table 4, an entry in column i and row j ($i \neq j$) indicates the percentage of Pareto points of the method of column i that is dominated by at least one Pareto point of the method of row j. We see that, on average, only 7.2% of the Pareto points of VEBBO are weakly dominated by at least one Pareto point from the other four MOO methods. Therefore, VEBBO ranks first in terms of relative coverage. GMOFS ranks second and performs better than SPBBO, NSBBO, and NPBBO. In addition, Table 5 shows that VEBBO and GMOFS rank first and second in terms of normalized hypervolume,

respectively. GMOFS ranks first in terms of the number of Pareto points. These results verify the competitive performance of GMOFS compared to the other four MOO methods.

Most importantly, in terms of the advantage of GMOFS, it requires the execution of only 43 classifier training procedures (due to the number of λ increments), while each of the other four EA-based MOO methods require 100,000 training procedures (due to the combination of population size and generation limit).
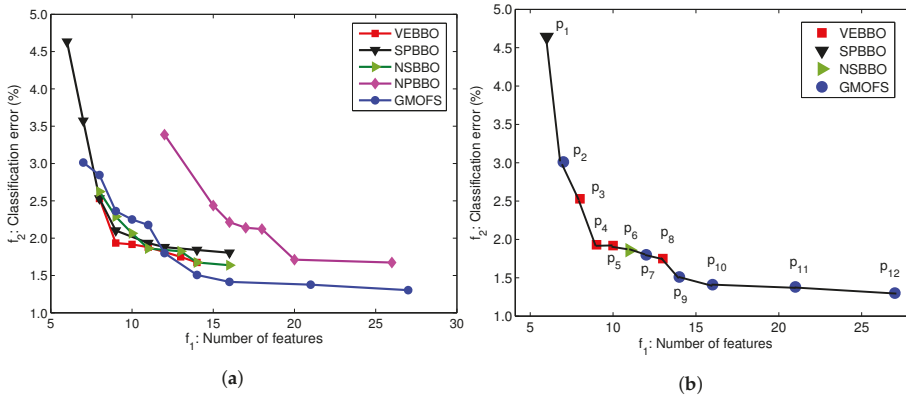
Figure 8. (**a**) Pareto fronts obtained from MOO methods with an SVM classifier with linear kernels using AB01 training data; (**b**) combined Pareto front obtained from non-dominated Pareto points in (**a**).

Table 4. Comparison of Pareto fronts using relative coverage (RC). Only 7.2% and 30% of the VEBBO and GMOFS points, respectively, are dominated by other Pareto points; so VEBBO and GMOFS rank first and second, respectively, in terms of RC.

	VEBBO	SPBBO	NSBBO	NPBBO	GMOFS
VEBBO	—	62.5	75.0	85.7	40.0
SPBBO	0.0	—	25.0	71.4	40.0
NSBBO	14.3	50.0	—	100.0	40.0
NPBBO	0.0	0.0	0.0	—	0.0
GMOFS	14.3	50.0	50.0	100.0	—
Mean RC (%)	7.2	40.4	37.5	89.3	30.0

Table 5. Comparison of Pareto fronts using normalized hypervolume. N_p is the number of Pareto points obtained by each MOO method. VEBBO and GMOFS rank first and second, respectively, in terms of normalized hypervolume, and GMOFS ranks first in terms of the number of points.

	N_p	Normalized Hypervolume
VEBBO	7	0.5026
SPBBO	8	0.5814
NSBBO	8	0.5676
NPBBO	7	0.8013
GMOFS	10	0.5332

The benefit of presenting the data of Figure 8b is that it allows us to find the best subset of features for an accurate and parsimonious classifier. Among the 12 Pareto points, we choose p_9 as a potential candidate solution. We could pick any other solution from the Pareto front depending on the priority of the problem objectives, but p_9 provides a good trade-off between classification error and number of features. Therefore, in Section 4.4, we will investigate classification performance with candidate solution p_9 for all human subject data AB01, AB02, AB03, AM01, AM02, and AM03. However, first, we will find the best classifier in the following section.

Figure 9 shows that the feature selection frequencies of GMOFS and VEBBO, taken across all Pareto points, are different. However, they both select significant features at a high frequency. There are five common features that appear in all of the GMOFS, VEBBO, NSBBO, NPBBO, and SPBBO Pareto points: WL from vertical hip position and thigh angle (features 6 and 7), VAR from thigh moment (feature 11), WAMP from vertical hip position (feature 18), and ANG from thigh moment (feature 32). Therefore, all five feature selection methods value the most informative features regardless of their selection criterion and machine learning method. For example, Pareto point p_8 (obtained by VEBBO combined with the LDA classifier) and candidate solution p_9 (obtained by the GMOFS combined with MLP classifier) have nine features in common out of a total of 13 and 14 features, respectively.

(**a**) VEBBO   (**b**) GMOFS
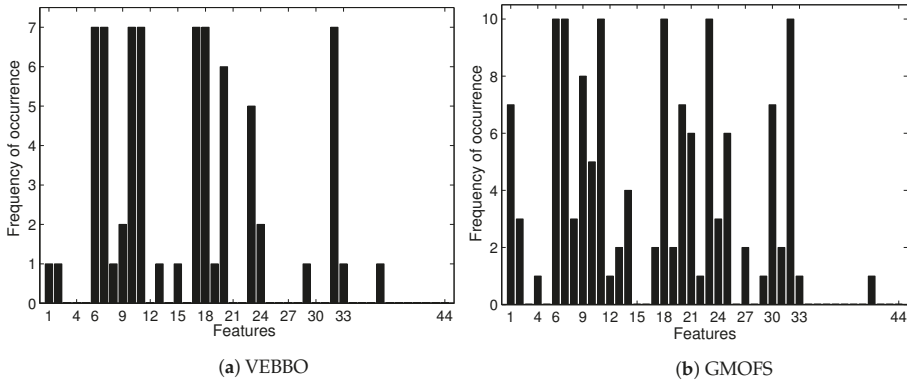
Figure 9. Selection frequency of 44 features by VEBBO and GMOFS. The plots show how many times each feature appears in the Pareto points of the given method. For instance, feature 6 is present in all 10 GMOFS Pareto points.

4.3. Comparison Results of Classification Algorithms

In this section, we use p_1 through p_{12} to statistically compare the performance of different classifiers for subject AB01. The objective is to find the best classifier for locomotion mode detection among linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM) with linear kernels (SVM-linear), SVM with RBF kernels (SVM-RBF), multi-layer perceptron (MLP), and decision tree (DT). The tuned parameter value of RBF kernel is $\sigma = 1$. Table 6 shows mean classification accuracy and standard deviation of each classifier trained with the features from each Pareto point using 10-fold cross validation (CV).

Table 6. Mean classification accuracy (ACC) and standard deviation (STD) for AB01 of classifiers trained with 13 different feature subsets. NF is the number of features in each set.

Pareto Point	NF	LDA		QDA		SVM-Linear		SVM-RBF		MLP		DT	
		ACC	STD	ACC	STD	ACC	STD	ACC	STD	ACC	STD	ACC	STD
p_1	6	93.56	0.740	94.33	0.852	95.37	1.218	98.33	0.421	97.34	0.698	96.15	1.16
p_2	7	95.31	0.829	96.06	0.711	96.99	0.775	98.88	0.216	98.29	0.539	96.35	1.00
p_3	8	96.69	0.835	96.82	0.410	97.47	0.694	98.86	0.378	98.20	0.411	96.67	1.18
p_4	9	96.86	0.684	96.95	0.484	98.07	0.576	99.31	0.234	98.34	0.459	96.73	1.25
p_5	10	97.04	0.657	96.99	0.427	98.08	0.430	98.90	0.406	98.47	0.645	96.56	1.22
p_6	11	96.84	0.536	97.15	0.654	98.14	0.692	98.94	0.293	98.62	0.578	96.32	1.28
p_7	12	96.93	0.656	97.36	0.372	98.20	0.497	99.05	0.356	98.87	0.406	97.21	1.00
p_8	13	96.61	0.384	97.62	0.554	98.25	0.534	99.14	0.305	95.76	9.180	96.86	1.36
p_9	14	96.76	0.485	97.79	0.311	98.49	0.500	98.88	0.290	98.90	0.355	97.15	0.78
p_{10}	16	96.95	0.525	97.84	0.578	98.59	0.467	98.38	0.449	98.66	0.371	96.99	0.72
p_{11}	21	97.13	0.501	97.93	0.716	98.62	0.564	98.40	0.392	99.00	0.432	97.30	0.54
p_{12}	27	97.41	0.568	97.77	0.861	98.70	0.422	97.58	0.663	99.07	0.373	96.91	0.64

Table 7 presents pairwise statistical comparisons using Wilcoxon signed-rank tests at a 5% significance level. If a pairwise *p*-value is less than 0.05, the mean performances of the two classifiers are statistically significantly different, and the classifier with larger mean prediction accuracy performs better than the other one. A pairwise *p*-value greater than 0.05 indicates no significant difference between the performance of the two classifiers. Table 7 shows that the classification performance of MLP and SVM-RBF are statistically equal, and are significantly better than the other methods. SVM-linear is statistically better than LDA, QDA, and DT. QDA performs better than LDA and similarly to DT. In summary, MLP and SVM-RBF are the best, SVM-linear is the second best, QDA and DT are the third best, and LDA is the worst for locomotion mode detection.

Table 7. Comparison of classification performance using Wilcoxon signed-rank tests (W.T.) at a 5% significance level. B or W indicates that the row method performs better or worse than the column method, respectively, while T shows that they tie with similar performance. ∗ indicates that the lower triangular half of the table is equal to its upper triangular half. These results are obtained using all the data from Table 6.

		DT		SVM-RBF		SVM-linear		QDA		LDA	
		p-Value	W.T.	*p*-Value	W.T.	*p*-Value	W.T.	*p*-Value	W.T.	*p*-Value	W.T.
MLP	vs.	2.44×10^{-4}	B	7.32×10^{-1}	T	8.50×10^{-3}	B	5.02×10^{-3}	B	2.44×10^{-4}	B
DT	vs.	−		1.23×10^{-4}	W	8.20×10^{-3}	W	1.33×10^{-1}	T	1.70×10^{-1}	T
SVM-RBF	vs.	∗		−		6.70×10^{-3}	B	2.44×10^{-4}	B	1.22×10^{-4}	B
SVM-linear	vs.	∗		∗		−		1.15×10^{-4}	B	1.25×10^{-4}	B
QDA	vs.	∗		∗		∗		−		2.44×10^{-4}	B

4.4. Performance Assessment of Selected Subset

In this section, we evaluate UIR for all able-bodied and transfemoral amputee subjects with feature subset p_9. All classifiers are trained with three representative methods (SVM-RBF, SVM-linear, and QDA). The RBF kernel tuning parameter is $\sigma = 1$ and $\sigma = 4$ for able-bodied and amputee subjects, respectively. In this section, we use multiple-fold CV to train and test UIR, where each walking sequence that consists of different gait modes is considered a fold (see Section 2.1). In training phase, we use all walking sequences except one to train UIR. We then test the UIR accuracy on the excluded walking sequence (fold). Accuracy is defined as the total number of correctly classified test patterns divided by the total number of test patterns. We repeat training and testing by shifting the excluded folds. We calculate the accuracy averaged over all folds to find the mean performance of UIR.

We use average classification error of *c*-fold cross validation (CV). In *c*-fold CV, we randomly divide the training set into *c* distinct folds. Then, we repeat training *c* times; each time, the model is trained using $c - 1$ folds and is tested with the remaining fold. The average of the *c* classification errors is used as the quality measure.

We saw in Section 2.2 that overlapped windowing with frame length $L_f = 250$ ms and increment $I = 50$ ms is the best data window option. For real-time operation, a conservative choice for parameter $q = 5$ satisfies the constraint $q \times I \leq 300$ ms. Therefore, we use a majority voting filter (MVF) with length $2 \times q + 1 = 11$. Results verify a fast processing time on a standard desktop computer of less than 50 ms, on average, including feature extraction and classification with each of the three classifiers.

Figure 10 illustrates the mean classification error of QDA, SVM-linear, and SVM-RBF trained with feature subset p_9. Training was conducted individually for able-bodied subjects AB01, AB02, and AB03 and amputee subjects AM01, AM02, and AM03, with and without MVF. Figure 10 indicates that: (1) SVM-RBF outperforms SVM-linear and QDA, which confirms the statistical results in Section 4.3; (2) MVF statistically significantly decreases classification error for locomotion mode detection ($p < 0.05$); and (3) p_9 is an effective feature subset and results in accurate as well as compact UIR. Feature subset p_9 uses only 14 features out of a total of 60 available features, which reduces the size of the feature set by 77%.

SVM-RBF was also trained for the able-bodied subjects with the full set of 60 features. When combined with MVF, it results in a mean classification accuracy of 98.54% ± 1.92%. In comparison, we achieve 97.14% ± 1.51% mean classification accuracy with feature subset p_9, which includes only 14 features. Statistical tests at 5% significance level indicate no significant difference between UIR performance when trained with the full feature set and subset p_9.

SVM-RBF was also trained for the amputee subjects with the full set of 60 features. When combined with MVF, it results in a mean classification accuracy of 99.37% ± 0.96%. In comparison, we achieve 98.45% ± 1.22% mean classification accuracy with feature subset p_9. As with the able-bodied subjects, statistical tests indicate no significant difference between UIR performance when trained with the full feature set and subset p_9. This indicates the satisfactory performance of our framework, which is able to eliminate unneeded features with no significant degradation in overall accuracy.

In this paper, we decoupled the optimization problem of window length and feature selection by dividing it into two smaller sequential optimization problems [35]. We may obtain a suboptimal feature subset with this approach, but this point is not critical since we were able to find accurate and simple UIR that has no meaningful performance difference than UIR designed with the full feature set.

(**a**) Trained with subset p_9 for able-bodied subjects   (**b**) Trained with subset p_9 for amputee subjects
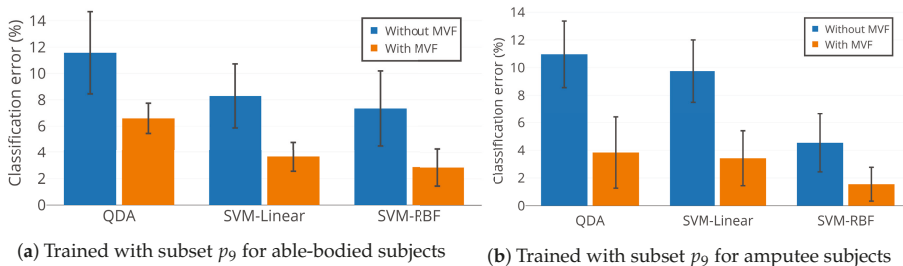
Figure 10. Classification performance of QDA, SVM-Linear, and SVM-RBF with feature subset p_9 for able-bodied subjects (AB01, AB02, and AB03) and amputee subjects (AM01, AM02, and AM03).

5. Conclusions

We presented a framework for designing a UIR system. We used experimental data collected from three able-bodied subjects and three above-knee amputee subjects to classify four and three different gait modes, respectively. Overlapped windowing with frame length 250 ms and increment 50 ms provided a good trade-off between classification performance and real-time computation. Several efficient TD and FD features were extracted from data frames to form the feature set. We performed feature selection in two steps. First, we excluded non-informative features with poor classification performance and high computational effort. Second, we used MOO to find an optimal feature subset from the remaining features to obtain a UIR system that was both parsimonious and accurate. For this purpose, GMOFS, a novel embedded multi-objective feature selection algorithm, was proposed and compared with four evolutionary MOOs on the basis of normalized hypervolume and relative coverage. Classification results confirmed the competitive performance of GMOFS. Several classifiers were trained with the optimal feature subsets that were selected by MOO, and SVM-RBF and MLP were found to be the best classifiers for UIR. The outputs of the classifiers were input to an MVF to improve classification accuracy and chattering between the identified classes.

For future work, more above-knee amputee subjects will be involved in data collection and classification. In addition, we will include other daily-life activities such as incline walking, stair ascent and descent, standing and sitting, etc. It is also of great interest to consider other informative features for classification, such as wavelet transform coefficients. Finally, it would be of interest to compare GMOFS with other state-of-the-art MOO methods, and to apply GMOFS to other MOO problems.

Abbreviations

The following abbreviations are used in this manuscript.

List of acronyms in order of appearance

Acronym	Definition	Acronym	Definition
UIR	User intent recognition	ZC	Zero crossing
MOO	Multi-objective optimization	WL	Waveform length
GMOFS	Gradient-based multi-objective feature selection	VAR	Variance
MOBBO	Multi-objective biogeography-based optimization	MAV	Mean absolute value
SVM	Support vector machine	RMS	Root mean square
RBF	Radial basis function	WAMP	Willison amplitude
MVF	Majority voting filter	SK	Skewness
sEMG	Surface electromyography	KU	Kurtosis
LDA	Linear discriminant analysis	COR	Correlation
QDA	Quadratic discriminant analysis	ANG	Angle
GMM	Gaussian mixture model	PSD	Periodogram spectrum density
ANN	Artificial neural network	MNF	Mean frequency
BBO	Biogeography-based optimization	MDF	Median frequency
VEBBO	Vector evaluated BBO	MAXF	Maximum frequency
NSBBO	Non-dominated sorting BBO	AR	Auto-regressive model
NPBBO	Niched Pareto BBO	CV	Cross validation
SPBBO	Strength Pareto BBO	AB01	Able-bodied subject 01
EA	Evolutionary algorithm	AM01	Amputee subject 01
MLP	Multilayer perceptron	PS	Preferred speed
TD	Time domain	ST	Standing
FD	Frequency domain	NW	Normal walking
FLDA	Fisher's linear discriminant analysis	SW	Slow walking
PCA	Principal component analysis	FW	Fast walking
DT	Decision tree		
SSC	Slope sign change		

References

1. Lawson, B.E.; Varol, H.A.; Huff, A.; Erdemir, E.; Goldfarb, M. Control of stair ascent and descent with a powered transfemoral prosthesis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 466–473. [CrossRef] [PubMed]

2. Khademi, G.; Richter, H.; Simon, D. Multi-objective optimization of tracking/impedance control for a prosthetic leg with energy regeneration. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 5322–5327.

3. Sup, F.; Varol, H.A.; Mitchell, J.; Withrow, T.J.; Goldfarb, M. Preliminary evaluations of a self-contained anthropomorphic transfemoral prosthesis. *IEEE/ASME Trans. Mechatron.* **2009**, *14*, 667–676. [CrossRef] [PubMed]

4. Tucker, M.R.; Olivier, J.; Pagel, A.; Bleuler, H.; Bouri, M.; Lambercy, O.; del R Millán, J.; Riener, R.; Vallery, H.; Gassert, R. Control strategies for active lower extremity prosthetics and orthotics: A review. *J. Neuroeng. Rehabil.* **2015**, *12*, 1. [CrossRef] [PubMed]

5. Huang, H.; Kuiken, T.A.; Lipschutz, R.D. A strategy for identifying locomotion modes using surface electromyography. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 65–73. [CrossRef] [PubMed]

6. Zhang, F.; Huang, H. Real-time recognition of user intent for neural control of artificial legs. In Proceedings of the 2011 MyoElectric Controls/Powered Prosthetics Symposium Fredericton, Fredericton, NB, Canada, 14–19 August 2011.
7. Stolyarov, R.; Burnett, G.; Herr, H. Translational motion tracking of leg joints for enhanced prediction of walking tasks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 763–769. [CrossRef] [PubMed]
8. Hargrove, L.; Englehart, K.; Hudgins, B. The effect of electrode displacements on pattern recognition based myoelectric control. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 2203–2206.
9. Winkel, J.; Jorgensen, K. Significance of skin temperature changes in surface electromyography. *Eur. J. Appl. Physiol. Occup. Physiol.* **1991**, *63*, 345–348. [CrossRef] [PubMed]
10. Zachariah, S.G.; Saxena, R.; Fergason, J.R.; Sanders, J.E. Shape and volume change in the transtibial residuum over the short term: Preliminary investigation of six subjects. *J. Rehabil. Res. Dev.* **2004**, *41*, 683. [CrossRef] [PubMed]
11. Varol, H.A.; Sup, F.; Goldfarb, M. Multiclass real-time intent recognition of a powered lower limb prosthesis. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 542–551. [CrossRef] [PubMed]
12. Liu, M.; Wang, D.; Huang, H.H. Development of an environment-aware locomotion mode recognition system for powered lower limb prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 434–443. [CrossRef] [PubMed]
13. Wang, X.; Wang, Q.; Zheng, E.; Wei, K.; Wang, L. A wearable plantar pressure measurement system: Design specifications and first experiments with an amputee. *Intell. Autom. Syst.* **2013**, *194*, 273–281.
14. Huang, H.; Zhang, F.; Hargrove, L.J.; Dou, Z.; Rogers, D.R.; Englehart, K.B. Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2867–2875. [CrossRef] [PubMed]
15. Young, A.J.; Simon, A.M.; Fey, N.P.; Hargrove, L.J. Classifying the intent of novel users during human locomotion using powered lower limb prostheses. In Proceedings of the IEEE International Conference on Neural Engineering, San Diego, CA, USA, 6–8 November 2013; pp. 311–314.
16. Ha, K.H.; Varol, H.A.; Goldfarb, M. Volitional control of a prosthetic knee using surface electromyography. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 144–151. [CrossRef]
17. Young, A.J.; Hargrove, L.J. A Classification Method for User-Independent Intent Recognition for Transfemoral Amputees Using Powered Lower Limb Prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 217–225. [CrossRef] [PubMed]
18. Young, A.; Kuiken, T.; Hargrove, L. Analysis of using EMG and mechanical sensors to enhance intent recognition in powered lower limb prostheses. *J. Neural Eng.* **2014**, *11*, 056021. [CrossRef] [PubMed]
19. Chen, B.; Zheng, E.; Wang, Q.; Wang, L. A new strategy for parameter optimization to improve phase-dependent locomotion mode recognition. *Neurocomputing* **2015**, *149*, 585–593. [CrossRef]
20. Simon, D.; Omran, M.G.; Clerc, M. Linearized biogeography-based optimization with re-initialization and local search. *Inf. Sci.* **2014**, *267*, 140–157. [CrossRef]
21. Simon, D. *Evolutionary Optimization Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
22. Xue, B.; Zhang, M.; Browne, W.N. Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Trans. Cybern.* **2013**, *43*, 1656–1671. [CrossRef]
23. Ghaemi, M.; Feizi-Derakhshi, M.R. Feature selection using forest optimization algorithm. *Pattern Recognit.* **2016**, *60*, 121–129. [CrossRef]
24. Hall, M.A.; Smith, L.A. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. In *FLAIRS Conference*; AAAI: Menlo Park, CA, USA, 1999; pp. 235–239.
25. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]
26. Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [CrossRef]
27. Ma, S.; Huang, J. Penalized feature selection and classification in bioinformatics. *Brief. Bioinform.* **2008**, *9*, 392–403. [CrossRef] [PubMed]
28. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
29. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]

30. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **1996**, *58*, 267–288. [CrossRef]
31. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc.* **2005**, *67*, 301–320. [CrossRef]
32. Van den Bogert, A.J.; Geijtenbeek, T.; Even-Zohar, O.; Steenbrink, F.; Hardin, E.C. A real-time system for biomechanical analysis of human movement and muscle function. *Med. Biol. Eng. Comput.* **2013**, *51*, 1069–1077. [CrossRef]
33. *Healthcare Inspection Prosthetic Limb Care in VA Facilities*; Technical Report; Department of Veterans Affairs Office of Inspector General: Washington, DC, USA, 2011.
34. Lara, O.D.; Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [CrossRef]
35. Oskoei, M.A.; Hu, H. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1956–1965. [CrossRef]
36. Hudgins, B.; Parker, P.; Scott, R.N. A new strategy for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* **1993**, *40*, 82–94. [CrossRef]
37. Valls-Solé, J.; Rothwell, J.C.; Goulart, F.; Cossu, G.; Munoz, E. Patterned ballistic movements triggered by a startle in healthy humans. *J. Physiol.* **1999**, *516*, 931–938. [CrossRef]
38. Stark, L. *Neurological Control Systems: Studies in Bioengineering*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
39. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. A novel feature extraction for robust EMG pattern recognition. *arXiv* **2009**, arXiv:0912.3973.
40. Oskoei, M.A.; Hu, H. Myoelectric control systems—A survey. *Biomed. Signal Process. Control* **2007**, *2*, 275–294. [CrossRef]
41. Englehart, K.; Hudgins, B. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 848–854. [CrossRef] [PubMed]
42. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
43. Oskoei, M.A.; Hu, H. GA-based feature subset selection for myoelectric classification. In Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics, Kunming, China , 17–20 December 2006; pp. 1465–1470.
44. Xue, B.; Zhang, M.; Browne, W.N.; Yao, X. A survey on evolutionary computation approaches to feature selection. *IEEE Trans. Evol. Comput.* **2016**, *20*, 606–626. [CrossRef]
45. Bishop, C.M. *Neural Networks for Pattern Recognition*; Oxford University Press; Oxford, UK, 1995.
46. Moore, B. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Autom. Control* **1981**, *26*, 17–32. [CrossRef]
47. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 2002.

Weight-Bearing Estimation for Cane Users by Using Onboard Sensors

Joaquin Ballesteros [1,*,†]**, Alberto Tudela** [2,†]**, Juan Rafael Caro-Romero** [2,†] **and Cristina Urdiales** [2,†]

[1]   Division of Intelligent Future Technologies, Mälardalen University, 721 23 Västerås, Sweden
[2]   Department of Electronic Technology, University of Malaga, 29071 Malaga, Spain;
      ajtudela@uma.es (A.T.); jrcaro@uma.es (J.R.C.-R.); acurdiales@uma.es (C.U.)
*   Correspondence: joaquin.ballesteros@mdh.se or jballesteros@uma.es
†   These authors contributed equally to this work.

Abstract: Mobility is a fundamental requirement for a healthy, active lifestyle. Gait analysis is widely acknowledged as a clinically useful tool for identifying problems with mobility, as identifying abnormalities within the gait profile is essential to correct them via training, drugs, or surgical intervention. However, continuous gait analysis is difficult to achieve due to technical limitations, namely the need for specific hardware and constraints on time and test environment to acquire reliable data. Wearables may provide a solution if users carry them most of the time they are walking. We propose to add sensors to walking canes to assess user's mobility. Canes are frequently used by people who cannot completely support their own weight due to pain or balance issues. Furthermore, in absence of neurological disorders, the load on the cane is correlated with the user condition. Sensorized canes already exist, but often rely on expensive sensors and major device modifications are required. Thus, the number of potential users is severely limited. In this work, we propose an affordable module for load monitoring so that it can be widely used as a screening tool. The main advantages of our module are: (i) it can be deployed in any standard cane with minimal changes that do not affect ergonomics; (ii) it can be used every day, anywhere for long-term monitoring. We have validated our prototype with 10 different elderly volunteers that required a cane to walk, either for balance or partial weight bearing. Volunteers were asked to complete a 10 m test and, then, to move freely for an extra minute. The load peaks on the cane, corresponding to maximum support instants during the gait cycle, were measured while they moved. For validation, we calculated their gait speed using a chronometer during the 10 m test, as it is reportedly related to their condition. The correlation between speed (condition) and load results proves that our module provides meaningful information for screening. In conclusion, our module monitors support in a continuous, unsupervised, nonintrusive way during users' daily routines, plus only mechanical adjustment (cane height) is needed to change from one user to another.

Keywords: smart cane; weight-bearing; gait analysis; health monitoring

1. Introduction

A 2017 United Nations report on world population ageing [1] shows that the number of persons over 60 years in 2050 will double with respect to 2017, i.e., in 2050 one out of five people worldwide will be seniors. Indeed, population ageing is already significant in Europe and Northern America, where currently more than one out of five persons are already over 60. Hence, healthy ageing has become the main concern. One of the main tools to promote healthy ageing is, reportedly, monitoring and feedback to users [2], especially for the most vulnerable population, like persons with some degree of disability. Mobility monitoring has attracted major interest, as it is fundamental to keep a healthy and active lifestyle and to remain autonomous. The simpler, most popular approach to

mobility monitoring are activity/fitness trackers, that provide data on distance walked or run, calorie consumption, and, in some cases heartbeat or quality of sleep. However, parameters like distance walked are only indirectly related to a person's condition. Alternatively, gait (and posture) analysis is widely acknowledged as a clinically useful tool for identifying problems with mobility. Poor gait may lead to musculoskeletal pain in different parts of the body, whereas a good gait will reduce the risk of joint problems, help good recovery after injury and/or surgery, improve mobility in the elderly and reduce fall risk [3]. Consequently, gait monitoring has attracted considerable attention.

Clinically, gait analysis is performed via scales like the Tinetti Mobility Test [4]: experts observe patients performing different tasks and manually provide scores for several indicators. However, this process is slow and requires significant time from patient and clinician. There are technology-based alternatives to this approach, like using force/pressure sensors on a walking surface like a treadmill or a walkway [5,6] or optical motion capture systems [7–12]. However, people need to be assessed in specific installations, and experts are often needed to set and/or calibrate equipment. Gait assessment during Activities of Daily Living (ADLs) may be achieved using wearable sensors, e.g., sole pressure sensors [13,14] and/or inertial sensors [15–17]. However, some of these wearables may not be comfortable nor easy to attach/calibrate [18] so people often do not carry them. Sensors can instead be attached to mobility aids if people require them for their ADLs [19,20]. Indeed, smart wheelchairs and rollators often include different sensors. However, attaching them to canes, the most spread mobility aid, with minor alteration of their ergonomics (Modifications that affect ergonomics or center of gravity in walking aids require full analysis, validation, and standardization of the modified device, a process usually costly and slow) is usually harder.

Load on a cane can be used to estimate a person's partial weight-bearing by calculating the differences between the users' weight and the load. Partial weight-bearing provides information about how much load the user supports on their affected side. This is an indirect measure of their condition and it can be used to monitor their evolution, provide feedback and, in clinical treatments, to adapt their rehabilitation process [21]. There are different solutions to measure load on a cane. In specific, fixed environments, a walkway with a dense matrix of force sensors can be used, e.g., Strideway^{TM} (Strideway System; Tekscan, Inc., South Boston, MA, USA) (Strideway System. Available online at: https://www.tekscan.com/products-solutions/systems/strideway-system?tab=applications). These systems measure weight-bearing accurately and, hence, the cane load by default. However, they only measure a very limited number of meters on a straight line, plus they are expensive and, as aforementioned, constrained to specific environments. Hence, this work focuses uniquely on onboard cane sensors. Currently, onboard sensors on most smart canes are located either on the handgrip [22,23], shaft [23–27] or tip [22,24]. Every location has advantages and disadvantages. Placing sensors on the handgrip or tip may involve major cane modifications [22,23]. As both locations affect how users support their weight, these modifications must be ergonomic. Unfortunately, granting ergonomics requires an extensively validation process. On the other hand, the shaft allows more space to place the electronic. However, this approach may involve changes in the cane center of gravity and, in its weight [23–25,27].

This work proposes an affordable add-on module for long-term monitoring of load on cane. The obtained value can be used for long-term assessment of users' condition, either for preventive healthcare or for evaluation of degenerative or rehabilitation processes. The module has been designed to be extensively available to as many potential users as possible, so its main goals are: (i) compatibility with existing commercial canes; (ii) to avoid any impact on cane ergonomics; (iii) to allow continuous, long-term use; and (iv) to keep the global cost as low as possible. Additionally, we plan to release the proposed system under a Creative Commons License to boost its reach and impact. Section 2 describes the proposed system, i.e., mechanical design and electronics. The load cane estimation is described in Section 3. Section 4 describes our methodology. Tests to validate that the module provides meaningful load estimation are presented in Section 5. Section 6 discusses our results. Finally, Section 7 presents the conclusions and future work.

2. Cane Adaptation

The main goal of this work is to design and develop a low-cost module that can be attached to commercial canes for load monitoring in a (fairly) simple way. Hence, we have the following specific subgoals: (i) use low-cost components; (ii) preserve cane ergonomics; and (iii) low power consumption. This is achieved through goal-oriented mechanical and electronic design.

2.1. Mechanical Design

Our module includes three different areas: sensors area, microcontroller area and charger area (Figure 1a). The sensors have been embedded inside the shaft, as close to the tip as possible at two different depths: 0.2 mm and 1 mm via a specially designed 3D printed plastic piece (Figure 1b). Thus, the tip and handgrip designs are not modified, and we preserve their original ergonomic properties. Placing electronics inside the shaft is actually much harder and makes access difficult, so they are externally attached. Their location has been split into two different areas to minimize changes to the center of gravity of the cane: the microcontroller (MCU) area (Figure 1c) and the charging area (Figure 1d). To connect the microcontroller (MCU) with the charger and with the sensor areas, only two holes are required in the shaft (H1 and H2 in Figure 1c,d). Hence, attaching the module only involves minor cane modifications.

The charging area has been designed to avoid connectors, so elderly people with visual impairment can use it in a straight, simple way. It relies on a *Qi* interface and it is divided into two different parts. The first part is a fixed piece that can be attached to a wall (**J** in Figure 1d). The second one is a flat piece attached to the shaft cane (**I** in Figure 1d). The *Qi* interface uses inductive charging over close distances. Hence, users do not need to plug any connector; they just need to support the cane on the wall piece and the battery starts charging. Each piece has been modelled using Autodesk® Fusion 360 and it has been 3D printed with Ultimaker2 using PLA plastic (All 3D models are freely available online at https://github.com/joaquinballesteros/Smart-Cane).

(**a**) General View   (**b**) Sensors Area   (**c**) MCU area   (**d**) Charger area
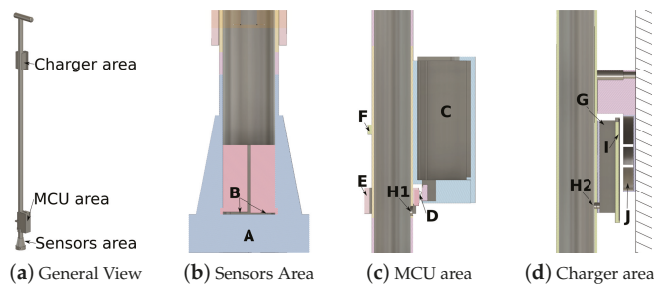
Figure 1. Mechanical design overview (**a**) and its parts (**b**–**d**): (A) rubber tip; (B) force sensors locations; (C) microcontroller (MCU) box; (D) connectors between charger-sensors and MCU; (E) connector ring; (F) clamp for MCU box; (G) battery and charger box; (H1) hole to wires from charger and sensors areas to MCU box; (H2) hole to wires from MCU box to charger area; (I) charging receiver location; and (J) charging transmitter location.

2.2. Electronic Design

There are different approaches to measure load on a cane. The most usual ones rely on force sensors on handgrip [22,23] and/or force sensors replacing part of the shaft [27]. Some of these solutions may affect the center of gravity and/or ergonomics of the cane depending on the sensor location, as commented. Nevertheless, the main drawback of most existing approaches is that they rely on relatively expensive electronic components, such as piezoelectric quartz force link with in-line amplifier (Kistler Instrument Corp., Novi, MI, USA) [27] or array of force sensors (FSR 402, Interlink Electronics, Los Angeles, CA, USA) [22], reducing affordability for end-users. Additionally,

solutions do not deal with the charging process, which can be a challenge for users with vision and/or cognitive impairments.

The main problem with force sensors is that their price increases significantly with their measurement range. Our proposed solution relies on combining two cheap force sensors at different depths into a single piece (**B** in Figure 1b) to increase their original range (see next section). We have designed a specific plastic piece to arrange them perfectly inside the shaft using a 3D printer. This approach is affordable because it only involves: two low-cost sensors (FSR 402, Interlink Electronics, Los Angeles, CA, USA), an inverter (TC7662B, Microchip Technology, Chandler, AZ, USA), an array of operational amplifiers (OPA347, Texas Instruments, Dallas, TX, USA) and some resistors and capacitors. Figure 2 shows the sensor board. The input signals to FS402 are inverted $-V_{ref}$ using TC7662B circuit. The operational amplifiers U1, U2, and U3 are adjusted with a 2.4 kΩ and 100 Ω resistors respectively to obtain the higher range of support measurement. The filtered output of FSR 402 force sensors U1, U2 are added to obtain the combination of both sensors U3. The sensor board provides a 50 Hz output.
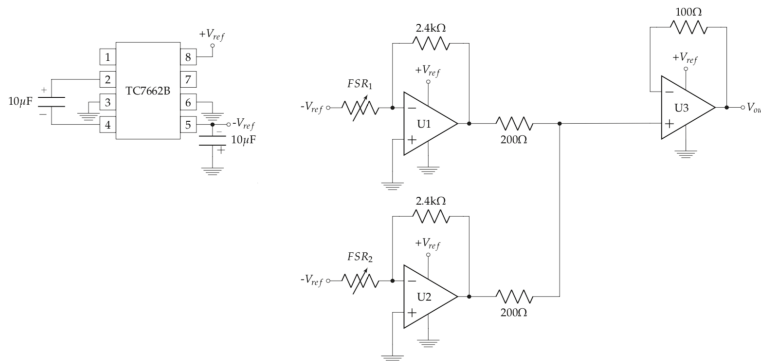
Figure 2. Sensor electronic board. The input signals to FS402 sensors are inverted. *U*1, *U*2 filter the force sensors outputs. Then, *U*3 sums both signal to obtain the output.

We work with a BLE nano v2 microcontroller (nrf51822, Nordic Semiconductor, Norway). It transmits packets of 20 Bytes (8 readings) at 6.25 Hz to a paired device. Long-term monitoring requires a working plan for at least 12 h without recharging. For this reason, we have chosen a standard Qi Wireless Charger receiver with a Lipo charger (TP4056, Nanjing Top Power ASIC Corp, Nankin, China) connected to a 1S1P 500 mAh Lipo battery that reportedly provides over 4 days of use without recharging.

To heuristically calculate the system battery working life, we have performed 10 runs of the same test, consisting of feeding the cane non-stop until communication with the microcontroller is cut off, i.e., a paired device does not receive packets any longer. The battery working life is inversely proportional to the load, i.e., higher loads will lead to shorter battery life. Hence, we have used a static weight on the cane to simulate the user's load. The period of loading versus non-loading and the amount of loaded weight have been mechanically changed to simulate a large variety of users' behaviors. After several tests, the range battery working life turned out to be 82–89 h, i.e., more than 3 days of continuous monitoring.

Overall, the cost of the proposed system is cheap when compared to other solutions (less than USD 100 in total), plus it requires only minor modifications in the cane.

3. Dynamic Weight-Bearing Calculation

Canes can be used to reduce pain in one lower extremity by allowing users to support some weight on them. Load on cane depends on the person's own weight, but also on their condition. To design a cane sensor, it is necessary to grant that its range is wide enough to capture this weight

variations. As commented, low-cost sensors have lower ranges and accuracies. However, in this section we propose a method to increase their range by combining two cheap sensors into a single sensing module.

Cane users can be categorized into contralateral and ipsilateral. Users who support weight on the cane and on the closest foot to the cane at the same time are ipsilateral. The others are contralateral. Ipsilateral users reportedly load up to 7% of their body weight on the cane. Contralateral users are more frequent and more critical, as they load up to 9% [28]. These percentages may increase when users have some physical issues. Our solution relies on cheap FSR 402 force sensors to measure loads. FSR 402 sensors have a limited range up to only 10 kg. Therefore, contralateral users could weigh as much as 111 kg, whose 9% is 10 kg, before sensors saturate. Unfortunately, this limit could be significantly more restricting for users with physical issues.

To increase the measurement range while keeping sensors affordable, we have placed two (Due to the limit in the inner cane diameter (maximum of 22 mm), only two force sensors can be placed) FSR 402 sensors at different depths on a circular 3D printed plastic piece. First, weight is distributed all over the surface of the piece (Figure 3a), so each sensor only receives part of the load, effectively increasing the global piece range. Additionally, the tip rubber applies different pressure on each sensor (Figure 3b), i.e., lower loads will not affect the deepest sensor. Unfortunately, factors like the nature of materials, non-linearity in sensors, and other physical variables, make unfordable to analytically calculate the sensor area output function. However, it is possible to estimate it. To obtain an estimated function of the sensors' outputs, we have calibrated the system through extensive testing using depths ranging from 0 mm to 5 mm.
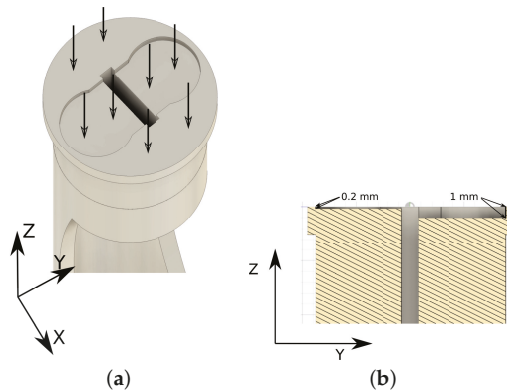
Figure 3. Pressure distribution is done at two different levels: (**a**) First the rubber tip pressure distribution among the force sensors and the sensor area; (**b**) Second, pressure distribution on the sensors depends on the depth.

Calibration consisted of automatically applying different static weights on the fully vertical cane when one FSR 402 was connected to the sensor electronic board (Figure 2) and the other was bypassed. The readings were performed using one Analog Digital Converter (ADC), range 0–1023, from the BLE nano v2 board. First, we tested different sensor depths to select the most appropriate one, i.e., depths that provide the best reading range. Figure 4 shows different depths where FSR 402 were located and, it shows the average load (kg) in those locations. It can be observed that deepest locations (1 mm, 1.5 mm or 2 mm) have a problem to detect lower loads as we commented previously. On the other hand, outer locations saturate the signal when higher load values are applied (0.15 mm, 0.2 mm or 0.3 mm). For these reasons, we selected 0.2 mm and 1 mm as a combination that deals successfully with low and high load on cane simultaneously.
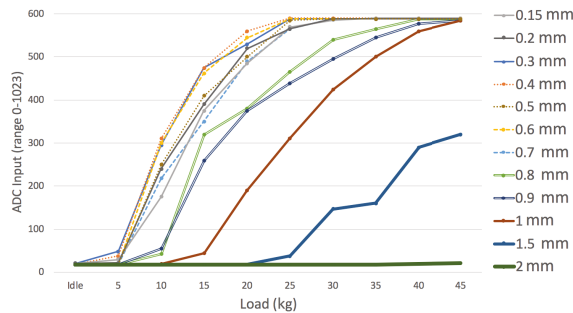
Figure 4. Sensor calibration for different sensor depths. Different static loads have been measured using the sensor electronic board (Figure 2) with one FSR 402 connected (the other one was bypassed).

Once both locations were selected (0.2 mm and 1 mm), we applied the same procedure to find the relation between applied weights and corresponding sensors readings, now with both FSR 402 connected. Figure 5 shows the obtained relation. It can be observed that readings below 2.5 kg present a bigger mean squared error (0.1389) than measurements over this value (0.0323). This is an effect of the rubber tip pressure distribution, as low weights only affect—partially—the outermost sensor. The obtained curve can be approximated by the following quadratic equation: $8.74 \cdot 10^{-11}x^4 - 1.21 \cdot 10^{-7}x^3 + 2.11 \cdot 10^{-5}x^2 + 3.61 \cdot 10^{-2}x + 0.5191$.
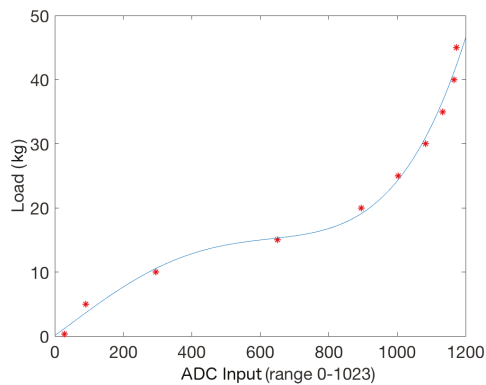
Figure 5. This figure shows the relation between the ADC connected to the sensor electronic board (Figure 2) and the load on the cane in kilograms when only one sensor is connected (the other one is bypassed). A 4th order polynomial has been used to estimate that transformation: $8.74 \cdot 10^{-11}x^4 - 1.21 \cdot 10^{-7}x^3 + 2.11 \cdot 10^{-5}x^2 + 3.61 \cdot 10^{-2}x + 0.5191$. The mean squared error on the estimation is 0.0515 kg.

The designed 2-sensors piece output range grows up to 45 kg (hardware reading equal to 1023) for the best depth difference. This means that a healthy contralateral user could weigh as much as 500 kg. Even though persons with disabilities support significantly more than 9% on the cane, this upper bound is high enough for most cases.

4. Methodology

While users walk, the weight they bear on the cane keeps changing. To calculate the maximum support that a given user needs while walking, gait cycles can be analyzed using the adapted cane. Figure 6a shows the typical cane movements for contralateral users while they walk. During a gait cycle, maximum force is applied on the cane when it is fully vertical, as the load vector is orthogonal to the force sensors plane. The cane support cycle corresponds to the elapsed time from a heel strike

of the opposite leg, with respect to the cane, to the next one. As expected, force sensors outputs fluctuate while users change their load on the cane (Figure 6b). As commented, maximum peak values correspond to the cane in a vertical position. A peak during a cane support period represents the upper bound support that a given user needs in each step. A sequence of peaks in time provides continuous information about the user load on the cane. We have used the *findpeaks* MATLAB R2016b (The MathWorks, Inc., Natick, MA, USA) function to detect all peaks during a given test. Users load on the cane during the affected leg step, but then they move the cane. The minimum time between two cane supports depends on the users' step times. More than 99,99% of the elderly population reported a value above 0.5 s [29]. For these reasons, the *findpeaks* function has been set with a *MinPeakDistance* parameter equal to 25 (0.5 s). An additional function parameter, the *MinPeakHeight*, has been defined. It is used to filter out spurious peaks from the input signal. Its value depends on users' load, so we have empirically set it to the average of the input signal exercise.
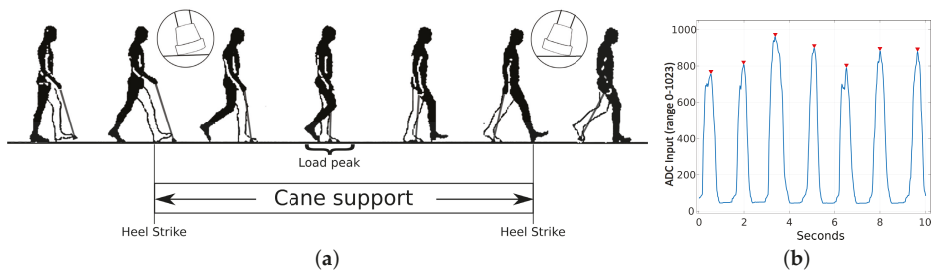
Figure 6. Contralateral user walking: (**a**) User walking in the support cane period; and (**b**) Input signal from force sensor during a 10 s interval in user 8 (range is 0–1023). Maximum peaks (corresponding to cane vertical position) are marked in ▼.

Our tests were carried out in two senior centers in Cordoba, Spain. As commented, we chose volunteers who require a cane for mobility. We chose to select people with different disabilities that required the cane either for balance or to reduce pain in limbs after surgery or due to degenerative conditions. We also chose people who favor, left, right, or no particular side when walking. Thus, we can check how the module responds to different load bearing profiles. In the end, we selected 10 volunteers: 8 men and 2 women. An additional participant who was initially eligible was excluded from the study, as the module showed that she did not load any weight while she walked, i.e., did not require weight support assistance. Participants were on average 83.7 years old (range 74–94 years). Table 1 shows their ages, genders, average gait speeds, and their physical diseases.

Table 1. Condition and characteristics per users.

Id	Age	Gender	Gait Speed	Physical Issues
1	80	M	0.615 m/s ± 0.0096 m/s	Visual impairment; osteoarthritis; low back pain
2	78	F	0.654 m/s ± 0.0109 m/s	Osteoarthritis (right shoulder and leg); spinal discs herniation
3	85	F	0.607 m/s ± 0.0094 m/s	Meniscus surgery in both knees
4	87	M	0.498 m/s ± 0.0063 m/s	Osteoarthritis (left knee)
5	86	M	0.687 m/s ± 0.0120 m/s	Heart surgery. Lower limbs weakness
6	91	M	0.597 m/s ± 0.0090 m/s	Vestibular disorder
7	76	M	0.763 m/s ± 0.0148 m/s	Visual impairment; low back pain
8	74	M	0.792 m/s ± 0.0160 m/s	Right knee prosthesis
9	86	M	0.749 m/s ± 0.0143 m/s	Osteoarthritis; low back pain; left meniscus surgery
10	94	M	0.563 m/s ± 0.0080 m/s	Anemia; pacemaker; left knee pain

Volunteers were asked to complete two different popular clinical tests sequentially. All cane data was collected by a mobile phone paired with the cane via BT. First, volunteers performed the 10 m test (Figure 7). Then, they kept walking during 1 more minute approximately, depending on their

condition. All tests were approved by the University of Malaga Institutional Ethical Committee and by the senior centers. Also, all volunteers signed an informed consent.

Figure 7. Cane user walking during a free walking test.

5. Experiments and Results

The target of the present work is to develop a low-cost, long-term module to estimate load on a cane while users walk. Hence, to validate the system, it is necessary to check that the readings of the sensor depend on the load on the cane during the gait cycle. It is important to test the device with its target population, as healthy volunteers reportedly use walking aids in different, non-consistent ways even when they try to simulate pathological gaits. Our test population includes only people who rely on a traditional cane for mobility in their ADLs (Figure 7).

We had already tested that the cane reliably measures static weights during the calibration process (Figure 4). The error was bigger when the load was below 2.5 kg (mean squared error of 0.1389). However, it is necessary to check how the module behaves during a gait cycle, with dynamically changing loads. As commented, the load on the cane depends not only on user weight but also on their condition. Hence, it is interesting to test the system with people with different conditions to check that the module works well for different gait abnormalities. Unfortunately, although weight can be normalized or accounted for, major load variations due to the condition cannot be quantified, i.e., we cannot directly predict how much weight a given person is going to support on the cane at each time instant to check whether our module is accurate or not. To correlate our readings with the condition, we assess users via their gait speed, which has been consistently reported as a meaningful parameter. We have measured manually the gait speed using two markers in the floor and a highly precise chronometer.

Figure 8 shows the cane load over time for each user presented in Table 1. We can observe that some users load significantly more weight on the cane with respect to others. For example, users 1 and 4 present peak values of 8.55 kg and 11.78 kg on average when compared to users like 2 or 5 (0.18 kg and 0.32 kg, respectively). The main reason for this variability is that, as commented, load depends largely on the users' condition, even more than on the users' weight, i.e., users with poor condition need more assistance. We can also notice that some users increase the load on the cane the longer they walk, like user 3 (meniscus surgery in both knees).
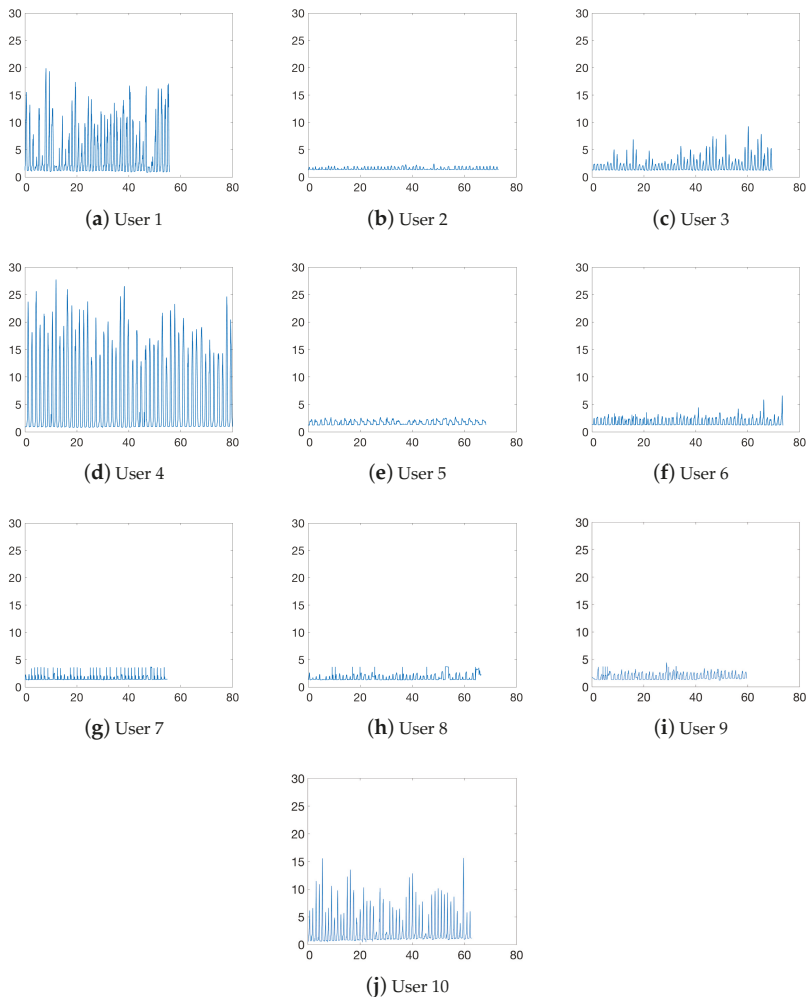
Figure 8. Load on cane *y*-axis (kg) over time *x*-axis (seconds). Users were suggested to walk for one minute, some of them (users 1 and 7) walked less than the minute and others more than the minute.

As weight bearing depends largely on condition, we had no benchmark function to determine how much weight each volunteer loaded on the cane at each time instant. Hence, we used gait speed as an indirect measure of disability to check that users with poorer conditions returned larger loads on the cane. People with severe dependencies typically present gait speeds below 0.6 m/s [30] and they are expected to bear more weight on the cane. Figure 9 shows the relationship between gait speed during our 10 m tests and the load peaks for each of our volunteers. As expected, gait speeds below 0.615 m/s are related with higher loads on the cane: 7.47 kg on average (ranging from 1.15 kg to 17.13 kg); whereas volunteers with gait speeds above its have an average load on the cane of 0.89 kg (ranging from 0.18 kg to 1.45 kg). Additionally, load variances for gait speeds under 0.615 m/s (variation range from 0.84 kg to 2.69 kg) are higher when compared to gait speeds above this limit (variation range from 0.17 kg to 0.68 kg). This gait speed relation with load on cane confirms that users with poor condition need more assistance than others with lower condition and that the need for

aid depends largely on their condition. Specifically, we obtained the Pearson correlation coefficient between gait speed and median load for our volunteers (H_0 is a correlation equal or greater than 0). The resulting coefficient to -0.7473 (*p*-value 0.0065), meaning that gait speed and load are inversely related, as reported by clinicians [31]. Volunteer 6 is an outlier in this analysis because he presents a vestibular disorder, i.e., he uses the cane for balance rather than for weight bearing. If we remove him from the correlation, the coefficient is equal to -0.7971 (*p*-value 0.0050).
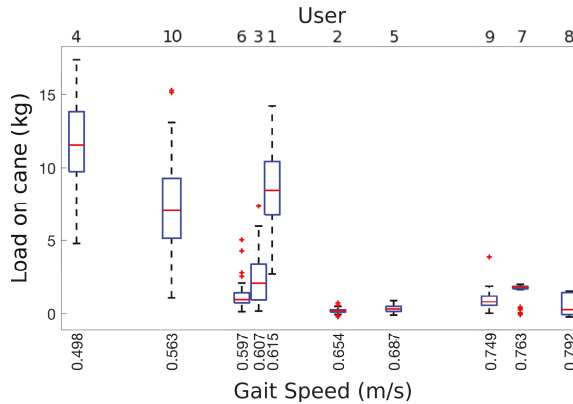
Figure 9. Load on cane vs users' gait speed.

Finally, we can observe in Figure 8 that the upper bound for our test volunteers is equal to 29 kg. This limit fits well the target 3-days of use without charge for 8 h per day of loading. In conclusion, the proposed module meets the required constrains: (i) it is cheap and easy to add to a commercial cane; (ii) it reliably estimates load on the cane in static and dynamic conditions; and (iii) it can be used for long-term monitoring.

6. Discussion

The target of this work was to present a low-cost device that can be attached to commercial canes for long-term load monitoring, so it can be used for condition assessment. The proposed device has been built and tested with its target users. Results have proven that indeed: (i) it provides continuous load monitoring; (ii) it does not affect cane ergonomics; and (iii) battery allows long-term use without recharging. Additionally, results show that the load on the cane is correlated with gait speed, which is a clinically reported condition estimator.

Tests were carried out in two senior facilities in Cordoba, Andalusia. Selected volunteers required a cane for mobility. Thus, we could check load estimation when the system operates under dynamic loads through different pathological gait cycles. We also obtained users' gait speed to check that estimated load is related to speed, i.e., condition. We did not notice any difference in users' gait when they were using the modified cane, nor they reported any change. Hence, we validated that changes in ergonomics were not significant. Also, we did not need to recharge the cane at all for the whole test duration, so we also validated that it can be continuously used for long-term monitoring. Load results were coherent with the hypothesis: people with more significant disabilities consistently bear more weight on their cane and walk slower. The device provides dynamic loads during the gait cycle. We checked that load variation was larger for people in poorer condition and that they increased their load on the cane when they started getting tired.

The proposed system is not as reliable as other load estimation devices, like treadmills or pressure plates. However, those systems are constrained to specific installations, so they do not provide information while users perform their ADLs. Our module has three main advantages when compared to similar cane-based systems. First, it is a low-cost add-on module that can be easily attached to any

commercial cane, rather than a stand-alone system that may present significantly higher costs and lower availability. The module has been designed to avoid changes in ergonomics and/or center of gravity of the original cane. Second, we have designed a novel sensor module that combines two low-cost load sensors at different depths into a single piece to increase their range. Optimal depth differences have been heuristically estimated to provide the best results even under extreme loads (up to 45 kg fully supported on the cane). Third, the system has been extensively tested under different loads to grant at least 3 days of operation without battery recharge, including power required by the communication module. Information can be gathered using any BT-equipped portable device like a smartphone or a tablet. We rely on an induction battery recharge system to facilitate its use to people with visual impairment. The measured error under static load is under 0.14 kg, which is suitable for the average load that our target population bears on their canes.

We expect that a significant number of these modules can be produced at a reduced cost to monitor specific target groups for long-term clinical gait analysis in hospital facilities and/or care centers. Unlike other equipment, the proposed system may be carried by users anywhere, to outside and inside environments, for days, so it could provide meaningful information about behavior trends and condition changes in people with specific disability profiles.

7. Conclusions and Future Work

This paper has presented a low-cost modular system to measure load on a cane for people with disabilities. The main goal of our work was to design an affordable prototype that could be easily attached to commercial canes and used by a wide population for long-term gait analysis. For this purpose, the system has been shared under an open license. All 3D models, plus the microcontroller software are freely available online under a Creative Commons Attribution 4.0.

The system has been tested in two senior facilities in Cordoba, Andalusia. All volunteers were elderly people that required a cane for everyday mobility and presented different disabilities. As load information is related to user's condition, it has significant clinical value. Load analysis provided by our system may help to monitor the evolution of people after surgery, during rehabilitation or suffering degenerative processes to correct their gait, change treatments or prevent accidents like falls.

Future work will focus on two different aspects: improving the module design and extracting more information from gathered data. The module can be improved by increasing battery life and reducing BT pairing time. We plan to induce a low power consumption stage when no loads are detected for a significant period of time. Additionally, the load on peak estimation will be processed and stored on the microcontroller to reduce communication time between the microcontroller and the mobile phone. Our preliminary tests also point out that further clinically relevant information can be extracted from the system, so we will evaluate parameters of interest in gait analysis and check how many we can extract either from the current module or from an improved version. Finally, we plan to perform exhaustive testing by the target population, including usability tests to assess acceptance.

References

1. *World Population Ageing 2017*; Technical Report ST/ESA/SER.A/408; United Nations, Department of Economic and Social Affairs, Population Division: New York, NY, USA, 2017.
2. *Global Strategy and Action Plan on Ageing and Health*; Technical Report 978-92-4-151350-0; World Health Organization: Geneva, Switzerland, 2017.
3. Hodgins, D. The importance of measuring human gait. *Med. Device Technol.* **2008**, *19*, 42–44. [PubMed]
4. Tinetti, M.E. Performance-oriented assessment of mobility problems in elderly patients. *J. Am. Geriatr. Soc.* **1986**, *34*, 119–126. [CrossRef] [PubMed]
5. Zeni, J.; Richards, J.; Higginson, J. Two simple methods for determining gait events during treadmill and overground walking using kinematic data. *Gait Posture* **2008**, *27*, 710–714. [CrossRef] [PubMed]
6. Webster, K.E.; Wittwer, J.E.; Feller, J.A. Validity of the GAITRite® walkway system for the measurement of averaged and individual step parameters of gait. *Gait Posture* **2005**, *22*, 317–321. [CrossRef] [PubMed]
7. Gomatam, A.N.M.; Sasi, S. Multimodal Gait Recognition Based on Stereo Vision and 3D Template Matching. In Proceedings of the International Conference on Information Sciences and Systems (CISST 2004), Princeton, NJ, USA, 17–19 March 2004; pp. 405–410.
8. Banach, M.; Wasilewska, A.; Dlugosz, R.; Pauk, J. Novel techniques for a wireless motion capture system for the monitoring and rehabilitation of disabled persons for application in smart buildings. *Technol. Health Care* **2018**, *26*, 671–677. [CrossRef] [PubMed]
9. Mostayed, A.; Mazumder, M.M.G.; Kim, S.; Park, S.J. Abnormal gait detection using discrete fourier transform. In Proceedings of the International Conference on Multimedia and Ubiquitous Engineering (MUE 2008), Busan, Korea, 24–26 April 2008; pp. 36–40.
10. Xue, Z.; Ming, D.; Song, W.; Wan, B.; Jin, S. Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognit.* **2010**, *43*, 2904–2910. [CrossRef]
11. Derawi, M.O.; Ali, H.; Cheikh, F.A. Gait Recognition using Time-of-Flight Sensor. In Proceedings of the Special Interest Group on Biometrics and Electronic Signatures (BIOSIG 2011), Darmstadt, Germany, 8–9 September 2011; pp. 187–194.
12. Gabel, M.; Gilad-Bachrach, R.; Renshaw, E.; Schuster, A. Full body gait analysis with Kinect. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 1964–1967.
13. Bae, J.; Tomizuka, M. A tele-monitoring system for gait rehabilitation with an inertial measurement unit and a shoe-type ground reaction force sensor. *Mechatronics* **2013**, *23*, 646–651. [CrossRef]
14. Hellstrom, P.; Folke, M.; Ekström, M. Wearable weight estimation system. *Procedia Comput. Sci.* **2015**, *64*, 146–152. [CrossRef]
15. Ferrari, A.; Rocchi, L.; van den Noort, J.; Harlaar, J. Toward the Use of Wearable Inertial Sensors to Train Gait in Subjects with Movement Disorders. In *Converging Clinical and Engineering Research on Neurorehabilitation*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 937–940.
16. Bruijn, S.M.; Warner, R.T.; Kate, T.; Faber, G.S.; Meijer, O.G.; Beek, P.J.; van Dieën, J.H. Estimating dynamic gait stability using data from non-aligned inertial sensors. *Ann. Biomed. Eng.* **2010**, *38*, 2588–2593. [CrossRef] [PubMed]
17. Ostaszewski, M.; Pauk, J. Estimation of ground reaction forces and joint moments on the basis on plantar pressure insoles and wearable sensors for joint angle measurement. *Technol. Health Care* **2018**, *26*, 605–612. [CrossRef]
18. Carcreff, L.; Gerber, C.N.; Paraschiv-Ionescu, A.; De Coulon, G.; Newman, C.J.; Armand, S.; Aminian, K. What is the Best Configuration of Wearable Sensors to Measure Spatiotemporal Gait Parameters in Children with Cerebral Palsy? *Sensors* **2018**, *18*, 394. [CrossRef] [PubMed]
19. Ballesteros, J.; Urdiales, C.; Martinez, A.B.; Tirado, M. Gait analysis for challenged users based on a rollator equipped with force sensors. In Proceedings of the 2015 IEEE/RSJ International Conference onIntelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 5587–5592.
20. Ballesteros, J.; Urdiales, C.; Martinez, A.B.; Tirado, M. Automatic Assessment of a Rollator-User's Condition During Rehabilitation Using the i-Walker Platform. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 2009–2017. [CrossRef] [PubMed]

21. Hesse, S.; Bertelt, C.; Jahnke, M.; Schaffrin, A.; Baake, P.; Malezic, M.; Mauritz, K. Treadmill training with partial body weight support compared with physiotherapy in nonambulatory hemiparetic patients. *Stroke* **1995**, *26*, 976–981. [CrossRef] [PubMed]

22. Wade, J.; Beccani, M.; Myszka, A.; Bekele, E.; Valdastri, P.; Flemming, P.; de Riesthal, M.; Withrow, T.; Sarkar, N. Design and implementation of an instrumented cane for gait recognition. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 5904–5909.

23. Mercado, J.; Chu, G.; Imperial, E.J.; Monje, K.G.; Pabustan, R.M.; Silverio, A. Smart cane: Instrumentation of a quad cane with audio-feedback monitoring system for partial weight-bearing support. In Proceedings of the 2014 IEEE International Symposium on Bioelectronics and Bioinformatics (ISBB), Chung Li, Taiwan, 11–14 April 2014; pp. 1–4.

24. Kumar, K.; Champaty, B.; Uvanesh, K.; Chachan, R.; Pal, K.; Anis, A. Development of an ultrasonic cane as a navigation aid for the blind people. In Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), Kanyakumari, India, 10–11 July 2014; pp. 475–479.

25. Rahman, M.; Abdullah, A.; Mohammad, A.; Nordin, M.; Ali, M.M.; Saaid, M. Smart cane with human detection capabilities for blind people. *J. Fundam. Appl. Sci.* **2018**, *10*, 989–1004.

26. Dang, D.C.; Suh, Y.S. Walking Distance Estimation Using Walking Canes with Inertial Sensors. *Sensors* **2018**, *18*, 230. [CrossRef] [PubMed]

27. Routson, R.L.; Bailey, M.; Pumford, I.; Czerniecki, J.M.; Aubin, P.M. A smart cane with vibrotactile biofeedback improves cane loading for people with knee osteoarthritis. In Proceedings of the 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3370–3373.

28. Aragaki, D.R.; Nasmyth, M.C.; Schultz, S.C.; Nguyen, G.M.; Yentes, J.M.; Kao, K.; Perell, K.; Fang, M.A. Immediate effects of contralateral and ipsilateral cane use on normal adult gait. *PM&R* **2009**, *1*, 208–213.

29. Hollman, J.H.; McDade, E.M.; Petersen, R.C. Normative spatiotemporal gait parameters in older adults. *Gait Posture* **2011**, *34*, 111–118. [CrossRef] [PubMed]

30. Fritz, S.; Lusardi, M. White paper: "Walking speed: The sixth vital sign". *J. Geriatr. Phys. Ther.* **2009**, *32*, 2–5. [CrossRef]

31. Bohannon, R.W. Relationship among paretic knee extension strength, maximum weight-bearing, and gait speed in patients with stroke. *J. Stroke Cerebrovasc. Dis.* **1991**, *1*, 65–69. [CrossRef]

Effects of Gait Strategy and Speed on Regularity of Locomotion Assessed in Healthy Subjects Using a Multi-Sensor Method

Marco Rabuffetti *, Giovanni Marco Scalera and Maurizio Ferrarin

IRCCS Fondazione Don Carlo Gnocchi, Milano 20121, Italy; gscalera@dongnocchi.it (G.M.S.);
mferrarin@dongnocchi.it (M.F.)
* Correspondence: mrabuffetti@dongnocchi.it; Tel.: +39-02-40308-544

Abstract: The regularity of pseudo-periodic human movements, including locomotion, can be assessed by autocorrelation analysis of measurements using inertial sensors. Though sensors are generally placed on the trunk or pelvis, movement regularity can be assessed at any body location. Pathological factors are expected to reduce regularity either globally or on specific anatomical subparts. However, other non-pathological factors, including gait strategy (walking and running) and speed, modulate locomotion regularity, thus potentially confounding the identification of the pathological factor. The present study's objectives were (1) to define a multi-sensor method based on the autocorrelation analysis of the acceleration module (norm of the acceleration vector) to quantify regularity; (2) to conduct an experimental study on healthy adult subjects to quantify the effect on movement regularity of gait strategy (walking and running at the same velocity), gait speed (four speeds, lower three for walking, upper two for running), and sensor location (on four different body parts). Twenty-five healthy adults participated and four triaxial accelerometers were located on the seventh cervical vertebra (C7), pelvis, wrist, and ankle. The results showed that increasing velocity was associated with increasing regularity only for walking, while no difference in regularity was observed between walking and running. Regularity was generally highest at C7 and ankle, and lowest at the wrist. These data confirm and complement previous literature on regularity assessed on the trunk, and will support future analyses on individuals or groups with specific pathologies affecting locomotor functions.

Keywords: wearable/inertial sensors; accelerometer; regularity; variability; human; motion; locomotion; autocorrelation

---

1. Introduction

Human functional periodic movements include locomotion, which is usually described in relation to its fundamental period [1], and upper limb activities, particularly those related to working tasks, sports, or art performances [2]. However, it is also possible to observe non-functional periodic movements such as tremor [3] or some cyclic gestures related to dystonic syndromes [4]. The periodicity of human locomotion is normally automatic and not necessarily consciously planned, controlled, and performed since specific neural structures, the central pattern generators, are in charge of controlling locomotor movements [5]. Moreover, inherent passive biomechanical and inertial characteristics of the different body parts naturally support the occurrence of passive pendulum-like periodic movements [6].

While, strictly speaking, a periodic phenomenon/signal repeats itself every predefined period of time T, a human cyclic movement and, particularly, kinematic and dynamic variables related to it

are only approximately periodic; small variations occur both in the time domain and in the physical domain of the phenomena (*y*), according to

$$y(t + T + \Delta T) = y(t) + \Delta y,\tag{1}$$

where *y* is a generic gait analysis variable (with the exception of point trajectories, but including their derivatives), *T* is the fundamental period, and *t* is the time variable. Smaller values of ΔT and Δy indicate the movement (described by a variable y) being closer to a periodic phenomenon. Therefore, since ΔT and Δy only tend to zero, it is more proper to refer to the "pseudo-periodicity" of human movements [7]. In the scientific literature, such a unique aspect was assessed by different quantitative methods and related numerical indexes, concerning variability [8–13], stability [13,14], and regularity [12,15–21].

The variability of a pseudo-periodic movement may be increased due to endogenous physiological factors, like dual-task interference [13], or by pathological factors, such as those associated with neurodegenerative diseases [10]. Therefore, the quantification of the alteration of movement periodicity, i.e., the regularity of variables related to the performance, may represent a pathological marker to be considered in the clinical decision-making process [22], but only if confounding factors are taken into consideration.

Referring to gait, many studies addressed the variability of gait-related temporal parameters both in healthy individuals, to explore the possible influence of age [15] and different gait strategies (i.e., walking and running) [23], and in specific pathologic conditions, e.g., in frail elderly people [24], in people with neurological diseases, either degenerative [10] or focal [25], and in persons suffering from orthopedic diseases [12]. Interestingly, gait analysis methodologies based on wearable sensors [26] can efficaciously support quantitative assessment of variability and regularity; acceleration and angular velocity of anatomical parts, which are directly measured by inertial units, are characterized by pseudo-periodic patterns according to the pseudo-periodic nature of the considered locomotor act [27].

The study of locomotion regularity is generally based on an established time-domain approach involving autocorrelation analysis; movement regularity is assumed as the degree of similarity between two consecutive patterns of variables or signals characterizing the same cyclic movement [28,29]. Particularly, the measured acceleration undergoes an autocorrelation analysis [17,28] or an autocovariance analysis [29], whose outcome functions are characterized by a peak in correspondence of the fundamental period of the signal itself; the greater the peak Y-value is, the more regularly the signal pattern repeats. This latter number, thus, represents a regularity index. This method was successfully applied to the assessment of various samples of people, including elderly [19,30] and persons with locomotor disturbances [17,21,31–33].

Many factors, particularly gait speed [9,20,27] and gait strategy, i.e., walking or running [23], but also cognitive loads [20] and shoe type [16], were identified as modulating gait regularity. Moreover, when tracking regularity during long-term monitoring, an effect of fatigue on regularity was shown [34]. Furthermore, in some specific cases such as a musical performance, the regularity can be voluntarily modulated [2].

As to the sensor position, few studies considered more than one sensor placed on the trunk or pelvis to study head stabilization [35] or to assess across-sensor agreement [36]. All the cited articles about regularity reported analysis on single components of acceleration, thus requiring the identification of the meaningful acceleration components and the related accurate sensor alignment with the considered anatomical plane. Moreover, only sensors located on the trunk or pelvis were considered, and were, thus, unable to report regularity for limb movements.

The present study's aims were as follows:

(A) To define a method to assess regularity by generalizing and further developing the already proposed method based on autocorrelation analysis [29]; innovative aspects include (1) the autocorrelation analysis being applied to the module of acceleration (i.e., norm of the acceleration vector) and not to one acceleration component, thus removing errors due to sensor misalignment;

(2) a multi-sensor approach which enables comparatively studying the regularity of anatomical parts, particularly of upper and lower limbs;

(B) To perform experiments and analyses on healthy subjects to quantify the effect of factors "strategy" (walking vs. running), "speed" (four speeds considered) [37], and "sensor location" (synchronized sensors located on pelvis, the seventh cervical vertebra (C7), and lower and upper limbs) on the regularity of human pseudo-periodic locomotor movements.

This study outcome may support future studies on human locomotion regularity by providing a robust methodology and reference data. Moreover, the application of the method to the study of non-functional periodic movements, such as tremors, is straightforward.

2. Materials and Methods

2.1. Measurement System

The measurement system used was the WaveTrack Inertial System (Cometa Systems, Italy; *https://www.cometasystems.com/*), which supported the four-sensor setting required for this study. The single device was a low weight (5.3 g), small-sized (32 mm × 24 mm × 7 mm), waterproof, inertial measurement unit (IMU; composed of a triaxial accelerometer, gyroscope, and magnetometer) with a wireless interface for real-time data streaming and a 1 GB of on-board memory, able to store up to 6 h of measurements. In the present study, only measurements of the triaxial accelerometer were considered; the sampling rate was 140 Hz with a 16-bit resolution analog-to-digital converter (ADC) and a full scale of ±16 g. Velcro elastic straps were provided for attaching sensors on selected anatomical landmarks.

2.2. Participants, Motor Tasks, and Raw Data

Twenty-five healthy subjects (14 males and 11 females, age range of 20–40 years) were recruited for the experiment (see details in Table 1). They signed an informed consent under the approval of the local Ethics Committee.

Table 1. Participants' demographic and anthropometric data. M—male; F—female.

Sex	Age (years)	Body Height (m)	Body Weight (kg)
M	31	1.62	52
M	26	1.76	85
F	22	1.61	45
F	29	1.60	49
M	32	1.88	86
M	40	1.62	48
F	27	1.67	53
M	30	1.70	61
M	24	1.81	75
F	28	1.85	73
M	35	1.75	75
M	26	1.70	57
F	20	1.74	59
F	23	1.79	65
M	24	1.70	70
M	25	1.89	83
M	26	1.80	74
M	25	1.70	75
F	22	1.62	46
M	29	1.74	75
F	26	1.61	45
F	23	1.60	51
F	23	1.60	55
M	24	1.80	79
F	23	1.70	55

Sensors were attached, by their flat surface, to the following landmarks: posterior aspect of spinous process of the seventh cervical vertebra (C7), midpoint between posterior superior iliac spines (pelvis), on dorsal aspect of the right wrist (wrist), and on lateral aspect of right ankle (ankle) (see Figure 1). Sensors were positioned in order to align their *X*-axes with longitudinal/vertical anatomical axes, sensor *Z*-axes were perpendicular to the sensor flat surface and were oriented accordingly.
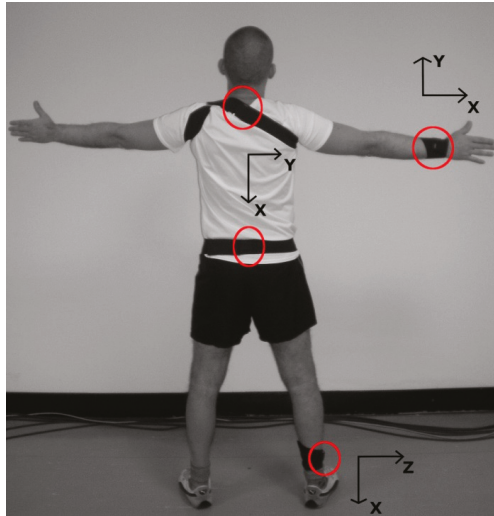
Figure 1. Subject equipped with four wearable triaxial accelerometers (Cometa Systems, Italy) positioned with elastic bands on the seventh cervical vertebra (C7), pelvis, wrist, and ankle. Sensor positions are circled in red, and axis directions are reported nearby.

The experiments were carried out on a treadmill and consisted of steady-state walking at 1.0, 1.4, and 1.8 m/s, and of steady-state running at 1.8 and 2.2 m/s (trials respectively labeled W1, W2, W3, R3, and R4). After taking between 30 and 60 s to reach steady-state locomotion, the duration of measurements was fixed at 70 s, the initial and final five seconds sections were discarded, and the further analysis was performed on the central one-minute-long recording. Trial sequences were randomly balanced among subjects; half of the participants performed the sequence W1, W2, W3, R3, and R4, and half performed the reversed sequence from R4 to W1.

Measured accelerometric components were low-pass filtered (fourth-order Butterworth filter, cutoff frequency 5 Hz, 0.5 dB of peak-to-peak ripple in pass-band, and 20 dB of attenuation in stop-band). Finally, acceleration module a was computed as the norm of the three measured orthogonal components (a_x, a_y, and a_z) according to the following formula:

$$a = \left| \vec{a} \right| = \sqrt{a_x^2 + a_y^2 + a_z^2}. \tag{2}$$

The experimental dataset consisted of 500 recordings obtained from 25 subjects, each performing five tasks, and equipped with four sensors.

2.3. Algorithm

The algorithm flowchart and sample data are presented in Figure 2. The input data consisted of the three acceleration components measured by a single sensor (Figure 2A). The analysis was applied on windowed finite time series of the acceleration module (Figure 2B) and of single acceleration components.
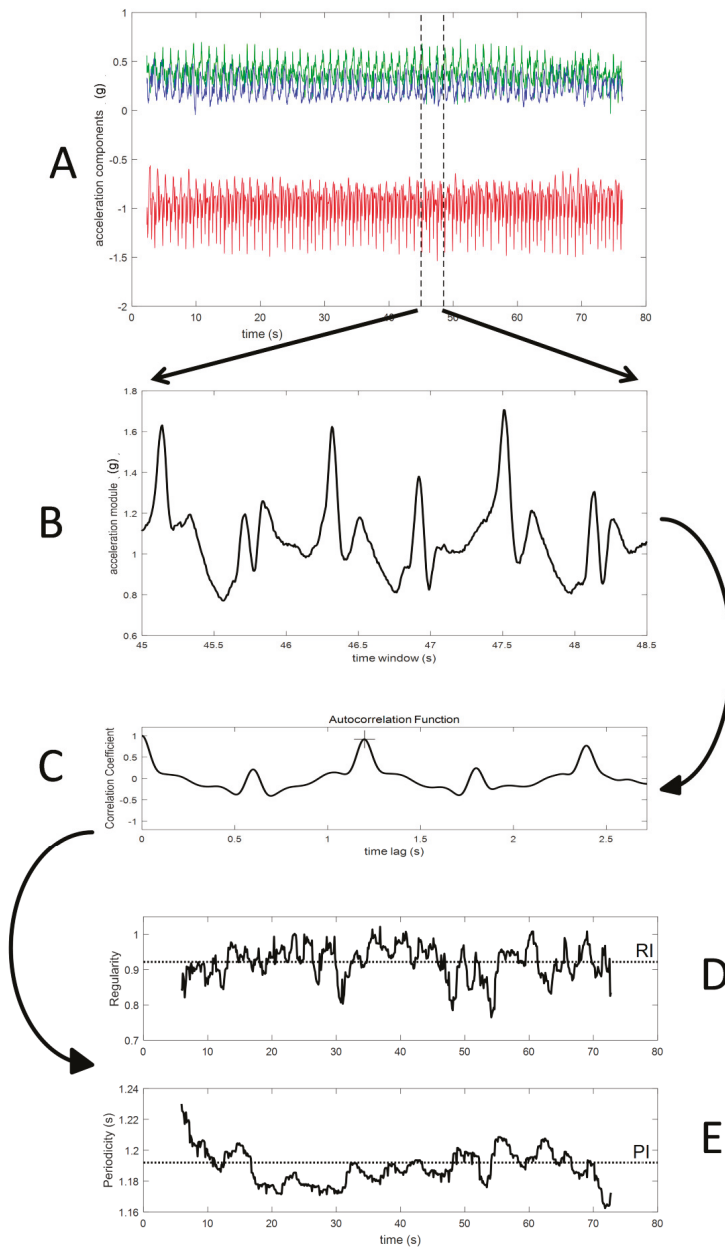
Figure 2. Method data flow: (**A**) raw recording from triaxial accelerometer; (**B**) window of acceleration module; (**C**) autocorrelation function computed on standardized windowed data (the cross marks the autocorrelation peak which corresponds to the window fundamental period); (**D**) the regularity time profile (the dotted line represents the window average value considered as a regularity index (RI)); (**E**) the fundamental period time profile (the dotted line represents the window average value considered as the period index (PI)).

The duration (*n* seconds, or alternatively *N* samples) of the signal window was fixed, on each trial, to three times the fundamental period preliminarily quantified on the entire one-minute recording (this duration ranged from about 2 s for the fastest running to about 4 s for the slower walking). Such a choice was sufficient for a robust computation of the autocorrelation function and avoided regularity variations across a higher number of strides may have affected the index estimation [31].

The windowed data were first standardized (removed average and then divided by standard deviation) and then an autocorrelation analysis was applied, which is equivalent to an autocovariance analysis since a unitary covariance is obtained when correlating the signal with itself at zero lag. Thus, a unitary value of the autocorrelation function reflects a perfect regularity, while a null value marks the absence of any cyclic component, i.e., the absence of any regularity.

The autocorrelation sequence for a given data window (Figure 2C), which is a function of the sample lag *m*, was implemented as follows:

$$\hat{R}_{\text{unbiased}}[m] = \frac{1}{N-m} \sum_{n=0}^{N-m-1} y[n+m]y[n], \tag{3}$$

where *R* is the autocorrelation function in the unbiased form, *y* is the analyzed variable, *N* is the number of data samples in the considered time window, and *m* is the number of samples quantifying the lag time of the autocorrelation function.

Notably, the previous equation includes a coefficient corrected for *m*, to compensate for the bias resulting from the *m* number of samples associated with the considered time lag (in fact, the larger the lag, the fewer samples support the computation) [29]. Negative lag values were not considered since they did not provide additional information, as autocorrelation is an even function.

Window periodicity, i.e., the fundamental period duration of the window data, corresponds to the time of the autocorrelation peak (Figure 2C). Window regularity corresponds to the peak value; the closer the input data are to a perfect periodic behavior, the closer the window regularity is to a unitary value.

The previous analysis was then repeated for the subsequent data window which was time-shifted from the previous window by 0.1 s. This implies that two subsequent data windows, for locomotion recordings, may overlap for more than 95% of their data. In order to speed up and increase algorithm robustness, when considering the next data window, the previous identification of the autocorrelation peak was considered as the starting seed in the search of the actual peak of the autocorrelation function, thus substantially avoiding the identification of peaks related to higher harmonics of the fundamental period.

The regularity index (RI) and period index (PI) were computed as the average values of the sequences of window regularity (Figure 2D) and periodicity (Figure 2E), respectively.

All analyses were implemented in Matlab (version R2017b, The MathWorks Inc, USA).

2.4. Statistical Analysis

The outcome set included both the regularity index RI and the period index PI for all subjects, all tasks, and all sensor locations. As data were non-normally distributed, we used non-parametric tests; the Friedman test was used for non-parametric analysis of variance and the Wilcoxon test was used for post hoc and planned comparisons. The significance values (significance set to $p \leq 0.05$) of all the tests inside the same analysis were corrected according to Holm–Bonferroni method. The experiment was designed to study the effect of three factors: a sensor factor, a speed factor, and a strategy (walking vs. running) factor. No analysis of interaction between factors was planned for two reasons: the sample data size was relatively small to allow efficaciously looking for interactions, and our interests were focused only on the main effects of the three factors.

As to the period index PI, a one-way non-parametric analysis of variance was performed to assess the across-sensor method robustness.

As to the regularity index RI, the experimental design allowed for assessing several factors. The analysis for the sensor factor was performed using one-way non-parametric analysis of variance. The speed factor was analyzed using the following planned comparison tests: W1 vs. W2, W2 vs. W3, and R3 vs. R4. Other comparisons were considered not relevant (W1 vs. W3) or inadequate (for example, W1 vs. R4) to explore the speed factor. The strategy factor (locomotion by walking vs. locomotion by running) was analyzed using a planned comparison of W3 vs. R3. Other comparisons (e.g., W2 vs. R4) report the effect of the strategy factor mixed with the effect of other factors (speed) and, therefore, were not included in the planned analysis.

Finally, in order to compare outcomes from the present module-based method with outcomes from a single-component approach in the regularity assessment, the algorithm was applied also to the single components of measured accelerations. Starting from RI_X, RI_Y, and RI_Z (regularity indexes of X, Y, and Z components, respectively), the minimum (worst index value) and the maximum (best index value) of the three single-component regularity indexes were identified for each subject/task/sensor, and multiple comparisons were performed between the module-based regularity index and the best and worst single-component ones.

3. Results

A summary of the experimental data is presented in Figure 3; the scatterplot of the regularity index vs. period index clearly shows the clusters related to speed and strategy, while the data concerning the four different sensors are partially overlapped. The non-normal distributions of RI values are made apparent by the ceiling effect toward the unitary value.
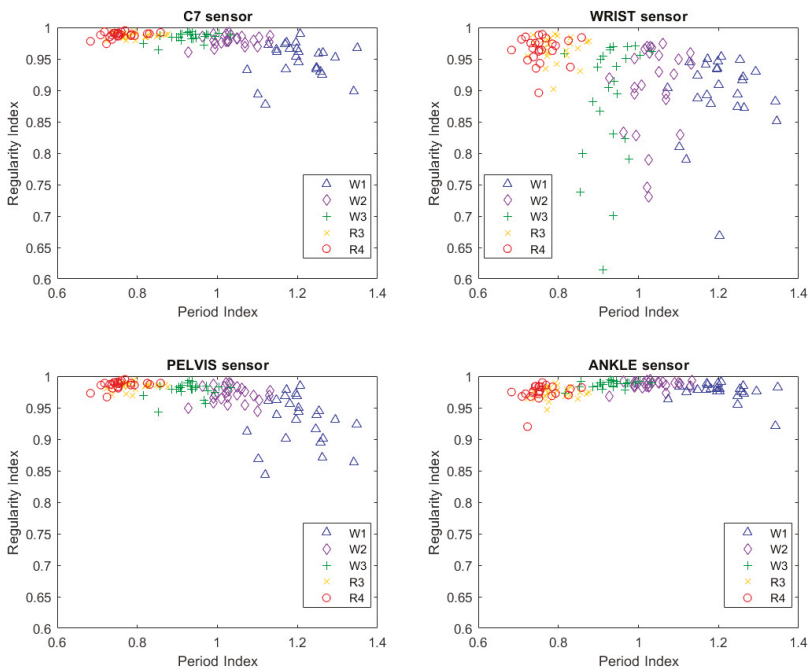
Figure 3. Summary scatterplot of all experimental outcomes: the period index is reported on the *X*-axis and the regularity index is reported on the *Y*-axis. Data from the four sensors are reported in separate plots. Different markers/color coding (refer to embedded legend) identify trial conditions (W1, W2, and W3 are for walking at 1.0, 1.4, and 1.8 m/s, respectively; R3 and R4 are for running at 1.8 and 2.2 m/s, respectively). The regularity index evidences an apparent ceiling effect.

3.1. Period Index

The fundamental period estimation, quantified by the PI value, showed the expected modulation in relation to task modality and speed (Figure 4A); period decreased with increasing speed and, at a matched speed of 1.8 m/s, the period was shorter when running compared to walking (R3 vs. W3). The period estimation as obtained on the same trial from different sensors may provide different values; when considering the four values relative to the four different anatomical landmarks, the difference between the two extreme values was null in 46% of the trials, and accounted for 1 ms in 44% of the trials; only in 13 out of 125 trials was the across-sensor period estimate difference between 2 and 5 ms. A data perusal allowed identifying the wrist and ankle as the sensor locations responsible for larger ($\geq$2 ms) deviations in period estimation.
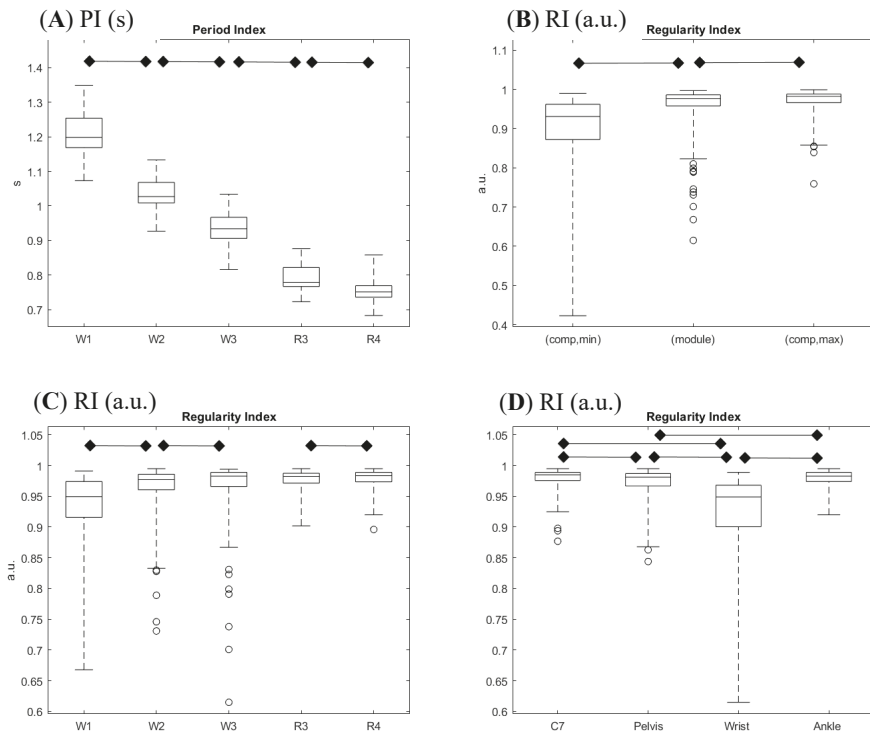
Figure 4. Composite presentation of results in 4 panels. (**A**) Period index (PI) distribution across tasks performed at different speeds and different modalities. (**B**) Comparison of the module-based regularity index (RI; center box) with the minimum (left box) and the maximum (right box) regularity indexes as computed on single components. (**C**) Regularity index distribution across tasks performed at different speeds and different modalities. (**D**) Regularity index distribution across sensor positions. Legends refer to tasks (W1, W2, and W3 for walking at 1.0, 1.4, and 1.8 m/s, respectively; R3 and R4 are for running at 1.8 and 2.2 m/s, respectively), to sensor positions (C7 on the posterior spinous process of the seventh cervical vertebra, pelvis on the midpoint between posterior superior iliac spine, wrist on the dorsal aspect of the right wrist, and ankle on the lateral aspect of right ankle), and to index definition (module for proposed computation on module of acceleration vector, min and max for extreme outcomes in single acceleration component computation). Boxes report medians and quartiles, whiskers represent extremes, and dots represent outliers. Horizontal bars with diamonds mark significant differences ($p \leq 0.05$, corrected for multiple tests) according to planned statistical analysis.

3.2. Regularity Index: Comparison of Module-Based and Component-Based Analyses

Figure 4B reports a summary of the regularity index values obtained from the analysis of acceleration module pattern, along with the min and the max regularity indexes obtained from single acceleration components. A perusal of outcomes shows that the module-based regularity index was higher or equal to any component-based index computed on the same recording in about half of the trials. The statistical analysis evidenced an intermediate rank of the module-based index, though its median value was much closer to the best component index (−1.0%) compared to the worst component index (+9.5%).

3.3. Regularity Index: Effect of Locomotion Speed

The effect of speed on regularity was highlighted by the comparisons between trials with the same locomotor strategy and with contiguous speeds (see Figure 4C). The tested couples were W1 vs. W2, W2 vs. W3, and R3 vs. R4. Regularity during walking trials significantly increased with increasing velocity, getting very close to the unit value for higher speeds. On the contrary, increasing velocity in running did not increase regularity. Nonetheless, it has to be noted that all values related to running showed a tendency to a ceiling effect with regularity indexes, limited in their values by the unitary upper limit, which marks a perfect regularity.

3.4. Regularity Index: Effect of Locomotor Strategy

The only test able to explore the effect of a different locomotor strategy was the comparison between the regularity indexes of W3 and R3, i.e., walking and running at the same velocity (1.8 m/s). This was the only feasible comparison, since healthy subjects can switch between walking and running without forcing themselves into a narrow range of speeds. No significant differences emerged from the Wilcoxon test, though a perusal of data distributions showed occasionally lower regularity in walking trials (see Figure 4C).

3.5. Regularity Index: Effect of Sensor Location

The statistical analysis showed a significant difference of regularity among locations; the highest regularity was found for C7 (median value 0.985) and the ankle (0.983), while the lowest was found for the wrist (0.949), and intermediate regularity was found for the pelvis (0.981) (Figure 4D). When further exploring the previous analysis by comparing the two locomotor strategies at a matched velocity, the median value of the regularity index increased in the running trials compared to the walking trials, at C7 (+1.0%), the pelvis (+1.6%), and the wrist (+4.8%), while it decreased (−1.0%) at the ankle.

4. Discussion

The present method for quantifying regularity in pseudo-periodic human movements was a direct evolution and a generalization of methods based on the autocorrelation analysis to study trunk oscillatory movements during locomotion [28,29]. The proposed method's novelty is in the application of the autocorrelation analysis to the module of acceleration and to the concurrent measurement of acceleration on four different anatomical points.

The autocorrelation analysis was applied to the module of acceleration, derived from measurements by a triaxial sensor, which was periodic to the extent that single components are periodic. This novel aspect was expected to imply a larger robustness than single-component methods since the latter require an accurate placement for sensor orientation and location (sensor axes have to be aligned with anatomical or functional axes, and this may be particularly prone to operator errors in experiments involving persons with abnormal anatomical features). On the contrary, the acceleration module is independent from sensor orientation, but relies only upon sensor location. As to the computational aspects, the autocorrelation was applied to a time-finite windowed signal after having standardized it (subtracted average and divided by standard deviation), implying a unitary value

for RI indicating perfect regularity. The average removal (equivalent to the autocovariance approach explicitly adopted by Moe-Nilssen [29]) is relevant, since an offset does corrupt, proportionally to the offset relevance/amplitude, the resulting autocorrelation coefficient and, thus, the regularity index. Therefore, this approach increases the robustness of the method.

The proposed sensor positions were not limited to location on the sagittal symmetry plane of the trunk–pelvis as in previous studies, but other body locations, particularly on either upper and lower limbs, were considered. This was possible since a periodic movement implies periodic patterns of the acceleration measured on any moving body location. A multi-sensor arrangement as already considered by Rispens et al. [36]; however, they only compared a regularity assessment of different locations on the trunk–pelvis segment, omitting the limbs. Obviously, locations outside the body's sagittal symmetry plane cannot support the analysis of symmetry according to Reference [29]; however, the multi-point approach here proposed has the advantage of disclosing how different body parts contribute to the movement regularity, thus supporting understanding and potentially targeting treatment or training on specific body segments.

The proposed autocorrelation analysis produced an autocorrelation time profile (see Figure 1E) whose average on a one-minute-long track was considered here as a regularity index (RI). However, analysis on longer recordings produced longer regularity profiles, which could be further studied for the occurrence of trends, potentially unveiling effects of fatigue on regularity. Schutte already dealt with this aspect by simply analyzing two spot assessments before and after a prolonged fatiguing task [34]. Moreover, the multi-sensor approach may further support a fatigue analysis allowing the identification of which anatomical functional part acts as a fatigue trigger.

The experimental data were obtained from one-minute-long locomotion trials at constant velocity on a treadmill, whereas previous studies considered recordings of walking for one minute or less [17,28,29]. The subjects were randomly requested to perform the sequence from slower walking to faster running or the reverse-order sequence, and no bias resulted in regularity index.

The experimental dataset also allowed, for all three components of acceleration singularly considered, to compute stride regularity according to the previous method [29]. For the sake of simplicity, the highest and the lowest outcome indexes as computed from single components were identified and then compared with the regularity indexes computed on the acceleration module pattern. The results show that the average difference between the highest single-component-based index and the module-based index was very little for the considered dataset, about 1% in relative terms, though statistically significant. However, no substantial bias was observed, since about half of the trials displayed a higher regularity index for the module-based index. Nonetheless, the option for a single-component regularity analysis, which showed validity in healthy subjects and in the study of selected locomotor disturbances [17,21,29], requires that this component be identified before the experiment takes place, which is a decision that may suffer from errors particularly when motor disturbances affect the performing subject. Ultimately, the proposed module-based analysis on a specific anatomical location [38] does not require that the researcher accurately align the sensor axes along predefined directions, which is particularly difficult when skeletal deformities affect the performing subject. Obviously, the adoption of triaxial accelerometers does not exclude the possibility to apply both the module-based approach and the single-component one.

Interesting findings emerged also from the analysis of period duration, a useful secondary outcome of the autocorrelation method. Firstly, it was confirmed that the estimate was very robust across sensors and across methods in accordance with published studies [39], and results were comparable with published normative data [40]. Secondly, the method provided a tracking of the fundamental period value along with the monitoring time; such information is supportive of studies concerning the gait variability [10]. Moreover, since the sensors may be placed on most parts of the body and adopted to any pseudo-periodic movements, this method appears to be of general applicability, while gait-specific methods for gait event identification cannot be extended to other motor tasks.

The application of the method was intended to disclose if different body segments showed different regularity during locomotion, and to analyze if locomotion strategy and locomotion speed influenced movement regularity.

Since the experiments involved the synchronized measurement of acceleration from four sensors, we were able to compare the autocorrelation coefficients observed on different anatomical sites; the highest regularity was found for C7 and the ankle, while the lowest (but still very high in absolute terms) was found for the wrist. The interpretation of C7's higher regularity may be related to biomechanical aspects (segments with larger mass and inertia, such as the trunk, show more regular movements, in the same manner that flywheels are expected to), to motor-control specific features (C7 movements are more regular because one of the objectives of locomotion is to keep the head, and the sensory organs inside it, the most stable and regular [35]) and to the balancing role played by limbs (particularly upper limbs) in keeping the body stability [41]. The reason for the lower regularity of upper limbs may include the possibility of accomplishing other episodic motor tasks, asynchronous with walking [42,43].

The effect of gait speed was assessed by comparing trials at different speeds but with the same strategy; a speed effect was observed for walking trials only, thus confirming that increasing speed is associated with increased regularity [20]. However, we observed that a faster locomotion resulted in the regularity index being closer to its maximum possible unitary value; therefore, it is reasonable to hypothesize the emergence of a ceiling effect at faster gait speeds, which could explain the missing detection of a speed effect for the running trials.

Interestingly, when comparing walking and running trials, the trunk (i.e., C7 and the pelvis) and, even more so, the upper arms tended to be more regular in running, according to the speed effect already evidenced, while lower limbs (ankle) were less regular. Such a decrease in regularity in lower limbs may be due to a technical artefact; the feet contacts on the ground are characterized, during running, by large impulsive impact forces which may transmit irregular oscillation on the nearby musculoskeletal structures and, therefore, also on the accelerometer tighten to the ankle [44].

As to the motor strategy, no between-strategy difference in regularity was observed at one single matched velocity close to the transition speed (RI_{W3} = RI_{R3} in Figure 4C), thus confirming previous studies concerning the analysis of spatio-temporal parameters [23]. While walking and running performances showed values of regularity index very close each other and close to the unitary upper limit value, a perusal of the whole dataset showed how RI never got below 0.9 when running, while 10 out of 100 RI values when walking were below 0.9, ranging down to approximately 0.7. Interestingly, all those ten values referred to regularity assessed by the wrist sensor. A possible explanation is that the role of upper limbs during running is more strictly related to locomotion than during walking, since they must counterbalance larger inertial forces which occur in the body [45]. Conversely, it can be expected that, during walking, the upper limbs keep the possibility to also perform some locomotor-unrelated tasks, characterized by less regular features [43].

5. Conclusions

The proposed method was proven to be feasible and reliable, to be able to quantify the regularity of movement of different anatomical parts, and to be able to track modulation of regularity determined by locomotor strategy and speed. The novel methodological approach of considering the acceleration module has the advantage over single-component methods [17,29] to be unaffected by sensor misalignment. The data from a group of healthy individuals may foster the collection of a larger dataset and provide reference in studies concerning pathological conditions.

Already planned future applications of the method include (a) a real-time application in providing biofeedback to the patient during periodic movements to help them in keeping a regular motor pattern (such as gait, but also upper limb movements), thus supporting motor learning rehabilitation protocols [46,47]; (b) studies on the modulation of regularity during prolonged performances, potentially induced by fatigue or by a voluntary change in motor strategy [48]; (c) studies on non

functional pseudo-periodic movements, such as tremor or dyskinesia, by means of regularity analysis of long-term actigraphic recordings [49].

Author Contributions: Conceptualization, M.R. and M.F.; methodology, M.R.; software, M.R. and G.M.S.; validation, M.R. and G.M.S.; formal analysis, M.R., G.M.S., and M.F.; investigation, G.M.S.; resources, G.M.S.; data curation, M.R. and G.M.S.; writing—original draft preparation, M.R.; writing—review and editing, M.R., G.M.S., and M.F.; visualization, M.R. and G.M.S.; supervision, M.F.; project administration, M.R. and M.F.; funding acquisition, M.R. and M.F.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bovi, G.; Rabuffetti, M.; Mazzoleni, P.; Ferrarin, M. A multiple-task gait analysis approach: Kinematic, kinetic and EMG reference data for healthy young and adult subjects. *Gait Posture* **2011**, *33*, 6–13. [CrossRef] [PubMed]

2. Demos, A.P.; Lisboa, T.; Chaffin, R. Flexibility of Expressive Timing in Repeated Musical Performances. *Front. Psychol.* **2016**, *7*, 1490. [CrossRef] [PubMed]

3. Deuschl, G.; Bain, P.; Brin, M. Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. *Mov. Disord.* **1998**, *13* (Suppl. 3), 2–23. [CrossRef] [PubMed]

4. Albanese, A.; Bhatia, K.; Bressman, S.B.; Delong, M.R.; Fahn, S.; Fung, V.S.C.; Hallett, M.; Jankovic, J.; Jinnah, H.A.; Klein, C.; et al. Phenomenology and classification of dystonia: A consensus update. *Mov. Disord.* **2013**, *28*, 863–873. [CrossRef] [PubMed]

5. Crenna, P.; Carpinella, I.; Lopiano, L.; Marzegan, A.; Rabuffetti, M.; Rizzone, M.; Lanotte, M.; Ferrarin, M. Influence of basal ganglia on upper limb locomotor synergies. Evidence from deep brain stimulation and L-DOPA treatment in Parkinson's disease. *Brain* **2008**, *131*, 3410–3420. [CrossRef] [PubMed]

6. Usherwood, J.R.; Channon, A.J.; Myatt, J.P.; Rankin, J.W.; Hubel, T.Y. The human foot and heel-sole-toe walking strategy: A mechanism enabling an inverted pendular gait with low isometric muscle force? *J. R. Soc. Interface* **2012**, *9*, 2396–2402. [CrossRef] [PubMed]

7. Pecoraro, F.; Mazzà, C.; Zok, M.; Cappozzo, A. Assessment of level-walking aperiodicity. *J. Neuroeng. Rehabil.* **2006**, *3*, 28. [CrossRef] [PubMed]

8. Hausdorff, J.M. Gait variability: Methods, modeling and meaning. *J. Neuroeng. Rehabil.* **2005**, *2*, 19. [CrossRef]

9. Jordan, K.; Challis, J.H.; Newell, K.M. Walking speed influences on gait cycle variability. *Gait Posture* **2007**, *26*, 128–134. [CrossRef]

10. Hausdorff, J.M. Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos* **2009**, *19*, 026113. [CrossRef]

11. Stergiou, N.; Decker, L.M. Human movement variability, nonlinear dynamics, and pathology: Is there a connection? *Hum. Mov. Sci.* **2011**, *30*, 869–888. [CrossRef] [PubMed]

12. Parker, K.; Hanada, E.; Adderson, J. Gait variability and regularity of people with transtibial amputations. *Gait Posture* **2013**, *37*, 269–273. [CrossRef] [PubMed]

13. Magnani, R.M.; Lehnen, G.C.; Rodrigues, F.B.; de Sá E Souza, G.S.; de Oliveira Andrade, A.; Vieira, M.F. Local dynamic stability and gait variability during attentional tasks in young adults. *Gait Posture* **2017**, *55*, 105–108. [CrossRef] [PubMed]

14. Bisi, M.C.; Riva, F.; Stagni, R. Measures of gait stability: Performance on adults and toddlers at the beginning of independent walking. *J. Neuroeng. Rehabil.* **2014**, *11*, 131. [CrossRef] [PubMed]

15. Kobsar, D.; Olson, C.; Paranjape, R.; Hadjistavropoulos, T.; Barden, J.M. Evaluation of age-related differences in the stride-to-stride fluctuations, regularity and symmetry of gait using a waist-mounted tri-axial accelerometer. *Gait Posture* **2014**, *39*, 553–557. [CrossRef] [PubMed]

16. Terrier, P.; Dériaz, O.; Meichtry, A.; Luthi, F. Prescription footwear for severe injuries of foot and ankle: Effect on regularity and symmetry of the gait assessed by trunk accelerometry. *Gait Posture* **2009**, *30*, 492–496. [CrossRef]

17. Tura, A.; Raggi, M.; Rocchi, L.; Cutti, A.G.; Chiari, L. Gait symmetry and regularity in transfemoral amputees assessed by trunk accelerations. *J. Neuroeng. Rehabil.* **2010**, *7*, 4. [CrossRef]
18. Kurz, M.J.; Hou, J.G. Levodopa influences the regularity of the ankle joint kinematics in individuals with Parkinson's disease. *J. Comput. Neurosci.* **2010**, *28*, 131–136. [CrossRef]
19. Kobayashi, H.; Kakihana, W.; Kimura, T. Combined effects of age and gender on gait symmetry and regularity assessed by autocorrelation of trunk acceleration. *J. Neuroeng. Rehabil.* **2014**, *11*, 109. [CrossRef]
20. Schaefer, S.; Jagenow, D.; Verrel, J.; Lindenberger, U. The influence of cognitive load and walking speed on gait regularity in children and young adults. *Gait Posture* **2015**, *41*, 258–262. [CrossRef]
21. Barden, J.M.; Clermont, C.A.; Kobsar, D.; Beauchet, O. Accelerometer-Based Step Regularity Is Lower in Older Adults with Bilateral Knee Osteoarthritis. *Front. Hum. Neurosci.* **2016**, *10*, 625. [CrossRef]
22. Cuzzolin, F.; Sapienza, M.; Esser, P.; Saha, S.; Franssen, M.M.; Collett, J.; Dawes, H. Metric learning for Parkinsonian identification from IMU gait measurements. *Gait Posture* **2017**, *54*, 127–132. [CrossRef]
23. Dingwell, J.B.; Bohnsack-McLagan, N.K.; Cusumano, J.P. Humans control stride-to-stride stepping movements differently for walking and running, independent of speed. *J. Biomech.* **2018**, *76*, 144–151. [CrossRef] [PubMed]
24. Mortaza, N.; Abu Osman, N.A.; Mehdikhani, N. Are the spatio-temporal parameters of gait capable of distinguishing a faller from a non-faller elderly? *Eur. J. Phys. Rehabil. Med.* **2014**, *50*, 677–691.
25. Sánchez, M.C.; Bussmann, J.; Janssen, W.; Horemans, H.; Chastin, S.; Heijenbrok, M.; Stam, H. Accelerometric assessment of different dimensions of natural walking during the first year after stroke: Recovery of amount, distribution, quality and speed of walking. *J. Rehabil. Med.* **2015**, *47*, 714–721. [CrossRef] [PubMed]
26. Muro-de-la-Herran, A.; Garcia-Zapirain, B.; Mendez-Zorrilla, A. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. *Sensors* **2014**, *14*, 3362–3394. [CrossRef] [PubMed]
27. Kim, Y.-K.; Joo, J.-Y.; Jeong, S.-H.; Jeon, J.-H.; Jung, D.-Y. Effects of walking speed and age on the directional stride regularity and gait variability in treadmill walking. *J. Mech. Sci. Technol.* **2016**, *30*, 2899–2906. [CrossRef]
28. Auvinet, B.; Berrut, G.; Touzard, C.; Moutel, L.; Collet, N.; Chaleil, D.; Barrey, E. Reference data for normal subjects obtained with an accelerometric device. *Gait Posture* **2002**, *16*, 124–134. [CrossRef]
29. Moe-Nilssen, R.; Helbostad, J.L. Estimation of gait cycle characteristics by trunk accelerometry. *J. Biomech.* **2004**, *37*, 121–126. [CrossRef]
30. Gillain, S.; Boutaayamou, M.; Dardenne, N.; Schwartz, C.; Demonceau, M.; Gerontitis, C.; Depierreux, F.; Salmon, E.; Garraux, G.; Bruyère, O.; et al. Data set of healthy old people assessed for three walking conditions using accelerometric and opto-electronic methods. *Aging Clin. Exp. Res.* **2017**, *29*, 1201–1209. [CrossRef] [PubMed]
31. Tura, A.; Rocchi, L.; Raggi, M.; Cutti, A.G.; Chiari, L. Recommended number of strides for automatic assessment of gait symmetry and regularity in above-knee amputees by means of accelerometry and autocorrelation analysis. *J. Neuroeng. Rehabil.* **2012**, *9*, 11. [CrossRef]
32. Yang, C.-C.; Hsu, Y.-L.; Shih, K.-S.; Lu, J.-M. Real-time gait cycle parameter recognition using a wearable accelerometry system. *Sensors* **2011**, *11*, 7314–7326. [CrossRef]
33. Demonceau, M.; Donneau, A.-F.; Croisier, J.-L.; Skawiniak, E.; Boutaayamou, M.; Maquet, D.; Garraux, G. Contribution of a Trunk Accelerometer System to the Characterization of Gait in Patients with Mild-to-Moderate Parkinson's Disease. *IEEE J. Biomed. Health Inf.* **2015**, *19*, 1803–1808. [CrossRef]
34. Schütte, K.H.; Maas, E.A.; Exadaktylos, V.; Berckmans, D.; Venter, R.E.; Vanwanseele, B. Wireless Tri-Axial Trunk Accelerometry Detects Deviations in Dynamic Center of Mass Motion Due to Running-Induced Fatigue. *PLoS ONE* **2015**, *10*, e0141957. [CrossRef]
35. Menz, H.B.; Lord, S.R.; Fitzpatrick, R.C. Acceleration patterns of the head and pelvis when walking on level and irregular surfaces. *Gait Posture* **2003**, *18*, 35–46. [CrossRef]
36. Rispens, S.M.; Pijnappels, M.; van Schooten, K.S.; Beek, P.J.; Daffertshofer, A.; van Dieën, J.H. Consistency of gait characteristics as determined from acceleration data collected at different trunk locations. *Gait Posture* **2014**, *40*, 187–192. [CrossRef]
37. Jin, L.; Hahn, M.E. Modulation of lower extremity joint stiffness, work and power at different walking and running speeds. *Hum. Mov. Sci.* **2018**, *58*, 1–9. [CrossRef]

38. Della Croce, U.; Leardini, A.; Chiari, L.; Cappozzo, A. Human movement analysis using stereophotogrammetry. Part 4: Assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait Posture* **2005**, *21*, 226–237. [CrossRef]

39. Fortune, E.; Lugade, V.; Morrow, M.; Kaufman, K. Validity of using tri-axial accelerometers to measure human movement—Part II: Step counts at a wide range of gait velocities. *Med. Eng. Phys.* **2014**, *36*, 659–669. [CrossRef]

40. Hansen, E.A.; Kristensen, L.A.R.; Nielsen, A.M.; Voigt, M.; Madeleine, P. The role of stride frequency for walk-to-run transition in humans. *Sci. Rep.* **2017**, *7*, 2010. [CrossRef]

41. Wu, Y.; Li, Y.; Liu, A.-M.; Xiao, F.; Wang, Y.-Z.; Hu, F.; Chen, J.-L.; Dai, K.-R.; Gu, D.-Y. Effect of active arm swing to local dynamic stability during walking. *Hum. Mov. Sci.* **2016**, *45*, 102–109. [CrossRef]

42. Carpinella, I.; Crenna, P.; Rabuffetti, M.; Ferrarin, M. Coordination between upper- and lower-limb movements is different during overground and treadmill walking. *Eur. J. Appl. Physiol.* **2010**, *108*, 71–82. [CrossRef]

43. Lamberg, E.M.; Muratori, L.M. Cell phones change the way we walk. *Gait Posture* **2012**, *35*, 688–690. [CrossRef]

44. Lafortune, M.A. Three-dimensional acceleration of the tibia during walking and running. *J. Biomech.* **1991**, *24*, 877–886. [CrossRef]

45. Arellano, C.J.; Kram, R. The metabolic cost of human running: Is swinging the arms worth it? *J. Exp. Biol.* **2014**, *217*, 2456–2461. [CrossRef]

46. Scalera, G.M.; Rabuffetti, M.; Marzegan, A.; Frigo, C.; Ferrarin, M. Regularity assessment of cyclic human movements: An innovative method based on wearable sensors. In Proceedings of the 2017 E-Health and Bioengineering Conference (EHB), Sinaia, Romania, 22–24 June 2017; pp. 454–457.

47. Jonsdottir, J.; Cattaneo, D.; Recalcati, M.; Regola, A.; Rabuffetti, M.; Ferrarin, M.; Casiraghi, A. Task-oriented biofeedback to improve gait in individuals with chronic stroke: Motor learning approach. *Neurorehabil. Neural Repair* **2010**, *24*, 478–485. [CrossRef]

48. Erdmann, W.S.; Lipinska, P. Kinematics of marathon running tactics. *Hum. Mov. Sci.* **2013**, *32*, 1379–1392. [CrossRef]

49. Rabuffetti, M.; Meriggi, P.; Pagliari, C.; Bartolomeo, P.; Ferrarin, M. Differential actigraphy for monitoring asymmetry in upper limb motor activities. *Physiol. Meas.* **2016**, *37*, 1798–1812. [CrossRef]

Feasibility of Home-Based Automated Assessment of Postural Instability and Lower Limb Impairments in Parkinson's Disease

Claudia Ferraris [1,4,*]**, Roberto Nerino** [1]**, Antonio Chimienti** [1]**, Giuseppe Pettiti** [1]**, Nicola Cau** [3]**, Veronica Cimolin** [2]**, Corrado Azzaro** [3]**, Lorenzo Priano** [3,4] **and Alessandro Mauro** [3,4]

[1] Institute of Electronics, Computer and Telecommunication Engineering, National Research Council, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; roberto.nerino@ieiit.cnr.it (R.N.); antonio.chimienti@ieiit.cnr.it (A.C.); giuseppe.pettiti@ieiit.cnr.it (G.P.)

[2] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; veronica.cimolin@polimi.it

[3] Istituto Auxologico Italiano, IRCCS, Department of Neurology and NeuroRehabilitation, S. Giuseppe Hospital, 28824 Piancavallo, Oggebbio (Verbania), Italy; n.cau@auxologico.it (N.C.); c.azzaro@auxologico.it (C.A.); lorenzo.priano@unito.it (L.P.); alessandro.mauro@unito.it (A.M.)

[4] Department of Neurosciences, University of Turin, Via Cherasco 15, 10100 Torino, Italy

[*] Correspondence: claudia.ferraris@ieiit.cnr.it; Tel.: +39-011-090-5405

Abstract: A self-managed, home-based system for the automated assessment of a selected set of Parkinson's disease motor symptoms is presented. The system makes use of an optical RGB-Depth device both to implement its gesture-based human computer interface and for the characterization and the evaluation of posture and motor tasks, which are specified according to the Unified Parkinson's Disease Rating Scale (UPDRS). Posture, lower limb movements and postural instability are characterized by kinematic parameters of the patient movement. During an experimental campaign, the performances of patients affected by Parkinson's disease were simultaneously scored by neurologists and analyzed by the system. The sets of parameters which best correlated with the UPDRS scores of subjects' performances were then used to train supervised classifiers for the automated assessment of new instances of the tasks. Results on the system usability and the assessment accuracy, as compared to clinical evaluations, indicate that the system is feasible for an objective and automated assessment of Parkinson's disease at home, and it could be the basis for the development of neuromonitoring and neurorehabilitation applications in a telemedicine framework.

Keywords: Parkinson's disease; UPDRS tasks; movement disorders; posture; postural stability; center of mass; RGB-depth; automated assessment; machine learning; at-home monitoring; neurorehabilitation

1. Introduction

Among chronic neurodegenerative diseases, Parkinson's disease (PD) is recognized as the second most common disorder after Alzheimer's disease. It causes an important negative impact on the quality of life characterized by a progressive impairment in motor functions [1].

Neurologists employ clinical assessment scales, such as the Part III of the Unified Parkinson's Disease Rating Scale (UPDRS) [2], as a common basis to assess the motor impairment severity and its progression over time. During the patient assessment, particular features of the movements (e.g., amplitude, speed, rhythm, hesitations) or of the posture (e.g., trunk flexion, one-side leaning and posture recover capabilities) are subjectively evaluated by neurologists on a discrete scale of five classes of increasing severity, with reliability limitations due to intra and inter-rater variability [3]. Aiming to

improve the clinical management and the quality of life of individuals with PD, more objective and automated approaches to disease assessment, also suitable for at home use, have been proposed.

The majority of these approaches employ wearable technologies [4,5], specifically in lower limbs UPDRS tasks assessment [6]; fewer of them are based on optical tracking, smartphones and other technologies [7,8]. Wearable and optical based technologies exhibit complementary aspects: the first ones are more ubiquitous but they are also more invasive and require more management efforts; the second ones are suited for spot assessment in a localized environment, but they are non-invasive. In general, for the assessment, both approaches make use of the correlation existing between the severity of the impairment, as assessed by UPDRS, and the static and/or kinematic parameters characterizing the pose and the movements [9,10].

Recently many low-cost, optical, body and hand tracking systems [11–14] have been employed successfully in the health care context. Among these, the Microsoft Kinect$^{®}$ v1 device has been used to monitor people with PD [15], in rehabilitation [16,17], in body sway and balance [18,19], in gait assessment [20] and gait anomalies detection [21], in identifying different subjects by kinematic signature [22], in hand tracking [23] and to prevent falls [24]. The Microsoft Kinect v2 is more robust and accurate as compared to Microsoft Kinect v1 [25], and it has been recognized a viable tool for tracking human movement in clinical applications [26], standing balance and postural stability [27], gait [28], body sway [29] and clinical motor functions [30]. In the specific context of neuro-degenerative diseases, it has been used for time up and go test [31], in assessing different types of PD patients [32], to classify gait disorders [33] and in neurological rehabilitation [34].

The work presented here is part of a more extensive project aimed to bring UPDRS compliant automated assessments at patients' home. In a former paper [35], we presented our work on the upper limb UPDRS tasks. Instead, in this case, the assessment is based on a Microsoft Kinect v2 device and it is focused on the analysis of posture and lower limb tasks, as specified by UPDRS [2]. This approach guarantees both to compare results with the standard clinical assessment scales, accepted and used by neurologists, and to define explicitly posture and movements to be performed. Furthermore, we select those UPDRS motor items which are suitable to be self-managed by patients at home, considering that some motor tasks are not feasible in any home environments. For example, according to task specifications [2], Gait task (UPDRS task 3.10) requires a safe straight walking path of 4–10 m, which is not usually available at home, as well as the Postural Stability task (UPDRS task 3.12) that cannot be self-administered being based on a retropulsion test.

Specifically, we examine the following UPDRS tasks (Section 3, items 3.8, 3.9 and 3.13): Leg Agility (LA), Arising from chair (AC) and Posture (Po). Concerning Postural Stability UPDRS task (PS$_{retrop}$, item 3.12), the standard retropulsion test used for the assessment is not a good predictor of fall risk. Furthermore, the related step count parameter is a too rough estimator of the postural instability [36]. Nevertheless, postural stability assessment is important to prevent falls and injury risk in PD, especially in advanced stages [37]. Postural stability deficits in PD subjects can be highlighted by concurrent cognitive tasks or by secondary motor tasks during steady standing stance tasks [38]. Several studies have found that, during the quite stance, the continuous movement of the center of mass (CoM), named "postural sway" or "body sway", contributes to balance control [39–41]. Alterations of body sway can reveal balance dysfunctions in PD, long before their clinical assessment [42] and they can be used to differentiate between motor subtypes of PD [43]. Recently, low cost RGB-Depth devices as Microsoft Kinect v2, have been used to assess objectively balance dysfunctions [44–46]. Given the importance of postural stability in PD progression and risk of fall prediction, we analyze the postural stability by CoM movements (PS$_{COM}$) during the posture task. Furthermore, we investigate also the potential correlation between this method and a standard clinical measure of postural stability PS$_{PIGD}$ based on the Postural Instability and Gait Difficulty (PIGD) subscale score of UPDRS, defined as the sum of the scores assigned to the AC, Gait, PS$_{retrop}$ and Po tasks of UPDRS [47]. To our knowledge, this is the first time, in the context of PD, that a set of UPDRS tasks (namely, LA, AC, Po) and the Postural

Stability (PS$_{COM}$) during quite stance are automatically assessed by the use of low-cost optical body tracking devices.

2. System Hardware and Software

The system hardware (Figure 1a) is built around a Microsoft Kinect v2 device (Microsoft Corporation, Redmond, WA, USA), which provides through its Software Development Kit (SDK) [12], RGB color and DEPTH streams at 30 frame/s, with resolution of 1920 × 1080 px and 512 × 424 px, respectively. The range of depth is from 0.5 m to 4.5 m. The device is connected, via an USB 3 port, to a NUC i7 Intel® mini-PC running Windows® 10 (64x) (Intel Corporation, Santa Clara, CA, USA) and equipped with a monitor to provide both a system management GUI and the visual feedback of the hand and body movements to the user (Figure 1b).

(**a**)   (**b**)
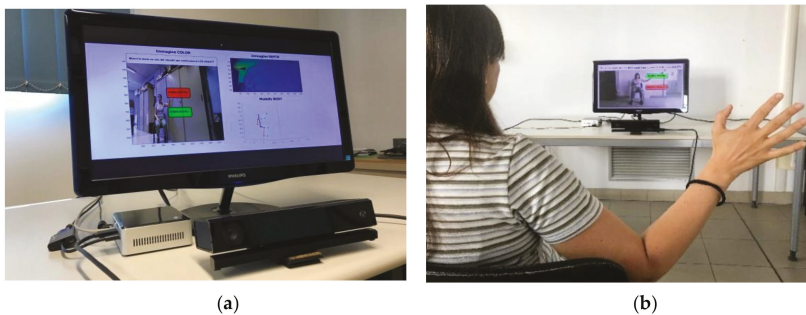
Figure 1. System for the lower limbs and postural tasks analysis: (**a**) RGB-Depth camera (Microsoft Kinect v2), NUC i7 Intel mini-PC and monitor (**b**) example of GUI with visual feedback.

The system software is made by custom scripts, written in C++, which run on NUC and access the SDK APIs, providing every 1/30 s RGB images and 25 three-dimensional (3D) coordinates of the skeleton model used by the SDK (Figure 2).

(**a**)   (**b**)
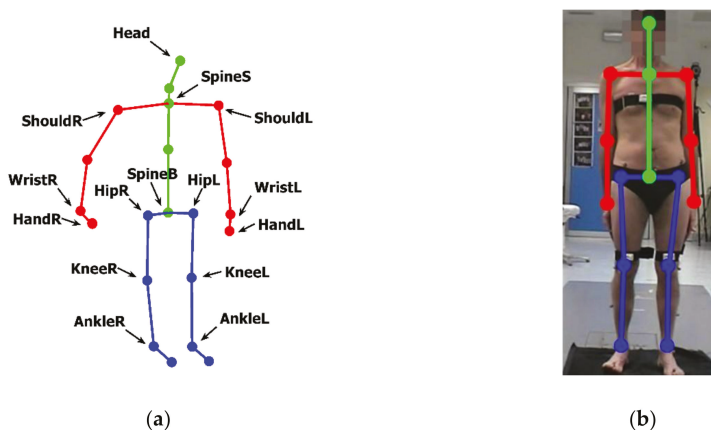
Figure 2. Positions of joints of the skeleton model from Microsoft Kinect SDK: (**a**) three-dimensional representation of joints and segments for body vertical axis (green), upper limbs (red), lower limbs (blue); (**b**) two-dimensional re-projection of the same joints and segments on the RGB image.

The data analysis and the supervised classifier training and testing phases are based on custom Matlab® scripts (Mathworks Inc, Natick, MA, USA). The software implements different functionalities

of the system: real-time interaction by a Human Computer Interface (HCI) based on hand joint tracking/processing and visual feedback; task movement analysis and characterization, by processing the 3D positions of specific task-dependent sets of skeleton joints; automated assessment of posture and lower limb tasks, through the implementation of trained supervised classifiers. Data of each acquisition session (consisting of video of each task performance, user inputs, trajectories of body movements and automated assessment scores) are encrypted and recorded to provide remote supervising facilities to authorized clinicians.

3. The Human Computer Interface

The HCI provides a natural interface suitable for subjects with limited computer skills and with motor impairments. It is implemented through a Graphical User Interface (GUI) and an interactive menu based on choice icons (Figure 3). During the interaction with the system, the user is guided by video and textual support. The 3D position of HandR joint, output by SDK, is tracked and re-projected onto the GUI screen, and the user selection is confirmed by considering the hand closure information provided by the SDK (Figures 1b, 2a and 3). At any time during an assessment session, the user can stop it and quit, for example when tired, to avoid the onset of stress and/or anxiety.

(a)   (b)
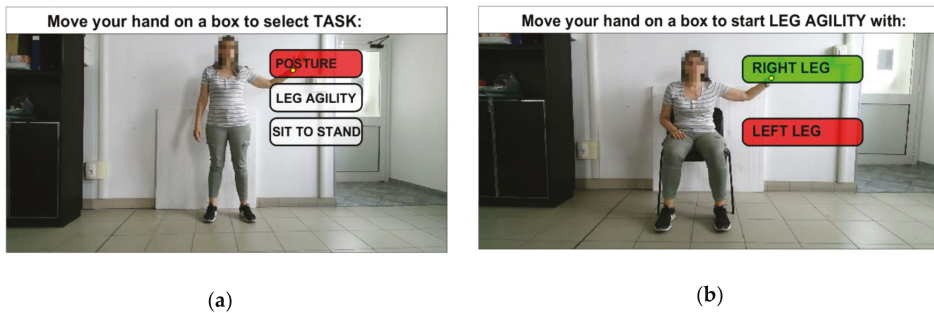
Figure 3. Gesture-based HCI: (**a**) GUI for the selection of lower limbs and postural tasks; (**b**) GUI for the selection of left/right leg before starting LA task.

4. Automated Assessment of UPDRS Tasks

4.1. Participant Recruitment

Two cohorts of subjects, consisting of fourteen PD patients and twelve Healthy Controls (HC) respectively were recruited. The PD patients were assessed for the LA, AC, Gait, PS_{retrop} and Po tasks (UPDRS tasks 3.8, 3.9, 3.10, 3.12 and 3.13, respectively) by two neurologists (N1, N2) expert in movement disorders. The postural stability score (PS_{PIGD}) was assessed by the PIGD subscale score obtained from the AC, Gait, PS_{retrop} and Po UPDRS tasks [47]. Motor impairment is sensitive to the time passed after the last drug intake; therefore, the OFF state (practically defined as that after 12 h without medication) was chosen as the reference for disease severity scoring. PD patients were excluded if they had previous neurosurgical procedures, tremor severity > 1 or cognitive impairment (Mini–Mental State Examination Score < 27/30). PD patients met the following criteria: Hoehn and Yahr average score 2.1 (min 1, max 3); age range 53–80 years (mean 69, std. dev. 7.5). disease duration range 3–10 years (mean 5.8, std. dev. 2.5), gender 8 men and 6 women. The HC subjects performed the same tasks, in the same environmental conditions and with the same system setup of PD patients. The HC cohort was selected trying to approximately match the PD cohort in age and gender, excluding subjects affected by neurological, motor and cognitive disorders.

Informed consent was obtained in accordance with the Declaration of Helsinki (2008). The study's protocol was approved by the Ethics Committee of the Istituto Auxologico Italiano (Protocol n. 2011_09_27_05).

4.2. Experimental Setup

An experimental setup has been built both to assess the accuracy of the system and to acquire relevant clinical and kinematic data useful for the automated assessment of the UPDRS tasks. The kinematic parameters evaluated by the system were compared with those evaluated by an optoelectronic system, considered as gold reference (BTS SMART DX400$^©$, eight TVC, 100–300 fps, BTS Bioengineering, Milan, Italy) [48]. For this experiment, reflective markers were attached to the body of the PD and HC subjects to evaluate kinematics of lower limbs, thorax, spine and head (Figure 4). The biomechanical measurements of the lower body were modelled according to the Helen Hayes Marker set [49,50] and those of the upper body were modelled according to the Plug In Gait model (Vicon$^®$ Motion Systems, Oxford, UK) [51], focusing the attention only on thorax, shoulders and spine. Three additional markers were put on the forehead (M_{HEAD}), on the right (MR_{WRS}) and the left wrist (ML_{WRS}) respectively, this to allow for the assessment of the head posture, CoM estimation and data synchronization between our system and the optoelectronic system. The body markers relevant for the accuracy assessment and their reference positions are presented in Table 1.
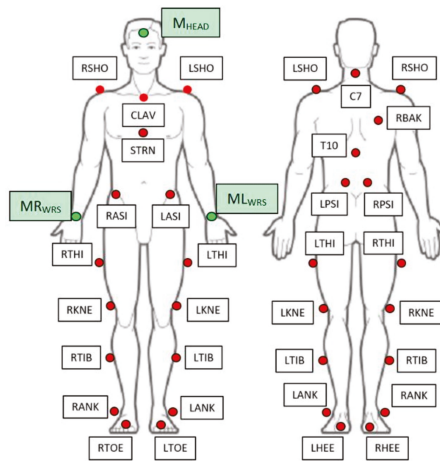
Figure 4. Details of the marker set placement positions.

Table 1. Markers of the optoelectronic system for the accuracy estimation.

Markers	Definitions	Positions Reference
C7	7th Cervical Vertebrae	Spinous process of the 7th cervical vertebrae
LPSI/RPSI	Left/Right PSIS	Placed over Left/Right posterior superior iliac spine
LSHO/RSHO	Left/Right Shoulder	Placed on Left/Right acromioclavicular joint
LASI/RASI	Left/Right ASIS	Placed over Left/Right anterior superior iliac spine
LKNE/RKNE	Left/Right Knee	Placed on lateral epicondyle of the Left/Right knee
LANK/RANK	Left/Right Ankle	Placed on lateral malleolus along an imaginary line that passes through the trans-malleolar axis
M_{HEAD}	Head	Placed on head (additional marker)
ML_{WRS}/MR_{WRS}	Left/Right Wrist	Placed on Left/Right wrist (additional markers)

4.3. Data Acquisition Procedure

During the experimental campaign, the HC and PD subjects were equipped with the set of reflective markers shown in Figure 4, and were instructed to perform the UPDRS tasks. Their performances were supervised and assessed by two neurologists and simultaneously acquired by the two systems. The neurologist' scores, the kinematic parameters and the videos of each task performance were recorded for the subsequent analysis.

The PD and HC subjects performed all the tests facing our system, that is, with the depth axis of the Kinect device perpendicular to the subject frontal plane. The proper position of the subject was verified both by the neurologists and by the system software, which checks the availability of the whole Kinect skeleton and its correct positioning. The device was placed 1.2 m height and at a distance of about 2 m from the subjects. No other object apart the chair, and limited to the AC and LA tasks, was allowed in the working volume. At the beginning of each test, the subject elevated the right arm three times, this to allow the synchronization between the two systems. The synchronization was performed by time-shifting the signals of the arm elevation angles measured by the systems such that their cross-correlation was maximized. The PD and HC subjects were told to perform the LA, AC and Po tasks as indicated by the UPDRS guidelines. Furthermore, the PD subjects performed also the Gait and the PS_{retrop} tasks, in order to evaluate the PIGD sub-score. For the LA task both legs were assessed independently. Two acquisition sessions were planned separated by an interval of thirty minutes to allow subjects to rest.

In the first session a total of five tasks were assessed by the neurologists for the fourteen PD subjects, and a total of three tasks were supervised for the twelve HC subjects. In the second session the same tasks of the first one were repeated in random order. The body sway of the CoM movements was measured during the Po task: in the first phase (indicated as Phase1) each subject was told to stand up straight for ten seconds. Then, during the second phase (indicated as Phase2) each subject was told to try to improve and maintain a more straight posture for other ten seconds: this can be considered a sort of secondary motor task, that potentially can highlight differences between PD and HC subjects [38].

4.4. Movement Characterization by Kinematic Parameters

The analysis and the related characterization of the considered UPDRS tasks make use of kinematic parameters which are mainly estimated from angles between pairs of body segments, involving femur, knee, tibia, spine and head. The body segments are defined by their distal and proximal points, which in our system are assumed to correspond to the joints of the skeleton model of Figure 2. The centroid of each segment is calculated as the midpoint between the proximal and distal extremities. The postural stability is assessed by the body CoM, which is estimated by the weighted average of the body segment centroids. In particular, the kinematic characterization of the LA, AC and Po tasks is based on the evaluation of the angles ANG_{KNEE} and ANG_{TRUNK}. Only for the Po task two further angles are considered: the forward $ANG_{FORHEAD}$ and lateral $ANG_{LATHEAD}$ bending angles of the head respect to spine direction. Specifically, with reference to Figure 5 for the proximal and distal 3D points relevant for the analysis, and to Figure 2a for the 3D skeleton joints involved, we considered:

- For the LA task (Figure 5a): the knee angle ANG_{KNEE} between the A-B and B-C segments, with A = HipR, B = KneeR and C = AnkleR for the right limb, and A = HipL, B = KneeL and C = AnkleL, for the left limb;

- For the AC task (Figure 5b): the knee angle ANG_{KNEE} defined above; the trunk angle ANG_{TRUNK} between the D-E segment and the vertical direction $\hat{n}$ (i.e., the red arrow), with D = SpineS and E = SpineB;

- For the Po task (Figure 5c): the knee angle ANG_{KNEE} and the trunk angle ANG_{TRUNK} defined above. The bending of the head respect to spine is evaluated by the angles $ANG_{FORHEAD}$ and $ANG_{LATHEAD}$, projections of the angle between the SpineS-SpineB (D-E) segment and the SpineS-Head segment on the sagittal and frontal body planes, respectively. The lateral body plane is approximately identified by the plane containing the segments ShouldR-ShouldL and SpineB-SpineS, while the sagittal body plane is perpendicular and contains the SpineB-SpineS segment. Note that while ANG_{TRUNK} has components in the sagittal and lateral body planes, $ANG_{FORHEAD}$ and $ANG_{LATHEAD}$ have components only in the sagittal and in the lateral planes, respectively.

(a)   (b)   (c)
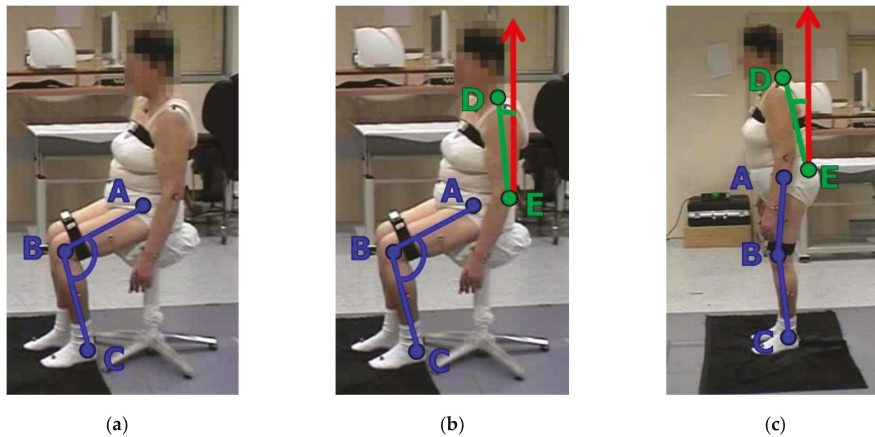
Figure 5. Segments involved in the estimation of the angular measures during lower limbs and postural tasks: (**a**) LA task (**b**) AC task (**c**) Po task. Note that the depth axis of the Kinect device is perpendicular to the subject frontal plane in all the tests, see Section 4.3).

The CoM is estimated during the Phase1 and the Phase2 of the Po task, both to evaluate the postural instability and to evidence the effects of the secondary tasks. A subject specific quasi-static center of mass $\mathbf{C_b}$ is evaluated by applying the kinematic method described in [52,53]. As indicated in Equation 1, $\mathbf{C_b}$ is obtained by the weighted average of the body segment centroids ($\mathbf{C_i}$), evaluated from the skeleton model, where the weights w_i are provided by standard body segment densities obtained from anthropometric data [54]:

$$\mathbf{C_b} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{C_i} * w_i \qquad (1)$$

The centroids $\mathbf{C_i}$ of the following segments made by pairs of skeleton joints are considered (Figure 2a): Head-SpineS, ShouldR-WristR, ShouldL-WristL, SpineS-SpineB, HipR-AnkleR, and HipL-AnkleL. Please note that $\mathbf{C_b}$ is a 3D point; but here only the transverse (or horizontal) plane components are evaluated for the analysis of the body sway. Concerning the evaluation of the kinematic parameters, the skeleton joints provided every 1/30 s by the Kinect SDK allow the estimation of the relevant parameters at the same rate. In particular, the angles ANG_{KNEE}, ANG_{TRUNK}, $ANG_{FORHEAD}$ and $ANG_{LATHEAD}$ were evaluated from the inner products of the pairs of unity vectors representing to the body segments involved. The vertical direction $\hat{n}$, used to evaluated ANG_{TRUNK}, was estimated by the normal to the floor plane. The 3D orientation of the plane was obtained by segmentation of the Kinect depth map using a RANSAC approach [55], with the upside direction of the Kinect skeleton and the feet location as priors. The angle signals were resampled both to remove the typical jitter of the Kinect sampling frequency, and to fit the sampling frequency of the optoelectronic system (100 Hz). The signals are filtered to reduce noise by a second order low-pass Butterworth filter with a cut-off frequency of 10Hz. Most of the significant kinematic parameters presented in the Results and used as input to the classifiers were obtained by standard signal processing algorithms applied to the sampled signals of the ANG_{KNEE}, ANG_{TRUNK}, $ANG_{FORHEAD}$ and $ANG_{LATHEAD}$ angles. The velocity parameters were evaluated as the derivatives of the spline approximations to the angle signals obtained through Matlab functions (unmkpp, mkpp and ppval). Specifically:

- For the LA task, the ANG_{KNEE} signal is segmented in a sequence of flexion/extension movements (cycles) by finding all the minimum-maximum-minimum sequences in the amplitude of the angle signal. The peak to peak amplitude, the speed and the duration of every flexion/extension movement of the leg is evaluated. Specifically, MKAm is the mean of the peak to peak amplitude

maxima and MKAv is its standard deviation; TDm is the mean of the cycle durations and TDv is its standard deviation; SPm is the mean of the speed maxima. Finally PM is the number of "poor movements", defined as the cycles whose amplitude and duration are both less than 25% of the MKAm and the TDm values. This last parameter tries to catch hesitations and very small amplitude cycles in a sequence of almost relevant cycles.

- For the AC task, the ANG_{TRUNK} signal is segmented in a sequence of forward/ backward bending movements (cycles) by finding all the minimum-maximum-minimum sequences in the amplitude of the angle signal. Usually, only one peak is present, but hesitation during the movement or some instability event can generate other peaks in the signal. These peaks are clinically relevant and, consequently, have been considered in the assessment through the parameter NPeaks. MBA is the maximum angular peak and TD is the duration of the main bending cycle containing MBA, while SPm is the mean speed during the AC movement.

- For the Po task, during the quasi-static Phase1 are evaluated the bending angle FTB of the spine respect to the vertical (mean of ANG_{TRUNK}), the forward bending angle FHB ($ANG_{FORHEAD}$) and the lateral bending angle LHB ($ANG_{LATHEAD}$) of the head. During the Phase2, the variations of these angles (FTB_Δ, FHB_Δ, LHB_Δ) respect to Phase1 are evaluated.

- For the PS_{COM} task, the sway is defined as the CoM component in the transversal plane (perpendicular to both the lateral and the sagittal body planes). In this plane, the Antero-Posterior (AP) and Medio-Lateral (ML) axis are defined as the intersection of sagittal and lateral planes with the transversal plane, respectively. The AP components of the range, total path length and velocity of the sway (APr, APt and APv, respectively), and the ML components of the range, total path length and velocity of the sway (MLr, MLt and MLv, respectively) are evaluated. Furthermore, the sway area SwayArea (convex hull of the sway path) is also evaluated.

The Pearson's correlation between the measures of the kinematic parameters provided by our system and those measured by the optoelectronic system was used to assess the body tracking accuracy. Because all the kinematic parameters for LA, AC, Po and PS_{COM} were obtained from the ANG_{KNEE}, ANG_{TRUNK}, $ANG_{FORHEAD}$ and $ANG_{LATHEAD}$ angles and from the CoM components in the transversal plane, only these last "essential" parameters were considered for the accuracy assessment.

The correspondences between optoelectronic markers (Figure 4 and Table 1) and Kinect joints (Figure 2a) we adopted for the comparison are shown in Table 2. The optoelectronic angular parameters corresponding to the essential ones were obtained by the marker correspondences of Table 2 and by the same procedure we used for the Kinect joints in Section 4.4. The CoM measured by the optoelectronic system was evaluated according to [51].

Table 2. Correspondences between body segments for Kinect and optoelectronic systems.

Parameter	Kinect Segments	Optoelectronic Segments
ANG_{KNEE} Left/Right	HipL/HipR-KneeL/R AnkleL/R-KneeL/R	LASI/RASI-LKNE/RKNE LANK/RANK-LKNE/RKNE
ANG_{TRUNK}	SpineS-SpineB	C7-MeanPSI [a]
$ANG_{FORHEAD}$	Head-SpineS	M_{HEAD}-C7
$ANG_{LATHEAD}$	Head-SpineS	M_{HEAD}-C7
CoM	Head-SpineS ShouldR-WristR ShouldL-WristL SpineS-SpineB HipR-AnkleR HipL-AnkleL	M_{HEAD}-C7 RSHO-MR_{WRS} LSHO-ML_{WRS} C7-MeanPSI [a] RASI-RANK LASI-LANK

[a] MeanPSI = (LPSI + RPSI)/2.

4.5. Discriminant Parameter Selection

The choice of the best parameters used to train the classifiers was performed by selecting the sets of kinematic parameters which best correlate with the UPDRS scores of subjects' performances.

The initial sets of parameters considered to characterize every single task consisted of more than ten parameters per set: they were chosen to be closely related to those features that are implicitly considered by neurologists to assess the motor performance. These initial sets could potentially include irrelevant and redundant parameters, which could hide the effects of the clinically relevant ones, reducing the predictive power of the classifiers used for the automated assessments. To avoid this, a feature selection (FS) procedure [56] is performed by the Elastic Net (EN) algorithm [57]. EN is a hybrid of Ridge regression and LASSO regularization. EN encourages a grouping effect on correlated parameters, and tends to be more conservative respect to LASSO or Ridge regression in removing correlated parameters, a process which can select incorrect data model. This capability is important when dealing with those features which are similar and tend to be moderately correlated. The EN implementation is based on Matlab scripts (lasso Matlab function). To avoid biasing the results by the different scaling, the PD parameters $p_{i\,PD}$ have been normalized (Equation (2)) by the corresponding average values of the HC parameters $p_{i\,HC}$. Then the normalized parameters range from the value 1 ($p_{i\,HC}$) to a maximum ($p_{i\,PD\,Norm\,MAX} > 1$), or to a minimum ($0 < p_{i\,PD\,Norm\,MIN} < 1$), depending if the value of the specific parameter increases or decreases when the severity of the impairment increases. The parameter Number of Poor movements (PM), whose minimum value is 0, was not normalized:

$$p_{i\,PD\,Norm} = p_{i\,PD}/p_{i\,HC}, \tag{2}$$

4.6. Statistical Analysis

Descriptive statistical analysis of the collected data by Mann-Whitney U and χ^2 tests did not show significant differences among PD and HC (age, gender, cognitive status), then the data were safely pooled into two groups (PD and HC) for the following analyses.

We note that, as common in feature selection algorithms, the previous selection of parameters based on the EN algorithm assumes the UPDRS scores are ratio data type, while they actually are ordinal data. Then, to confirm the relevance of these parameters in the context of the ordinal nature of the scores and to deal with their non-normal distributions, Spearman non-parametric rank correlation at a significant level $p < 0.05$ was applied. Only those parameters showing a Spearman's correlation coefficient ρ greater than 0.3 (as absolute value) with respect to the UPDRS scores assigned to the LA, AC and Po task performances were considered for the final sets. For the PS_{COM} task, the CoM parameters were correlated with the PIGD subscale scores (PS_{PIGD}).

Furthermore, as a support to the effectiveness of the selected parameters in the automated assessment, the statistical significance of each parameter in discriminating PD and HC was considered and verified by the Mann-Whitney U test with $p < 0.05$. All statistical analyses were performed using Matlab. For the correct application of the test, only the data of the second acquisition session were considered.

4.7. Supervised Classifier Training

Three different types of supervised classifiers have been considered for the automatic assessment of the LA, AC, Po and PS_{COM} tasks: k-Nearest Neighbours (kNN), Multinomial Logistic Regression (MLR) and Support Vector Machine (SVM) with polynomial kernel [58]. Two types of classification problems were considered: first, a binary classification problem, where the subjects are classified into the HC and the PD classes; second, a multiclass classification problem, where the subjects are classified into the three PD classes of increasing severity. The design of this second experiment was suggested by the distributions of the severity scores of the PD patients recruited for the study, which were essentially distributed among slight, mild and moderate UPDRS severity scores [2], corresponding to the UPDRS1, UPDRS2 and UPDRS3 classes, respectively. Furthermore, the severity scores distributions

were adequately balanced among the classes for all the tasks (Table 3). The classifiers were trained for each task (LA, AC, Po, and PS$_{COM}$) by using as input the sets of "selected kinematic parameter vector—UPDRS score" pairs obtained from the reference dataset of performances. In particular, the PS$_{COM}$ classifiers was trained using the PIGD subscale scores (PS$_{PIGD}$) as UPDRS score.

Table 3. Distribution of the severity scores among the UPDRS tasks.

UPDRS Task	UPDRS Severity Scores		
	UPDRS1 (Slight)	UPDRS2 (Mild)	UPDRS3 (Moderate)
LA[a]	16	22	18
AC	12	11	5
Gait	12	8	8
PS$_{retrop}$	8	6	14
Po	14	8	6

[a] In the LA task, both legs were assessed.

The input data have been normalized both to make more stable the training procedure and to simplify the behavior of the parameters in the parameter space. Specifically, the parameters $p_{i\,PD\,Norm}$ whose values increases with the worsening of the performance from 1 to their maximum ($p_{i\,PD\,Norm\,MAX} > 1$), are scaled in the range 0 to 1, while those whose values decrease with the worsening of the performance from 1 to the their minimum ($0 < p_{i\,PD\,Norm\,MIN} < 1$) are first reversed and then scaled in the range 0 to 1. The score values are scaled in the range 0 to 1 as well.

The kNN classifiers were employed as baseline and implemented and tested by Matlab scripts (fitcknn function). The classifiers were tested with parameter k = 1,3,5,7 using the Euclidean distance metric. The tie breaking algorithm adopted was to decrease k by 1 until the tie is broken.

The MLR classifiers were implemented and tested by Matlab scripts (fitmnr function for ordinal data with probit link function).

The SVM classifiers were implemented and tested by Matlab scripts with the support of the LibSVM library package [59]. The kernel function of SVM classifier is polynomial with parameters: γ (gamma), r (bias) and d (polynomial degree) and C (cost). Every SVM multiclass classifier uses the one-versus-one coding design with majority voting scheme and is made by three binary SVM models, all with the same parameters [60]. A grid-search and cross-validation method were used to find the optimal values of the SVM parameter C, γ, r and d for the three binary classifiers.

4.8. System Reliability and Accuracy Evaluation

A commonly accepted measure of reliability in the context of clinical assessments is the Intra Class Correlation coefficient (ICC). Accordingly, the reliability of the system assessments respect to the neurologist ones was evaluated by the Intra Class Correlation coefficient ICC$_{N12-SY}$ (two-way random effects model with an absolute agreement) [3]. The inter-rater agreement ICC$_{N12}$ between the two neurologists was evaluated and compared as a baseline with the inter-rater agreement ICCN$_{12-SY}$ among neurologists and system, considering the system as a third "virtual" neurologist.

In the evaluation of ICC$_{N12}$, the scores of the neurologists for the LA, AC, Po tasks and for the subscale PS$_{PIGD}$ were considered, while for ICCN$_{12-SY}$ both the neurologist scores and the corresponding system scores were used. Concerning the reliability of the remote video-based assessments, motor examination of video recorded UPDRS tasks has already been demonstrated to be a sufficiently accurate alternative to in field ones [61]. In machine learning context, it is more common to assess the reliability of classifiers by their accuracy. Then, we evaluated also this measure of system performance considering the mean accuracies of each classifier, both in discriminating between PD from HC subjects (binary classification problem) and in classifying PD subjects into different severity classes (multi-classes classification problem) [62].

5. Results

5.1. Clinical Assessment Results

After collecting the clinical assessments at the end of the experiment, none of the performances of the PD cohort were scored with normal (score 0) or severe impairment (score 4) for all the UPDRS tasks considered. The distributions of the severity scores assigned to the PD patients among slight, mild and moderate responses were relatively balanced for all the tasks (Table 3).

5.2. Accuracy of the Kinematic Parameter Evaluation

The measurement accuracy of Microsoft Kinect v2 in clinical estimation of motor functions [27,28,30] and body CoM [18,19,25,27,29] has been previously assessed: this was confirmed also by our experiment. The comparison of the parameter measurement respect to the gold reference system cannot be performed directly because the Kinect skeleton model and the optoelectronic marker set have different body landmark positions. Furthermore, for every parameter, we want to estimate an average accuracy based on all the trials per task acquired.

Then, for every essential parameter i, characterizing one or more tasks, the parameter samples of each associated task trials were joined together into a single parameter sample sequence (PSS_i). The PSS_i sequences measured by the two systems were then compared by evaluating the Pearson correlation coefficient r_i. In Table 4, the r_i coefficients indicate a significant correlation that ranges from good to strong for all the examined parameters. In Figures 6 and 7 are shown two examples of the ANG_{KNEE} and the ANG_{TRUNK} variations for a LA and an AC task trial. In Figure 6, it is interesting to point out that the last movement is characterized by a significant reduction in amplitude and duration. In Figure 7, the presence of a secondary peak indicates that the PD subject had an instability event at the corresponding time. These anomalies are hardly identified by neurologists.

Table 4. Mean and standard deviation of the Pearson's correlation coefficients for essential parameters estimated by the two systems.

Parameter	Pearson's Correlation Coefficient r [a]	
	Mean $\pm$ Std. Dev.	p-Value[a]
ANG_{KNEE}	0.94 ± 0.07	9.09×10^{-3}
ANG_{TRUNK}	0.87 ± 0.10	6.72×10^{-3}
$ANG_{FORHEAD}$	0.73 ± 0.20	$3.98e \times 10^{-2}$
$ANG_{LATHEAD}$	0.71 ± 0.23	3.57×10^{-2}
CoM_{AP}	0.84 ± 0.11	3.18×10^{-3}
CoM_{ML}	0.90 ± 0.09	8.94×10^{-3}

[a] Significance level $p < 0.05$.
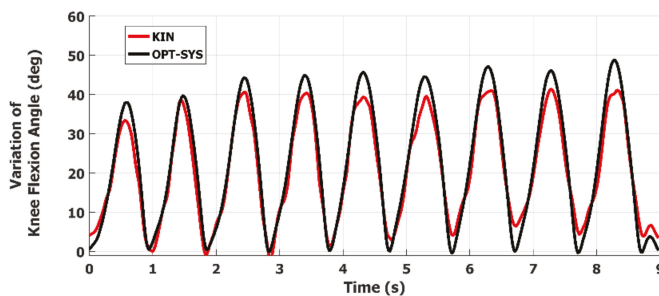
Figure 6. Example of the ANG_{KNEE} variations during the LA task performance of a PD subject: the last movement at 8.9 s is characterized by significant reduction in both amplitude and duration.
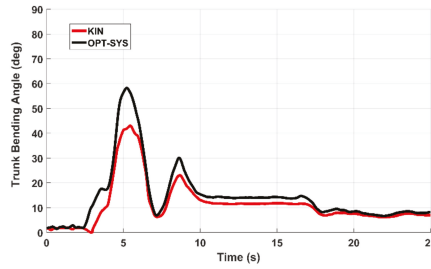
Figure 7. Example of the ANG$_{TRUNK}$ variations during the AC task performance: the secondary peak at 8.5 s indicates the presence of an instability event in the final standing stance.

5.3. Discriminant Kinematic Parameter Selection and Validation

The Spearman correlation values between selected parameters and UPDRS scores, and the Mann-Whitney U test values concerning their significance in discriminating PD and HC subjects are shown in Tables 5–8 for the LA, AC, Po and PS$_{COM}$ tasks, respectively. We remark that, for PS$_{COM}$ task, the Spearman correlation was evaluated respect to the PS$_{PIGD}$ subscale scores. For the correct application of the U test, only the data of the second acquisition session were considered. The results in the tables show that all the selected parameters correlate with UDRS score ($|\rho > 0.3|$, $p < 0.05$). Furthermore, they are all significant for Mann-Whitney test ($p < 0.05$), even though at different significance levels. The mean values of the selected parameters respect to the UPDRS severity classes are shown in the radar graphs of Figure 8 for all the tasks.

Table 5. Parameters of the LA task: discriminant power and correlation with UPDRS scores.

		Mann-Whitney U Test				Spearman Coefficient	
Name	Meaning (Unit)	Median HC	Median PD	Z	*p*-Value [a]	ρ	*p*-Value [a]
MKAm	Mean of Maximum Knee Angle (degree)	32.41	25.02	1.93	5.37×10^{-2}	−0.72	9.99×10^{-6}
MKAv	Var. of Maximum Knee Angle (-)	0.07	0.13	1.81	7.03×10^{-2}	0.49	6.72×10^{-3}
TDm	Mean of movement Total Duration (s)	0.26	0.42	2.88	3.95×10^{-3}	0.43	1.98×10^{-2}
TDv	Var. of movement Total Duration (-)	0.10	0.12	1.68	9.19×10^{-2}	0.43	2.07×10^{-2}
SPm	Mean Speed of movement (degree/s)	114.8	64.20	3.00	2.66×10^{-3}	−0.84	8.18×10^{-9}
PM	Num. of poor movements (#)	0.00	1.00	1.99	4.69×10^{-2}	0.74	3.94×10^{-6}

[a] Significance level $p < 0.05$.

Table 6. Parameters of the AC task: discriminant power and correlation with UPDRS scores.

		Mann-Whitney U Test				Spearman Coefficient	
Name	Meaning (Unit)	Median HC	Median PD	Z	*p*-Value [a]	ρ	*p*-Value [a]
MBA	Maximum Bending Angle (degree)	17.50	31.26	3.18	1.44×10^{-3}	0.75	4.00×10^{-7}
TD	Total Duration of S2S movement (s)	0.90	2.42	2.86	4.17×10^{-3}	0.80	1.08×10^{-8}
SPm	Mean Speed of S2S movement (degree/s)	21.85	12.92	2.76	5.84×10^{-3}	-0.69	6.26×10^{-6}
NPeaks	Number of Bending Peaks (#)	1.00	1.00	1.13	2.59×10^{-1}	0.63	5.65×10^{-5}

[a] Significance level $p < 0.05$.

Table 7. Parameters of the Po task: discriminant power and correlation with UPDRS scores.

		Mann-Whitney U Test				Spearman Coefficient	
Name	Meaning (Unit)	Median HC	Median PD	Z	*p*-Value [a]	ρ	*p*-Value [a]
FTB	Forward Trunk Bending (degree)	0.38	−5.69	2.71	9.88×10^{-4}	−0.70	1.36×10^{-4}
FTB$_\Delta$	Var. of Forward Trunk Bending (degree)	0.35	0.27	0.18	8.55×10^{-1}	0.43	5.54×10^{-2}
FHB	Forward Head Bending (degree)	−1.83	−6.86	1.92	5.23×10^{-2}	−0.78	5.90×10^{-6}
FHB$_\Delta$	Var. of Forward Head Bending (degree)	0.46	0.53	0.22	8.17×10^{-1}	0.27	3.62×10^{-1}
LHB	Absolute Lateral Head Bending (degree)	2.05	3.02	0.53	6.07×10^{-1}	0.59	2.39×10^{-3}
LHB$_\Delta$	Var. of Lateral Head Bending (degree)	0.19	0.43	1.53	1.25×10^{-1}	0.43	6.54×10^{-2}

[a] Significance level $p < 0.05$.

Table 8. Parameters of the PS$_{COM}$ task: discriminant power and correlation with UPDRS scores.

Name	Meaning (Unit)	Mann-Whitney U Test				Spearman Coefficient [b]	
		Median HC	Median PD	Z	*p*-Value [a]	ρ	*p*-Value [a]
APr	CoM AP sway Range (cm)	0.59	1.13	1.80	7.20×10^{-2}	0.59	3.24×10^{-3}
APt	CoM AP sway Total (cm)	1.49	3.28	2.23	2.50×10^{-2}	0.65	2.54×10^{-2}
MLt	CoM ML sway Total (cm)	0.98	3.48	2.24	2.53×10^{-2}	0.48	1.88×10^{-2}
APv	CoM AP sway Velocity (cm/s)	0.72	1.32	1.86	6.34×10^{-2}	0.56	4.92×10^{-2}
MLv	CoM ML sway Velocity (cm/s)	0.48	1.49	2.24	2.53×10^{-2}	0.42	4.25×10^{-2}
SwayArea	CoM Sway Area (cm²)	0.30	0.85	1.58	1.13×10^{-1}	0.59	2.92×10^{-3}

[a] Significance level $p < 0.05$; [b] The Spearman correlation was evaluated respect to the PS$_{PIGD}$ subscale scores.
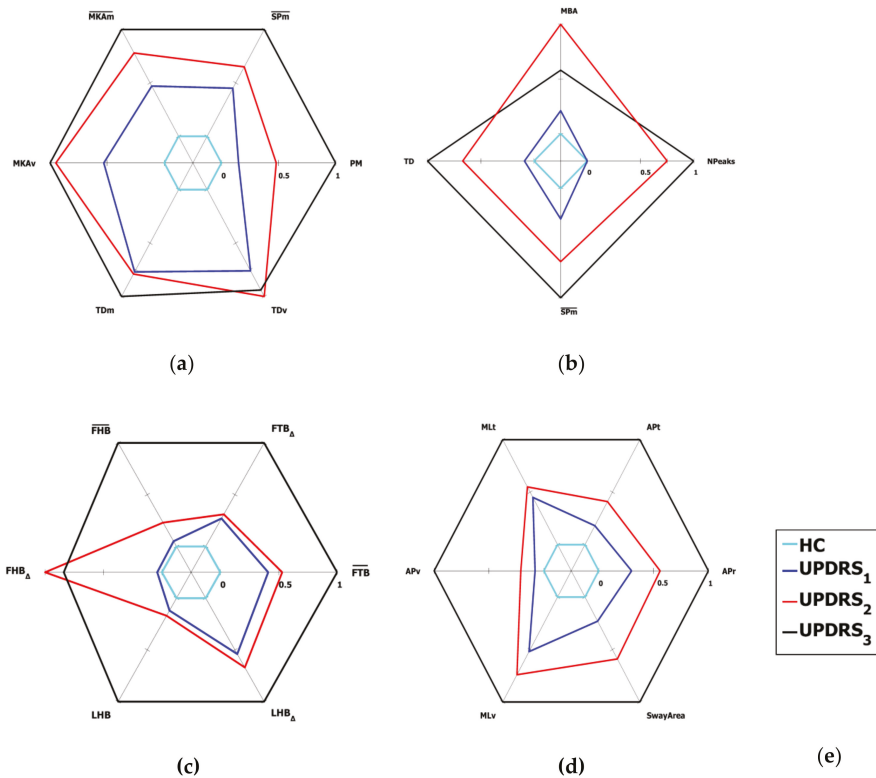
Figure 8. Radar graphs of the mean values of the normalized kinematic parameters of HC and UPDRS severity classes for the lower limbs and postural tasks: (**a**) Leg Agility (LA); (**b**) Arising from Chair (AC); (**c**) Posture (Po); (**d**) Postural Instability (PS$_{COM}$); (**e**) Legend for the radar plots. See Section 5.3 for further details of the graph representation.

The parameters have been represented such that an increasing values indicate a worsening of the performance, highlighted by a corresponding expansion of the related graph. For this reason, the parameters of Tables 5–8 are represented in Figure 8 directly (with the original parameter name) or inversely (with an overscore on the original parameter name), depending if the parameter value increases or decreases when the severity of the impairment increases.

Furthermore, with reference to Section 4.5, the parameters are scaled in such a way that the parameter values corresponding to the best performance ($p_{i\,PD\,Norm} = p_{i\,HC}$) are represented on

the innermost circle (i.e., value = 0) and those corresponding to the worst one ($p_{i\,\text{PD Norm MAX}}$, or $1/p_{i\,\text{PD Norm MIN}}$, depending on the parameter) are represented on the outermost circle (i.e., value = 1).

Finally, it should be noted that almost all the parameters are able to discriminate the different UPDRS classes for the LA, AC, Po and PS_{COM} tasks, pointing out the increasing severity of motor impairment by the corresponding increasing of their values. The graphs are encapsulated and do not overlap, which means that a monotonic increase of the parameter value corresponds to an increase (and so a worsening) of the UPDRS score.

The Pearson correlation analysis of the CoM movements, as measured by our system and by the optoelectronic system, shows that they are correlated both in the Antero-Posterior (AP) and in Medio-Lateral (ML) components (Table 4). These values confirm the feasibility of Kinect in the accurate estimation of center of mass movements. In Figure 9a, an example of CoM trajectories as measured at the same time by the two systems is shown; the trajectory of center of mass resembles the gold reference one, even if a scale factor is present. Figure 9b shows an example of the two phases of PS_{COM} task: in particular, the CoM trajectory measured by the optoelectronic system while a PD subject is performing the Phase1 (solid cyan line) and the Phase2 (solid red line) respectively. In Figure 9c, the same movement as measured by our system is shown.
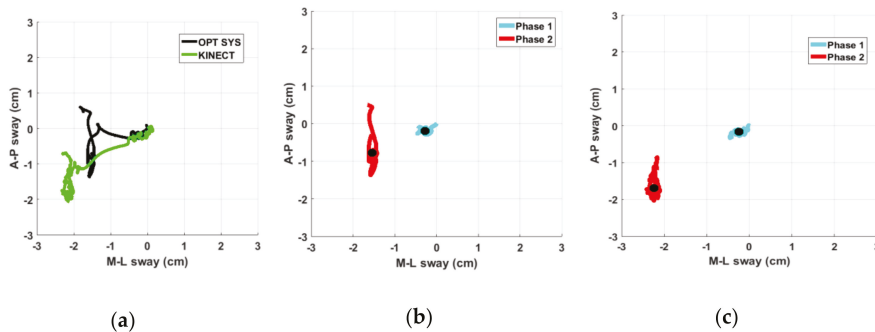
Figure 9. (**a**) Example of CoM trajectory of a PD subject represented in the Antero-Posterior (AP) and Medio-Lateral (ML) components during the Po task, as measured by our system (green line) and by optoelectronic system (black line); (**b**) Details of the trajectories during the first (cyan line) and second phase (red line) of PS_{COM} task with the respective centroids (black dots) as measured by optoelectronic system; and (**c**) as measured at the same time by our system.

In both figures, the secondary motor task (during which the PD subject is trying to improve and then maintain a straighter posture) clearly increases the body sway along the AP direction, supporting the hypothesis of a performance degradation for PD subjects respect to HC in this context. The shapes of trajectories are quite similar: this confirms the feasibility of our system in acquiring the body CoM in agreement with the gold standard. Again, there is a mild scaling and an offset between the centroids of the trajectories measured by the two systems: this is probably due to the different landmark positions of the body skeleton models considered and to the different algorithms used to estimate the CoM position. Nevertheless, we remark that the CoM parameters we chose are independent from these biases. Furthermore, they convey useful information which well correlates with clinical evaluations, discriminating between PD from HC subjects, as indicated in Table 9. This is evident for almost all the PD subjects, on AP and/or ML directions; on the contrary, this is negligible for HC subjects, as confirmed by the values in the second and third column of Table 9. Furthermore, the differences of the CoM parameters (Phase2 respect to Phase1) between PD and HC subjects are significant at level $p < 0.05$, both for the U test and for the T test (column 5 and 6, Table 9).

Table 9. Average differences of CoM parameters between Phase1 and Phase2 of the Po task for HC and PD subjects.

Name	PD Subjects Phase2-Phase1	HC Subjects Phase2-Phase1	Mann-Whitney U Test Z	Mann-Whitney U Test *p*-Value [a]	*T* Test *p*-Value [a]
APt	1.61	0.54	2.04	4.17×10^{-2}	2.22×10^{-3}
MLt	1.87	0.27	2.04	4.17×10^{-2}	6.33×10^{-3}
APv	1.14	0.52	2.11	3.44×10^{-2}	9.86×10^{-4}
MLv	1.04	0.45	1.97	4.89×10^{-2}	1.11×10^{-3}
SwayArea	1.88	0.25	2.13	3.28×10^{-2}	7.19×10^{-3}

[a] Significance level $p < 0.05$.

5.4. Reliability of the Assessments of the System and the Neurologists

The values for the inter-rater agreement between the neurologists N1 and N2 (ICC_{N12}) and among the neurologists and the system ($ICC_{N12\text{-}SY}$) are shown in Table 10 (ρ values and 95% confidence intervals). According to [3], the ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively. In the evaluation of the $ICC_{N12\text{-}SY}$ for PS_{PIGD}, we use the system evaluated PS_{COM}, stressing both the interpretation of PS_{COM} as a posture stability score and its good correlation with PS_{PIGD}.

Table 10. Intra Class Correlations for the system and the neurologists assessment reliability.

Reliability/ Task	LA	AC	Po	PS_{PIGD}
ICC_{N12} [a]	0.80	0.82	0.77	0.73
$ICC_{N12\text{-}SY}$ [a]	0.77	0.80	0.74	0.65

[a] Significance level: $p < 0.05$.

The $ICCN_{12}$ values, as reported in Table 10, indicate a generally good agreement between the two neurologists, with differences per task compatible with literature results. The results for $ICCN_{12\text{-}SY}$ show the system does not degrade significantly the inter-rater agreement between the neurologists, except for PS_{PIGD}. This could be due to the limited number of subjects examined, or to the use of CoM parameters not completely superimposable to PIGD subscale assessments. In fact, CoM parameters are evaluated only in PS_{COM} task, e.g., during postural adjustments from the quiet stance and not in other more challenging dynamic domains of postural stability characterization [37].

5.5. Accuracies of the Supervised Classifiers

The classification accuracies of the LA, AC, Po and PS_{COM} tasks are shown in Table 11. They are obtained applying the leave-one-out and 10-fold cross validation method for the MLR, SVM and kNN classifiers. Accuracies refer to two different classification goals per task: discriminating PD from HC (two-classes classifier, binary problem) and classifying PD subjects into three UPDRS severity classes (three-classes classifier, multiclass problem).

The kNN with k = 3 and SVM with polynomial degree d = 2 gave the best performance using the leave-one-out cross validation, then these values were chosen for the system classifiers. In general, the accuracies of SVM classifiers are better than the kNN and MLR ones. Furthermore, the results of binary classification problem, in classifying HC and PD subjects, are quite better than the multiclass classification ones. This behavior was not unexpected because, in general, the classifiers perform worse on the same training data when the number of classification labels (i.e., classes) increases. In reporting the multiclass classification accuracy is more appropriate to indicate the per-class accuracy (fourth column for leave-one-out and sixth column for 10-fold of the Table 11), where the classification accuracies are averaged over the classes [62]. The absolute classification error (e_c) was defined as the difference between the UPDRS class C, assigned by the neurologists, and the estimated class C' assigned by the system to each motor performance i ($e_c = |Ci - C'i|$). The e_c value for the kNN and

MLR classifiers is sometimes larger than 1 UPDRS class, even when their average accuracies are better than that of the SVM classifiers. On the contrary, the e_c value for the best SVM classifiers was never greater than 1 UPDRS class for all the tasks; this means that the automatic assessments are always close to the neurologist' ones. This is also an important feature for the system reliability respect to an average greater agreement but with large spot disagreements.

Table 11. Classification accuracies for the supervised classifiers.

Task	Classifier	LEAVE-ONE-OUT		K-FOLD (10) [a]	
		HC-PD (2-Classes)	UPDRS (3-Classes)	HC-PD (2-Classes)	UPDRS (3-Classes)
LA	SVM	95.6	68.9	96.5	73.6
	KNN (k = 3)	94.5	51.7	96.5	58.0
	MLR	89.6	68.9	89.6	70.5
AC	SVM	88.2	66.3	88.2	69.9
	KNN (k = 3)	86.0	60.0	88.2	67.5
	MLR	94.1	70.5	96.8	73.3
Po	SVM	91.6	68.0	93.5	68.2
	KNN (k = 3)	95.8	70.8	95.0	68.9
	MLR	83.3	62.5	81.7	58.8
PS$_{COM}$	SVM	95.2	58.3	93.2	59.6
	KNN (k = 3)	92.8	41.6	95.7	45.8
	MLR	95.8	50.0	91.9	52.1

[a] 100 iterations.

In addition, the results in Table 11 show that the two-classes accuracy is higher for LA and PS$_{COM}$, while is slightly lower for the other two tasks. This is in agreement with the Figure 8, in which the AC and Po graphs show more overlapping between UPDRS classes as compared to the LA and PS$_{COM}$ ones. The partial incoherence of some parameters in separating the different classes has probably an impact on the classifier performance. The behavior of the two-classes classification accuracy is not repeated in case of the three-classes classification, for which the worst performance is obtained for PS$_{COM}$ task. This could be due again to CoM parameters not directly comparable to PIGD subscale assessments. Looking at the error distribution, we obtain a big contribution from UPDRS 3 class (i.e., most impaired PD subjects). The limited number of observations assessed as UPDRS 3 suggests that some significant parameters, which should have been considered, are probably missed, and could be included among the selected features only by increasing the number of UPDRS 3 observations in the training set.

6. Discussion

The availability of low-cost home-based solutions for the reliable and automated assessment of motor symptoms in Parkinson's disease is highly desirable since it could provide several advantages, among which: reduction of costs and patient discomfort; better and prompt supervising and adjustment of the therapy; healthcare analytics for patient care improvements. Surely, among the features that these solutions should exhibit, particularly important are: a non-invasive approach to the assessment; a user-friendly interaction suitable to motor impaired users; an objective, continuous and automated evaluation of patient status, strongly correlated with the standard clinical assessments; an improvement of the reliability respect to the typical intra and inter-rater variability of the clinical evaluations.

In this paper, a self-managed system for the automated assessment of Parkinson's disease which tries to implement many of the aforementioned features is presented. The developed system is focused on posture instability and motor impairments of lower limbs and it is one of the elements of a larger project aimed to bring an overall automated assessment of UPDRS tasks at home [35].

As a first step, we addressed both the non-invasiveness and user-friendly interaction by a low-cost system based on an RGB-D optical device, then providing a gesture based human computer interface for the self-management of the assessment procedures. The usability of the interface was tested and verified by PD users during a campaign of data acquisition sessions. Then, the accuracy of the kinematic measures, as obtained by the system, was validated successfully by comparison with a gold

standard equipment (i.e., an optoelectronic system). This was a necessary preliminary requirement, since an objective evaluation of the patient status is based on the strong correlation existing between motor impairments and kinematic parameters extracted from patient's movements.

To reliably refer the system assessments to the clinical ones, the analysis of possible movements was constrained to those specified by the UPDRS tasks. An experimental protocol was designed in which PD patients and healthy controls were assessed at the same time both by neurologists and by the system during the execution of the specific standard tasks defined by UPDRS. A feature selection procedure yielded to sets of optimal parameters, both correlated to UPDRS clinical scores and statistically significant in discriminating PD subjects from healthy controls. As shown in Figure 8, not all these parameters have the same discriminant power to separate subjects among the different PD severity classes; this is true especially for the AC and Po tasks. This is probably due to the limited number of PD subjects examined: consequently, further experiments could improve the current results.

Following related works based on wearable systems [40,41], the postural stability of PD subjects was characterized by CoM movements. We analyzed the CoM trajectories during the two phases of the Po task (named PS_{COM}), assuming the Phase2 as a mild secondary motor task [38]. As in [40,41], large differences in CoM trajectories of PD respect to HC were found. Differently to [41], a good correlation between PS_{COM} parameters and the standard postural stability test (PIGD) was observed. This result can be explained because of the different physical quantity and derived parameters considered by the two approaches: CoM displacements in our case, derivative of CoM accelerations in [41]. On the other hand, the CoM parameters in [40] have a closer physical relationship with ours: respect to us, the authors did not find a significant correlation between the PIGD scores and the parameters they selected, but this could probably be due to the exclusion of the retropulsion task from their analysis. In conclusion, we found that the PS_{COM} parameters are related to PIGD score and are also statistically significant: in fact, they clearly discriminate PD subjects from healthy controls, supporting the initial hypothesis of a worsening of PD stability during the execution of secondary tasks.

The automated assessment of UPDRS tasks is performed by means of kNN, MLR and SVM supervised classifiers, trained on the sets of selected parameters and the corresponding UPDRS scores from reference datasets of performances of PD and HC cohorts. In general, the accuracy of the SVM classifiers is better than those of the MLR and kNN classifiers. Besides, the binary-classification (i.e., HC versus PD) gives quite better results than the multiclass-classification, as expected. Moreover, in the last case, the classification error for the optimized SVM was never greater than 1 UPDRS class for all the tasks, and on the average well below of this value. This indicates that chosen classifiers are robust and, in any case, they do not make assessments too far from neurologists. Furthermore, these results agree with Table 10 about the measure of the inter-rater agreement ICC_{N12-SY}, which indicate that the system performs almost as a third neurologist, except for PS_{PIGD} task. For this task, the lower value of ICC_{N12-SY} as compared to ICC_{N12} can be due to CoM parameters that are not directly comparable to PIGD subscale assessments or to the limited number of PD subjects included in the training set.

Due to the novelty of our approach, based on low-cost optical RGB-D device, we cannot compare directly the results of the classification accuracy with other similar works. Furthermore, a limited attention has been devoted to the automated assessment of specific UPDRS tasks by motion capture technologies. Then, we decided to refer to approaches based on wearable devices employing supervised classifiers [63]. Even if not directly comparable with our tasks, Timed Up and Go (TUG) test in [64] discriminate PD from HC by machine learning approach, with accuracy of 77.5%, which is lower than the value we have obtained (Table 11). In [6] the accuracy values for the multiclass classification of the LA and AC tasks are about of 43%, which are lower than ours (Table 11), even if care must be taken because the number of classes considered is different.

Summarizing, to our knowledge this is the first time that posture instability and lower limb motor tasks were assessed with reference to the clinical UPDRS context by a system based on optical RGB-D device. The results on the classifier accuracies and on ICC show that the automated assessments of the system are comparable with the clinical ones, then demonstrating their effectiveness. Furthermore, it is

also the first time that a system based on low-cost optical device characterizes CoM movements for the assessment of Parkinson's Disease. Finally, another original feature is the interpretation of the posture improvement during quite stance as secondary motor task, and the findings about its effectiveness in assessing postural instability in PD subjects.

Certainly, some aspects of this work require a further investigation. For instance, the number of analyzed subjects should be increased to obtain a more robust characterization of each single task and a better accuracy in the automated assessments. Furthermore, the PD subjects should be distinguished in phenotypes to verify if different sets of parameters could characterize different subtypes of parkinsonians; other balance tests should be considered to assess balance instability. These will be the next steps of our activity; the current findings encourage us to continue along this line of research to achieve a comprehensive system for the automatic and reliable assessment of PD status, suitable for the home monitoring of disease progression.

Limitations

Recently Microsoft announced that the Kinect device was discontinued [65], even if there is a cooperation with Intel to provide a transition from Kinect to Intel RealSense [66] or Orbbec cameras [67]. Even though our current implementation relies on the Kinect for body tracking, the Orbbec SDK or the sensor independent NUI Tracker middleware [68] are equivalent replacements for the purpose of this work. Furthermore, according to the specifications, Intel RealSense D415 combined with the NUI Tracker environment can output skeleton information at a double rate (60 fps) respect to Kinect device, providing more accuracy for fast movements.

7. Conclusions

In this paper, a self-managed system for the automated assessment of Parkinson's disease at home is presented. The automated assessment is focused on lower limbs, posture and postural stability tasks as specified by standard clinical assessment scales. A high usability of the system is guaranteed to motor impaired users by a gesture based human computer interface. The patient movements are characterized by sets of selected kinematic parameters which best correlate with clinical UPDRS scores, collected in an experimental campaign conducted on PD subjects. The data acquired have been used to train supervised classifiers employed for the automated assessment of new task instances. For the first time, in the context of Parkinson's disease, low-cost optical tracking devices are used to characterize center of mass movements as an index of postural instability. Preliminary results on the assessment accuracy, as compared to standard clinical evaluations, suggest that the proposed system is suitable for an objective assessment of posture and lower limb UPDRS tasks, also in a domestic environment, and then it could be the basis for the development of neuromonitoring and neurorehabilitation applications in a telemedicine framework.

Author Contributions: C.F. and R.N. designed and developed the system, analyzed the PD data and wrote the paper; A.C. and G.P. gave technical support on the development and contributed to review the paper; N.C. and V.C. provided the optoelectronic facilities and data; C.A., L.P. and A.M. designed and supervised the clinical experiment on PD subjects and assessed the patient performance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pal, G.; Goetz, C.G. Assessing bradykinesia in Parkinsonian Disorders. *Front. Neurol.* **2013**, *4*, 1–5. [CrossRef] [PubMed]
2. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martinez-Martin, P.; Dubois, B. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **2008**, *23*, 2129–2170. [CrossRef] [PubMed]

3. Richards, M.; Marder, K.; Cote, L.; Mayeux, R. Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examination. *Mov. Disord.* **1994**, *9*, 89–91. [CrossRef] [PubMed]

4. Espay, A.J.; Bonato, P.; Nahab, F.B.; Maetzler, W.; Dean, J.M.; Klucken, J.; Eskofier, B.M.; Merola, A.; Horak, F.; Lang, A.E.; et al. Movement Disorders Society Task Force on Technology. Technology in Parkinson's disease: Challenges and opportunities. *Mov. Disord.* **2016**, *31*, 1272–1282. [CrossRef] [PubMed]

5. Rovini, E.; Maremmani, C.; Cavallo, F. How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review. *Front. Neurosci.* **2017**, *11*, 555. [CrossRef] [PubMed]

6. Parisi, F.; Ferrari, G.; Giuberti, M.; Contin, L.; Cimolin, V.; Azzaro, C.; Albani, G.; Mauro, A. Body-sensor-network-based kinematic characterization and comparative outlook of UPDRS scoring in leg agility, sit-to-stand, and Gait tasks in Parkinson's disease. *J. Biomed. Heal Inf.* **2015**, *19*, 1777–1793. [CrossRef] [PubMed]

7. Godinho, C.; Domingos, J.; Cunha, G.; Santos, A.T.; Fernandes, R.M.; Abreu, D.; Goncalves, N.; Matthews, H.; Isaacs, T.; Duffen, J.; et al. A systematic review of the characteristics and validity of monitoring technologies to assess Parkinson's disease. *J. NeuroEng. Rehabil.* **2016**, *13*, 24. [CrossRef] [PubMed]

8. Hasan, H.; Athauda, D.S.; Foltynie, T.; Noyce, A.J. Technologies assessing limb bradykinesia in Parkinson's disease. *J. Parkinson's Dis.* **2017**, *7*, 65–77. [CrossRef] [PubMed]

9. Heldman, D.A.; Giuffrida, J.P.; Chen, R.; Payne, M.; Mazzella, F.; Duker, A.P.; Sahay, A.; Kim, S.J.; Revilla, F.J.; Espay, A.J. The modified bradykinesia rating scale for Parkinson's disease: Reliability and comparison with kinematic measures. *Mov. Disord.* **2011**, *26*, 1859–1863. [CrossRef] [PubMed]

10. Espay, A.J.; Giuffrida, J.P.; Chen, R.; Payne, M.; Mazzella, F.; Dunn, E.; Vaughan, J.E.; Duker, A.P.; Sahay, A.; Kim, S.J.; et al. Differential Response of Speed, Amplitude and Rhythm to Dopaminergic Medications in Parkinson's Disease. *Mov. Disord.* **2011**, *26*, 2504–2508. [CrossRef] [PubMed]

11. Han, J.; Shao, L.; Shotton, J. Enhanced computer vision with Microsoft Kinect sensor: A review. *IEEE Trans. Cybern.* **2013**, *43*, 1318–1334. [CrossRef] [PubMed]

12. Microsoft Kinect 2 SDK. Available online: https://developer.microsoft.com/en-us/download/driver (accessed on 4 December 2018).

13. Leap Motion Controller. Available online: https://www.leapmotion.com (accessed on 3 August 2018).

14. Intel Developer Zone. Available online: https://software.intel.com/en-us/realsense/previous (accessed on 3 August 2018).

15. Galna, B.; Barry, G.; Jackson, D.; Mhiripiri, D.; Olivier, P.; Rochester, L. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait Posture* **2014**, *39*, 1062–1068. [CrossRef] [PubMed]

16. Mousavi Hondori, H.; Khademi, M. A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation. *J. Med. Eng.* **2014**, *8465*. [CrossRef] [PubMed]

17. Anton, D.; Berges, I.; Bermúdez, J.; Goñi, A.; Illarramendi, A. A Telerehabilitation System for the Selection, Evaluation and Remote Management of Therapies. *Sensors* **2018**, *18*, 1459. [CrossRef] [PubMed]

18. Yeung, L.F.; Cheng, K.C.; Fong, C.H.; Lee, W.C.; Tong, K.Y. Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway. *Gait Posture* **2014**, *40*, 532–538. [CrossRef] [PubMed]

19. Yang, Y.; Pu, F.; Li, Y.; Li, S.; Fan, Y.; Li, D. Reliability and validity of Kinect RGB-D sensor for assessing standing balance. *IEEE Sens. J.* **2014**, *14*, 1633–1638. [CrossRef]

20. Springer, S.; Seligmann, G. Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors* **2016**, *16*, 194. [CrossRef] [PubMed]

21. Nguyen, T.N.; Huynh, H.H.; Meunier, J. Skeleton-Based Abnormal Gait Detection. *Sensors* **2016**, *16*, 1792. [CrossRef] [PubMed]

22. Blumrosen, G.; Miron, Y.; Intrator, N.; Plotnik, M. A Real-Time Kinect Signature-Based Patient Home Monitoring System. *Sensors* **2016**, *16*, 1965. [CrossRef] [PubMed]

23. Ferraris, C.; Nerino, R.; Chimienti, A.; Pettiti, G.; Pianu, D.; Albani, G.; Azzaro, C.; Contin, L.; Cimolin, V.; Mauro, A. Remote monitoring and rehabilitation for patients with neurological diseases. In Proceedings of the 10th International Conference on Body Area Networks (BODYNETS 2014), London, UK, 29 September–1 October 2014; pp. 76–82, ISBN 978-1-63190-047-1. [CrossRef]

24. Stone, E.E.; Skubic, M. Fall detection in homes of older adults using the Microsoft Kinect. *IEEE J. Biomed. Health Inf.* **2015**, *19*, 290–301. [CrossRef] [PubMed]

25. Wang, Q.; Kurillo, G.; Ofli, F.; Bajcsy, R. Evaluation of Pose Tracking Accuracy in the First and Second generations of Microsoft Kinect. In Proceedings of the International Conference on Healthcare Informatics, Dallas, TX, USA, 21–23 October 2015; ISBN 978-1-4673-9548-9. [PubMed]

26. Napoli, A.; Glass, S.M.; Ward, C.R.; Tucker, C.A.; Obeid, I. Performance analysis of a generalized motion capture system using microsoft kinect 2.0. *Biomed. Signal Process. Control* **2017**, *38*, 265–280. [CrossRef]

27. Clark, R.A.; Pua, Y.H.; Oliveira, C.C.; Bower, K.J.; Thilarajah, S.; McGaw, R.; Mentiplay, B.F. Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control. *Gait Posture* **2015**, *42*, 210–213. [CrossRef] [PubMed]

28. Müller, B.; Ilg, W.; Giese, M.A.; Ludolph, N. Validation of enhanced Kinect sensor based motion capturing for gait assessment. *PLoS ONE* **2017**, 12. [CrossRef] [PubMed]

29. Mishra, A.K.; Skubic, M.; Willis, B.W.; Guess, T.M.; Razu, S.S.; Abbott, C.; Gray, A.D. Examining methods to estimate static body sway from the Kinect V2. 0 skeletal data: Implications for clinical rehabilitation. In Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, Barcelona, Spain, 23–26 May 2017; pp. 127–135. [CrossRef]

30. Otte, K.; Kayser, B.; Mansow-Model, S.; Verrel, J.; Paul, F.; Brandt, A.U.; Schmitz-Hübsch, T. Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Functions. *PLoS ONE* **2016**, *11*, e0166532. [CrossRef] [PubMed]

31. Kähär, H.; Taba, P.; Nõmm, S.; Medijainen, K. Microsoft Kinect-based differences in lower limb kinematics during modified timed up and go test phases between men with and without Parkinson's disease. *Acta Kinesiologiae Universitatis Tartuensis* **2017**, *23*, 86–97. [CrossRef]

32. Rocha, A.P.; Choupina, H.; Fernandes, J.M.; Rosas, M.J.; Vaz, R.; Silva Cunha, J.P. Kinect v2 based system for Parkinson's disease assessment. In Proceedings of the IEEE 37th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 1279–1282. [CrossRef]

33. Li, Q.; Wang, Y.; Sharf, A.; Cao, Y.; Tu, C.; Chen, B.; Yu, S. Classification of gait anomalies from kinect. *Vis. Comput.* **2018**, *34*, 229–241. [CrossRef]

34. Knippenberg, E.; Verbrugghe, J.; Lamers, I.; Palmaers, S.; Timmermans, A.; Spooren, A. Markerless motion capture systems as training device in neurological rehabilitation: A systematic review of their use, application, target population and efficacy. *J. Neuroeng. Rehabil.* **2017**, *14*, 61. [CrossRef] [PubMed]

35. Ferraris, C.; Nerino, R.; Chimienti, A.; Pettiti, G.; Cau, N.; Cimolin, V.; Azzaro, C.; Albani, G.; Priano, L.; Mauro, A. A Self-Managed System for Automated Assessment of UPDRS Upper Limb Tasks in Parkinson's Disease. *Sensors* **2018**, *18*, 3523. [CrossRef] [PubMed]

36. Bloem, B.R.; Grimbergen, Y.A.; Cramer, M.; Willemsen, M.; Zwinderman, A.H. Prospective assessment of falls in Parkinson's disease. *J. Neurol.* **2001**, *248*, 950–958. [CrossRef] [PubMed]

37. Shoneburg, B.; Mancini, M.; Horak, F.; Nutt, J.G. Framework for understanding balance dysfunction in Parkinson's disease. *Mov. Disord.* **2013**, *28*, 1474–1482. [CrossRef] [PubMed]

38. Morris, M.; Iansek, R.; Smithson, F.; Huxham, F. Postural instability in Parkinson's disease: A comparison with and without a concurrent task. *Gait Posture* **2000**, *12*, 205–216. [CrossRef]

39. Frenklach, A.; Louie, S.; Koop, M.M.; Bronte-Stewart, H. Excessive postural sway and the risk of falls at different stages of Parkinson's disease. *Mov. Disord.* **2009**, *24*, 377–385. [CrossRef] [PubMed]

40. Ozinga, S.J.; Machado, A.G.; Miller Koop, M.; Rosenfeldt, A.B.; Alberts, J.L. Objective assessment of postural stability in Parkinson's disease using mobile technology. *Mov. Disord.* **2015**, *30*, 1214–1221. [CrossRef] [PubMed]

41. Mancini, M.; Horak, F.B.; Zampieri, C.; Carlson-Kuhta, P.; Nutt, J.G.; Chiari, L. Trunk accelerometry reveals postural instability in untreated Parkinson's disease. *Parkinsonism Relat. Disord.* **2011**, *17*, 557–562. [CrossRef] [PubMed]

42. Ferrazzoli, D.; Fasano, A.; Maestri, R.; Bera, R.; Palamara, G.; Ghilardi, M.F.; Pezzoli, G.; Frazzitta, G. Balance Dysfunction in Parkinson's Disease: The Role of Posturography in Developing a Rehabilitation Program. *Parkinson's Dis.* **2015**, *2015*, 520128. [CrossRef]

43. Rezvanian, S.; Lockart, T.; Frames, C.; Lieberman, R.; Soangra, A. Motor Subtypes of Parkinson's Disease Can Be Identified by Frequency Component of Postural Stability. *Sensors* **2018**, *18*, 1102. [CrossRef] [PubMed]

44. Grooten, W.J.A.; Sandberg, L.; Ressman, J.; Diamantoglou, N.; Johansson, E.; Rasmussen-Barr, E. Reliability and validity of a novel Kinect-based software program for measuring posture, balance and side-bending. *BMC Musculoskelet. Disord.* **2018**, *19*, 6. [CrossRef] [PubMed]

45. Napoli, A.; Glass, S.M.; Tucker, C.; Obeid, I. The automated assessment of postural stability: Balance detection algorithm. *Ann. Biomed. Eng.* **2017**, *45*, 2784–2793. [CrossRef] [PubMed]
46. Dehbandi, B.; Barachant, A.; Smeragliuolo, A.H.; Long, J.D.; Bumanlag, S.J.; He, V.; Lampe, A.; Putrino, D. Using data from the Microsoft Kinect 2 to determine postural stability in healthy subjects: A feasibility trial. *PLoS ONE* **2017**, *12*, e0170890. [CrossRef] [PubMed]
47. Jankovic, J.; McDermott, M.; Carter, J.; Gauthier, S.; Goetz, C.; Golbe, L.; Huber, S.; Koller, W.; Olanow, C.; Shoulson, I.; et al. Variable expression of Parkinson's disease: A base-line analysis of the DATATOP cohort. The Parkinson study group. *Neurology* **1990**, *40*, 1529–1534. [CrossRef] [PubMed]
48. BTS S.p.A. Products. Available online: http://www.btsbioengineering.com/it/prodotti/smart-dx (accessed on 11 August 2018).
49. Kadaba, M.P.; Ramakrish, H.K.; Wooten, M.E. Measurement of lower extremity kinematics during level walking. *J. Othop. Res.* **1990**, *8*, 383–391. [CrossRef] [PubMed]
50. Davis, R.B.; Ounpuu, S.; Tyburski, D.J.; Gage, J.R. A gait analysis data collection and reduction technique. *Hum. Mov. Sci.* **1991**, *10*, 575–587. [CrossRef]
51. Gutierrez, E.M.; Bartonek, A.; Haglund-Akerlind, Y.; Saraste, H. Centre of mass motion during gait in person with myelomengocele. *Gait Posture* **2003**, *18*, 37–46. [CrossRef]
52. Lafond, D.; Duarte, M.; Prince, F. Comparison of three methods to estimate the center of mass during balance assessment. *J. Biomech.* **2004**, *37*, 1421–1426. [CrossRef]
53. González, A.; Hayashibe, M.; Bonnet, V.; Fraisse, P. Whole Body Center of Mass Estimation with Portable Sensors: Using the Statically Equivalent Serial Chain and a Kinect. *Sensors* **2014**, *14*, 16955–16971. [CrossRef] [PubMed]
54. Clauser, C.E.; McConville, J.T.; Young, J.W. Weight, Volume, and Center of Mass Segments of the Human Body. *J. Occup. Med.* **1971**, *13*, 270.
55. Van Crombrugge, I.; Mertens, L.; Penne, R. Fast Free Floor Detection for Range Cameras. In Proceedings of the VISIGRAPP, Porto, Portugal, 27 February–1 March 2017; Volume 4, pp. 509–516. [CrossRef]
56. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
57. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2005**, *67*, 301–320. [CrossRef]
58. Data Mining: Practice Machine Learning Tools and Techniques. Available online: https://www.cs.waikato. ac.nz/ml/weka/book.html (accessed on 28 September 2018).
59. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27. [CrossRef]
60. Rocha, A.; Goldenstein, S.K. Multiclass from binary: Expanding One-Versus-All, One-Versus-One and ECOC-Based Approaches. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 289–302. [CrossRef] [PubMed]
61. Louis, E.D.; Levy, G.; Côte, L.J.; Mejia, H.; Fahn, S.; Marder, K. Diagnosing Parkinson's disease using videotaped neurological examinations: Validity and factors that contribute to incorrect diagnoses. *Mov. Disord.* **2002**, *17*, 513–517. [CrossRef] [PubMed]
62. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
63. Kubota, K.J.; Chen, J.A.; Little, M.A. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Mov. Disord.* **2016**, *31*, 1314–1326. [CrossRef] [PubMed]
64. Palmerini, L.; Mellone, S.; Avanzolini, G.; Valzania, F.; Chiari, L. Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 664–673. [CrossRef] [PubMed]
65. Microsoft Kinect for Windows. Available online: https://developer.microsoft.com/en-us/windows/kinect (accessed on 20 December 2018).
66. Intel. Intel Real Sense. Available online: https://realsense.intel.com/stereo/ (accessed on 20 December 2018).
67. Orbbec. Available online: https://orbbec3d.com/products/ (accessed on 20 December 2018).
68. NUI Track SDK. Available online: https://nuitrack.com/ (accessed on 20 December 2018).

Heel Height as an Etiology of Hallux Abductus Valgus Development: An electromagnetic Static and Dynamic First Metatarsophalangeal Joint Study

Rubén Sánchez-Gómez [1,2,*], Ricardo Becerro de Bengoa-Vallejo [2], Marta Elena Losa-Iglesias [3], César Calvo-Lobo [4], Carlos Romero-Morales [1], Eva María Martínez-Jiménez [2], Patricia Palomo-López [5] and Daniel López-López [6]

[1] Faculty of Sports Sciences, Universidad Europea de Madrid, 28670 Villaviciosa de Odón, Madrid, Spain; carlos.romero@universidadeuropea.es

[2] Facultad de Enfermería, Fisioterapia y Podología, Universidad Complutense de Madrid, 28040 Madrid, Spain; ribebeva@ucm.es (R.B.d.B.-V.); eva.hache2@hotmail.com (E.M.M.-J.)

[3] Faculty of Health Sciences, Universidad Rey Juan Carlos, 28922 Alcorcón, Spain; marta.losa@urjc.es

[4] Nursing and Physical Therapy Department, Institute of Biomedicine (IBIOMED), Faculty of Health Sciences, Universidad de León, Ponferrada, 24071 León, Spain; ccall@unileon.es

[5] University Center of Plasencia, Universidad de Extremadura, 10600 Plasencia, Spain; patibiom@unex.es

[6] Research, Health and Podiatry Unit, Department of Health Sciences, Faculty of Nursing and Podiatry, Universidade da Coruña, 15403 Ferrol, Spain; daniellopez@udc.es

* Correspondence: rusanc02@ucm.es

Abstract: Background: Hallux abductus valgus (HAV) is a forefoot condition produced by extrinsic and intrinsic factors. Shoes with a high heel height and a typical narrow tip toe box can induce deviations in both the proximal phalanx of the hallux (PPH) and the first metatarsal (IMTT) bones. Nevertheless, the isolated role of heel height remains unclear in the development of HAV pathology. **Objectives:** The goal was to determine if the heel height increase of shoes without a narrow box toe could augment the PPH and IMTT deviation in frontal, sagittal, and transverse planes toward the first metatarsophalangeal joint (MPJ) and the first metatarsocuneiform joint (MCJ), respectively, during static and dynamic conditions in relation to precursor movements of HAV. **Methods:** Women with an average age of 25.10 ± 4.67 years were recruited in this cross-sectional study to assess the three planes of motion of PPH and IMTT while wearing high heels with heights at 3, 6, 9 cm and unshod conditions via sandals. The measurements used an electromagnetic goniometer device with sensors placed on medial aspects of the PPH and IMTT bones under static and dynamic conditions. **Results:** Wearing shoes with a 6 cm heel in dynamic condition may increase the PPH valgus and abduction deviation from $3.15 \pm 0.10°$ to $3.46 \pm 0.05°$ ($p < 0.05$) and from $1.35 \pm 0.28°$ to $1.69 \pm 0.30°$ ($p < 0.001$), respectively. In addition, a PPH abduction increase from $1.01 \pm 0.36°$ to $1.31 \pm 0.46°$ ($p < 0.05$) after wearing shoes with a 6 cm heel height was observed under static conditions. **Conclusions:** Wearing shoes with a heel height of 6 cm without a narrow box toe interference may produce PPH abduction and valgus deviations related to HAV formation.

Keywords: hallux abductus valgus; high heel; proximal phalanx of the hallux; abduction; valgus

---

1. Introduction

Hallux Abductus Valgus (HAV) is a pathological subluxation of the first metatarsophalangeal joint (IMPJ) with lateral deviation of the proximal phalanx of the hallux bone (PPH) toward abduction and valgus direction in both the transverse and frontal planes of motion, respectively. There is also deviation of the first metatarsocuneiform joint (IMCJ) through the first metatarsal bone (IMTT) toward

adduction and valgus direction in both transverse and frontal planes of motion, respectively. There are occasionally bony enlargements of the first metatarsal head (also called a "bunion") [1–3]. The IMPJ bears 80% of the body load without help of any structure in heel-off phase, and this makes it a more sensitive joint to biomechanics deformities such HAV [1].

The etiology of HAV can have multiple origins, and there are intrinsic and extrinsic factors in play [4–9]. Intrinsic factors include hyperpronation [5,10], soft tissue weakness [1], and hyperlaxity with medial longitudinal arch collapse [11,12]; Windlass mechanism failure [7], first ray hypermobility [8], and female sex [13–15] have been linked with HAV growth. As extrinsic factors, the use of high-heel shoes has also been detected as a possible cause of HAV development: The typical high-heeled shoe for women can lead to bad body repercussions and be detrimental to bone mineralization [16], rear foot instability [17], body mass center changes [18], biomechanics gait changes [19,20], and general damage health [21].

Some authors have reported increases in concentrate load under the forefoot with high heeled shoes [22–25], and this condition can predispose the subject to HAV development [26]. A few studies have speculated that the current narrow box of high-heeled shoes are not the only cause of HAV because isolated high-heel shoes might cause weight to be placed on the forefoot, and this may overstretch the toes and lead to the development of splayfoot [1,27,28]. However, this has not been reported in the literature.

The 2D [29] and 3D kinematic movements of PPH and I MTT bone deviation during gait in subjects with and without HAV has shown the relationship between the rearfoot and midfoot eversion with respect to the first ray hypermobility and the presence on HAV [10,30]; other groups have studied the kinematic effects of improvements in taping in subjects with HAV during gait [31] or the negative effects of HAV surgery to normal ambulation [32]. Other work studied the kinematics effects on IMPJ using foot orthoses that incorporated forefoot and rearfoot posting—the results showed no negative effects on mobility [33]. One prior study [34] identified kinetic evidence of wearing 5 cm-heeled shoes during gait can lead to HAV development and an increase in hallux dorsiflexion in the final push-off phase; however, no report has described the transverse or frontal plane motion of the PPH or IMTT.

Therefore, the goal of this study is to determine how the heel height affects PPH and IMTT bone deviations either in the three planes of motion in static conditions and during the dorsiflexion of the IMPJ sequence (push-off phase of the dynamic condition) regardless of toe box of the shoe. The results can show if there is some movement related to HAV deviation that is characterized by PPH in abduction (away to medial body line) in transverse plane and valgus deviation in frontal plane toward the IMPJ [2] and/or the adduction deviation of the IMTT (toward medial body line) regarding the second MTT in transverse plane and the presence of the valgus in the frontal plane toward the IMCJ [2].

2. Materials and Methods

2.1. Subjects

The institutional review board at Rey Juan Carlos University approved the study. All subjects signed the informed consent form prior to beginning this study. The following inclusion criteria were required to participate in this study [35]: (1st) to have at least $10°$ dorsiflexion of the ankle after complete knee fully dorsiflexion; (2nd) to have at least $30°$ motion in the functional subtalar joint; (3rd) to have at least $15°$ of motion along the mid-tarsal joint longitudinal axis; (4th) to reach at least 8 mm of non-weight bearing motion of the first ray; (5th) older than 18 years and younger than 38 years; (6th) no lower limb pathology or chronic condition at the time of recruitment and measurement; and (7th) prior experience of the use of shoes with high heel height during at least the prior year [36]. The exclusion criteria included established hallux valgus; prior history of foot and lower limb traumas, fractures, or surgeries; and congenital deformities or existing diagnoses of neurological, inflammatory, metabolic or vascular diseases. Subjects were also excluded if they showed less than $40°$ passive

dorsiflexion of the first MPJ (as a non-weight bearing technique previously described according to Buell et al. [37]). This range has been reported to be the IMPJ dorsiflexion range used during normal propulsion [38,39] and indicates joint structural limitations. In addition, male sex was an exclusion criterion because they rarely wear high-heeled shoes.

2.2. Sample Size

The sample size was calculated with software from Unidad de Epidemiología Clínica y Bioestadística. Complexo Hospitalario Universitario de A Coruña. Universidade A Coruña (www.fisterra.com) to detect a significant difference in the deviation of the axes of IMTT and PPH during static and dynamic test wearing high heels versus barefoot similar in another study with 15 recruited subjects [34] where hallux dorsiflexion decreased while wearing high heels from $26.6 \pm 2.33°$ to $22.5 \pm 1.62°$ in barefoot condition with 80% statistical power. Here, $\beta = 20\%$ with a 95% interval of confidence and $\alpha = 0.05$ using a 2-tailed test. Thus, at least 61 participants were required. Furthermore, assuming a loss to follow up rate of 15%, at least 71 participants were included in the study. However, recruited 80 subjects.

2.3. Instrumentation; Assessment of the IMTT and PPH

Three-dimensional degrees of the angular deviation of the IMTT bone movement; dorsiflexion/plantar flexion, adduction/abduction, and inversion/eversion at the first MCJ; dorsiflexion/plantar flexion, adduction/abduction, and inversion/eversion of PPH bone; and dorsiflexion/plantarflexion, adduction/abduction and inversion/eversion at the first MPJ, with and without wearing high-heeled shoes were studied during static and push-off phases using a 6Space Fastrak® (Polhemus, Inc., Colchester, VT, USA) (Figure 1). This system had a 120 Hz sampling rate and a excellent validity and reliability ($r > 0.99$) [38] with ICC = 0.88–0.99 and 0.95–0.99 (SEM = 0.7–0.8 mm), respectively [40], according to previous studies; the equipment contained two sensors that send an electromagnetic field with 6° of freedom relative to the electromagnetic emitter. A 30-foot serial cable connected the sensors and emitter to a receiver module that collected, filtered, and managed all the signal's information.

Figure 1. Polhemus device. From left to the right: sensors, receiver module, emitter module.

The signal from each sensor was captured by the receiver module. It was digitally transformed by the software to generate spatial orientation data from each sensor. The system showed a 8-mm static accuracy relative to the sensor position and 0.15° regarding sensor orientation; the error was 1.6% [38,41]. The relation between range vs. resolution in orientation aspect was 0.3 m to 1.16 m; in addition, the relation between range vs. orientation aspect was 0.3 to 0.0038°.

In this study, the electromagnetic emitter was 96 cm high at the midway point on a 6-m raised walkway. There were no metallic elements near the electromagnetic device or in the subject's walking path to avoid possible interference [42]. One of the sensors was placed at the medial aspect of the head of the IMTT bone (Figure 2), and the second one was located along the medial aspect of the PPH (Figure 3) on the right foot; the Polhemus Fastrack® (Polhemus, Inc., Colchester, VT, USA) did not measure angles but did measure bone displacements. Thus, we assessed the mobility of the PPH regardless of IMTT mobility to determine which areas were affected by the high heels. Both sensors were attached with tape and secured to the skin with an auto-adhesive bandage (Figure 4); this medial location was selected according to a protocol devised by Welsh et al. [43] to minimal overlying soft tissue of extensor hallucis longus tendon's excursion [41,44] All cables were fixed via straps to the thighs with a belt.
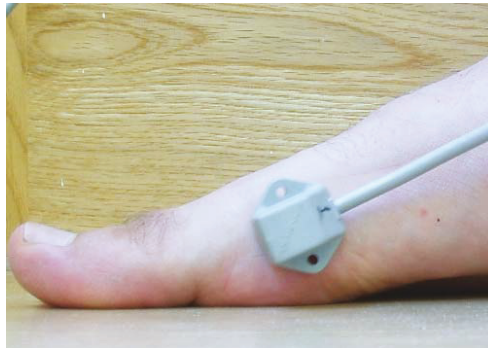

Figure 2. Sensor one. Location placed on the medial head of first metatarsal bone.

Figure 3. Sensor two. Location placed on the border of the proximal phalanx of the hallux.

Figure 4. Fastening of the sensors. White bandages are auto-adhesive fixatives.

2.4. Procedure

Each of 80 participants began static or dynamic study randomly after choosing a sealed envelope that assigned them to one group or another; there were 40 total in each group. They then selected four sealed envelopes with each of the four conditions of the study (unshod, 3 cm, 6 cm, and 9 cm). These dictated the order of use at random. The 40 participants of the dynamic group did the static test in randomly order with four sealed envelopes and vice versa with the 40 participants in the static test (Figure 5).

Before beginning either angular measurements, position "zero" was achieved to calibrate the assessment and determine a reference position from which to begin the angular record. We asked the participants to remain in a relaxed standing position near the generated field for a few seconds until the software determined the Reference [35].

Participants used a pair of sandals with three different prefabricated high heels (Figure 6). The sandals (Figure 7) provided a strap in the first digital web space, and the rearfoot had a semi-rigid bowl that permitted the heels to be held into the shoes. The participants used the sandals and tested the sensations with the cables before starting the study; they then walked along the walkway at a self-selected speed. When they were comfortable, they began the static or dynamic measurements depending on the order chosen by the random chance.

For the dynamic condition study, participants initiated gait for 1 m at a self-selected speed before entering the 1.5 m calibrated capture volume. They then continued walking for a further 1.5 m (two steps); this capture was repeated for 5 trials with each of 3 different types of prefabricated heel heights (3 cm, 6 cm and 9 cm) and unshod condition. Care was taken when inserting and removing the prefabricated heel height so that the sensor devices was not disturbed or displaced. For the static condition study, subjects were asked to stand on their tiptoe (to reach the maximal range of motion of the joint) for 2 s. There were five trials with each of three different types of prefabricated heel heights (3 cm, 6 cm and 9 cm) and unshod condition. The mean of the five trials was used in the posterior analysis for each static and dynamic test.
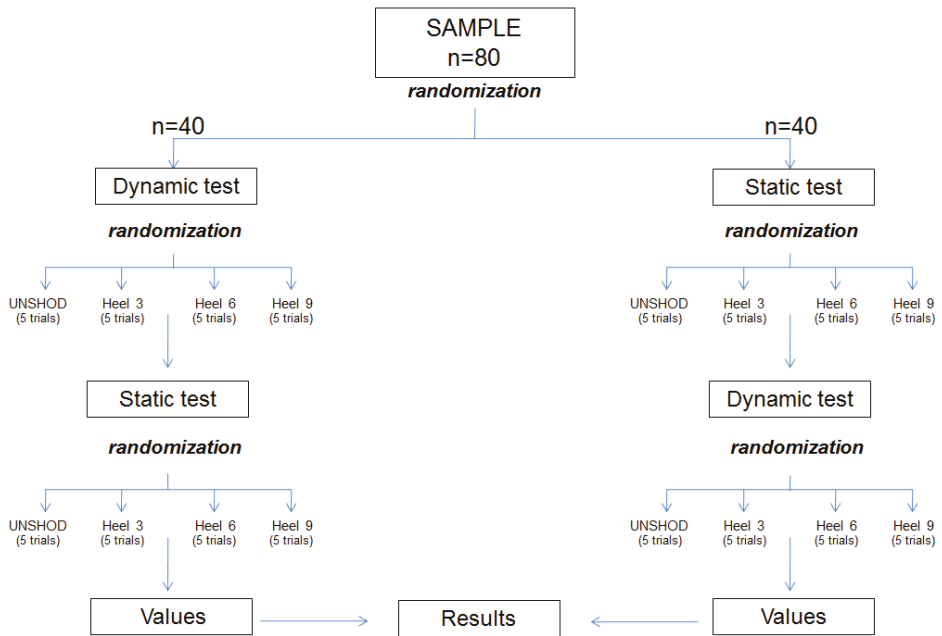
Figure 5. Randomized flow chart.

Figure 6. Prefabricated high-heels. The heels were made on hard silicone. The additional height was added over the top.

Figure 7. Sandal model used in the study. A free toe-tip-box is present to allow movement of proximal phalanx of the hallux and first metatarsal bone.

Later, a "cardan system" with three possibilities of rotation (X-Y-Z) in two ways was used to draw angular and linear information from both PPH and IMTT bones movements: mediolateral axis (X), defined dorsiflexion (DF), and plantar flexion (PF) movements on the sagittal plane as well as frontal plane inversion (VR) and eversion (VL) movements through an antero-posterior axis (Y). Movement around a craneocaudal axis (Z) was considered to be abduction (ABD) and adduction (ADD) movements on a transverse plane. The anatomical landmarks and coordinate systems were previously detailed [35,44,45]. All bone movements were recorded after considering the laboratory coordinate system. All assessments used the same heel height in contralateral extremity under the foot to maintain the body balance during these trials [46]. Independent assessment of the bone movements showed that the results were only produced by high heel effects and not by bone interference.

2.5. Statistical Analysis

We recorded the within-day trial-to-trial intraclass correlation coefficients (ICC) and standard errors of measurement (SEM) [47] values for the participants wearing the different heel height (T3, T6, T9) vs. unshot in each plane of motion for PPH bone and IMTT bone in static and dynamic conditions. Landis and Koch [47] proposed that coefficients below 0.20 indicate slight agreement, and coefficients from 0.20 to 0.40 indicate fair reliability. Coefficients from 0.41 to 0.60 indicate moderate reliability, coefficients from 0.61 to 0.80 indicate substantial reliability, and coefficients from 0.81 to 1.00 indicate almost perfect reliability. We considered coefficients of 0.90 or larger to reflect a sufficient magnitude of reliability because they increase the likelihood that a measure is also reasonably valid. To check the concordance correlation between high heels' variables, Concordance Correlation Coefficient (CCC) [48] was done, where statistically significant p-values (<0.05) would mean a perfect correlation [48]. SEM was utilized to determine the minimal detectable change (MDC) for all evaluations. This was also considered the Reliable Change Index (RCI). We utilized the RCI as a statistical method to determine the clinical significance according to Jacobson and Truax [49].

An initial Kolmogorov-Smirnov test showed that the data was not normally distributed ($p < 0.05$). The p-values for multiple comparisons were corrected with a non-parametric paired Friedman test to prove that all high-heels variables were different. Bivariate correlations with a Wilcoxon test were carried out to determine whether there were significant differences between "unshod" vs. "with high heels of T3, T6 and T9" in static and dynamic conditions in FPH and IMTT bones; an alpha level of 0.01 was established for all tests of significance. All data were studied to establish the effect of heel height in 3 axes of movement with the medial line of the body taken as reference movement. In addition, Spearman's Rho (ρ) rank correlation coefficient both in static and dynamic conditions was done to check the possible relation between the height of the heels and the different bones' deviations. The results showed descriptive summaries as the mean $\pm$ SD. Analyses of total 96 variables led to p-values <0.05 (within a 95% confidence interval) that considered statistically significant. We conducted data analysis with SPSS software version 19.0 (SPSS Science, Chicago, IL, USA).

3. Results

None of the data were normally distributed ($p < 0.05$). The means that the 96 measured variables were different when compared between each other ($p < 0.001$). Participants in this cross-sectional study were recruited from an orthopedic clinic in Madrid (Spain) over a 2-year period (May 2015 to May 2017). Of the 163 subjects who initially volunteered to participate in the study, 68 subjects did not meet the inclusion criteria. An additional 15 subjects did not present for testing. The remaining 80 subjects participated in the study (Figure 8). The participants were only females; sociodemographic data are shown in Table 1.

All Tables are shown with corresponding values of IMTT and PPH to the static test as well as independent of the dynamic values.
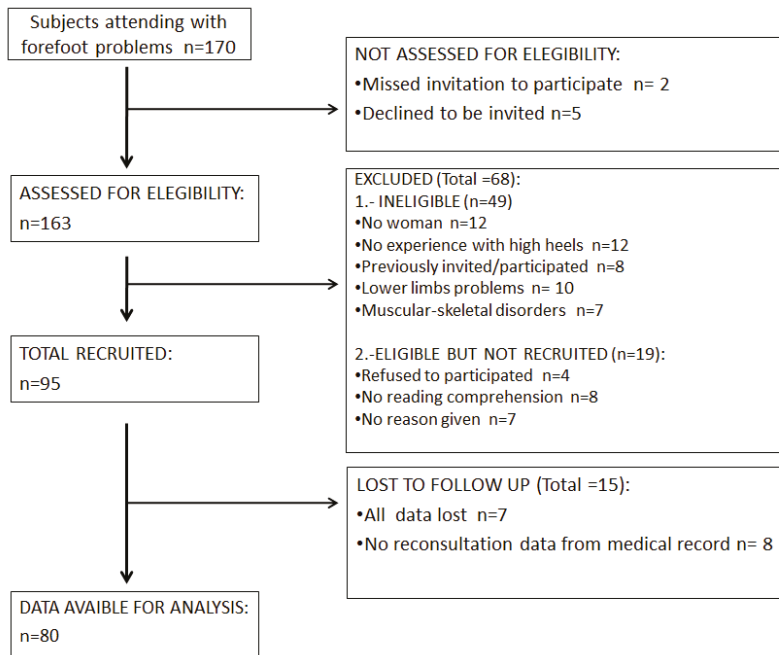
Figure 8. Participant flow chart.

Table 1. Sociodemographic characteristics of the participants.

Variable	Female $n = 80$
	Mean $\pm$ SD (95% CI)
Age (years)	25.1 ± 4.67 (26.11–24.08)
Height (cm)	164.03 ± 5.44 (165.2–162.85)
Weight (kg)	57.53 ± 6.304 (58.90–56.15)
Foot Size (EC)	38.13 ± 1.184 (38.37–37.88)
BMI (kg/m^2)	20.2 ± 1.74 (20.58–19.82)

Abbreviations: EC = European countries; BMI = body mass index; SD = Standard Deviation; CI = Confidence Interval.

The reliability of the static variables in unshod and heels of 3 cm, 6 cm and 9 cm high are summarized in Table 2. Each condition showed overall excellent reliability [50] with ICCs ranging from 0.902 to 0.997 indicating high reliability. In addition, in general terms, CCC showed a strong statistically significant correlation ($p < 0.001$) between unshod and either high heel in all movements of PPH and MTT except those related to a sagittal plane. The reliability of dynamic variables in unshod subjects for heel heights of 3 cm, 6 cm and 9 cm is summarized in Table 3. Each condition showed an overall excellent reliability [50] with ICCs ranging from 0.898 to 0.999 indicating high reliability. The CCC showed a strong correlation ($p < 0.001$) in all movements of PPH and MTT in dynamic tests, as well. The inter-rater MDC 95% values to static (Table 2) conditions ranged from 14.915° to 0.74° for sensor 1 (IMTT) and from 7.90° to 1.63° for sensor 2 (PPH); the inter-rater MDC 95% values to dynamic (Table 3) conditions ranged from 26.78° to 1.12° for sensor 1 (MTT) and from 27.16° to 2.16° for sensor 2 (PPH).

The SEM values for 6 cm high heel in static abduction of PPH (Table 2) was 0.591° and 0.782° for dynamic abduction condition (Table 3); in addition, the SEM was 2.02° for valgus movement of PPH in dynamic condition with high heels of 6 cm and 2.33° in static condition.

Table 2. Reliability ICC and CCC of static variables in unshod versus heels that were 3 cm, 6 cm, or 9 cm high.

Variable	Unshod ICC (95%CI)	Heel 3 cm ICC (95%CI)	Unshod-Heel 3 cm CCC (95%CI)	SEM	MDC 95%	Heel 6 cm ICC (95%CI)	Unshod-Heel 6 cm CCC (95%CI)	SEM	MDC 95%	Heel 9 cm ICC (95%CI)	Unshod-Heel 9 cm CCC (95%CI)	SEM	MDC 95%
MTT-AD	0.986 (0.981–0.991)	0.968 (0.983–0.991)	0.612 (0.493–0.7) **	1.14	3.161	0.991 (0.988–0.994)	0.477 (0.341–0.594) **	1.275	3.533	0.984 (0.978–0.0989)	0.184 (0.106–0.26) **	1.455	4.034
MTT-AB	0.904 (0.866–0.933)	0.936 (0.91–0.956)	0.15 (0.04–0.33) **	0.64	1.774	0.985 (0.965–0.862)	0.258 (0.042–0.45)	0.598	1.658	0.961 (0.945–0.973)	0.136 (0.012–0.277) **	0.711	1.97
PPH-AD	0.981 (0.964–0.992)	0.95 (0.93–0.965)	0.096 (0.039–0.153) **	1.357	3.762	0.969 (0.956–0.978)	0.054 (0.11–0.13) **	1.65	4.574	0.944 (0.922–0.961)	0.147 (0.081–0.212)	1.298	3.598
PPH-ABD	0.94 (0.916–0.958)	0.908 (0.872–0.937)	0.006 (0.01–0.019) **	0.828	2.296	0.926 (0.897–0.949)	0.239 (0.040–0.41) **	0.591	1.639	0.969 (0.957–0.979)	0.168 (0.044–0.365) **	0.63	1.747
MTT-PF	0.997 (0.995–0.998)	0.994 (0.992–0.996)	0.792 (0.727–0.842)	2.493	6.911	0.991 (0.988–0.994)	0.162 (0.033–0.286)	5.633	15.615	0.99 (0.986–0.993)	0.19 (0.104–0.272) *	5.381	14.915
MTT-DF	0.902 (0.864–0.932)	0.928 (0.90–0.95)	0.414 (0.219–0.577)	0.304	0.844	0.918 (0.886–0.944)	0.487 (0.30–0.63)	0.27	0.749	0.968 (0.955–0.978)	0.107 (0.008–0.218) *	0.579	1.604
PPH-PF	0.983 (0.977–0.989)	0.982 (0.974–0.987)	0.034 (0.16–0.23)	1.825	5.059	0.988 (0.984–0.992)	0.091 (0.10–0.279)	1.67	4.63	0.979 (0.971–0.986)	0.19 (0.013–0.356) **	1.574	4.363
PPH-DF	0.956 (0.939–0.970)	0.978 (0.969–0.984)	0.381 (0.182–0.549)	1.213	3.361	0.99 (0.986–0.993)	0.26 (0.066–0.435)	1.628	4.512	0.982 (0.975–0.988)	0.153 (0.037–0.67) **	2.274	6.303
MTT-VL	0.975 (0.965–0.983)	0.994 (0.992–0.996)	0.043 (0.004–0.091) **	2.471	6.848	0.961 (0.946–0.983)	0.158 (0.069–0.244) **	1.202	3.333	0.973 (0.962–0.981)	0.082 (0.033–0.13) **	1.277	3.539
MTT-VR	0.995 (0.993–0.996)	0.976 (0.967–0.984)	0.178 (0.29–0.31) *	2.419	6.706	0.99 (0.986–0.993)	0.64 (0.513–0.739) **	1.463	4.054	0.971 (0.960–0.98)	0.43 (0.33–0.52) **	1.463	4.747
PPH-VL	0.989 (0.985–0.993)	0.993 (0.99–0.995)	0.455 (0.328–0.567) **	2.023	5.608	0.985 (0.979–0.990)	0.162 (0.035–0.347) **	2.336	6.476	0.981 (0.974–0.987)	0.128 (0.051–0.299) **	2.143	5.94
PPH-VR	0.98 (0.972–0.986)	0.978 (0.969–0.985)	0.388 (0.198–0.55)	1.395	3.866	0.975 (0.966–0.983)	0.152 (0.058–0.35) **	1.921	5.326	0.989 (0.984–0.992)	0.13 (0.047–0.18) **	2.851	7.902

Abbreviations: ICC = Intraclass Correlation Coefficient; CI = Confidence Interval; CCC = Concordance Correlation Coefficient; SEM = standard error of measurement; MDC = minimal detectable change; MTT ADD = first metatarsal bone in Adduction, I MTT toward to medial body line; MTT ABD = first metatarsal bone in Abduction, I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexion, I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexion, I MTT away to the floor; MTT VL = first metatarsal bone spins to inside, toward medial body line; MTT VR = first metatarsal bone spin to outside, away to medial body line; PPH AD = proximal phalanx of hallux in adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexion: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexion: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward to medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; $\pm$SD = standard deviation. Movements and axes refer to the medial line of the body; all values are expressed in degrees; * p-value < 0.05; ** p-value < 0.001.

Table 3. Reliability ICC and CCC of dynamic variables in unshod versus heels that were 3 cm, 6 cm, or 9 cm high.

Variable	Unshod ICC (95%CI)	Heel 3 cm ICC (95%CI)	Unshod-Heel 3 cm CCC (95%CI)	SEM	MDC 95%	Heel 6 cm ICC (95%CI)	Unshod-Heel 6 cm CCC (95%CI)	SEM	MDC 95%	Heel 9 cm ICC (95%CI)	Unshod-Heel 9 cm CCC (95%CI)	SEM	MDC 95%
MTT-AD	0.988 (0.983–0.992)	0.986 (0.981–0.990)	0.191 (0.123–0.25) **	1.953	5.414	0.968 (0.955–0.978)	0.25 (0.171–0.327) **	1.785	4.949	0.986 (0.981–0.99)	0.232 (0.15–0.30)	1.813	5.026
MTT-AB	0.993 (0.991–0.995)	0.993 (0.990–0.995)	0.041 (0.48–0.131)	3.529	9.782	0.997 (0.995–0.998)	0.151 (0.055–0.245) *	3.499	9.699	0.992 (0.988–0.944)	0.28 (0.18–0.46)	2.715	7.526
PPH-AD	0.938 (0.913–0.957)	0.958 (0.941–0.971)	0.049 (0.11–0.17) **	1.051	2.914	0.965 (0.951–0.976)	0.026 (0.10–0.157) **	1.334	3.698	0.977 (0.969–0.984)	0.197 (0.06–0.321) *	1.117	3.096
PPH-ABD	0.976 (0.966–0.983)	0.983 (0.976–0.988)	0.04 (0.03–0.11) **	1.852	5.132	0.981 (0.973–0.987)	0.51 (0.33–0.65) **	0.782	2.169	0.994 (0.991–0.996)	0.295 (0.17–0.41) **	1.793	4.971
MTT-PF	0.998 (0.998–0.999)	0.997 (0.996–0.998)	0.834 (0.77–0.88)	4.421	12.255	0.998 (0.997–0.999)	0.799 (0.714–0.861)	5.051	14.001	0.992 (0.988–0.994)	0.268 (0.14–0.38)	9.662	26.782
MTT-DF	0.935 (0.910–0.955)	0.898 (0.857–0.929)	0.356 (0.186–0.521)	0.404	1.121	0.933 (0.907–0.954)	0.198 (0.015–0.368)	0.416	1.153	0.953 (0.935–0.968)	0.453 (0.262–0.61)	0.417	1.155
PPH-PF	0.999 (0.999–1.00)	0.999 (0.998–0.999)	0.723 (0.608–0.808)	6.755	18.723	0.998 (0.998–0.999)	0.857 (0.787–0.9)	4.929	13.662	0.996 (0.995–0.997)	0.425 (0.298–0.538)	9.801	27.166
PPH-DF	0.981 (0.974–0.987)	0.917 (0.844–0.943)	0.175 (0.002–0.185)	1.181	3.272	0.928 (0.9–0.950)	0.198 (0.015–0.36)	1.17	3.243	0.958 (0.941–0.971)	0.231 (0.038–0.40) *	1.195	3.313
MTT-VL	0.985 (0.979–0.990)	0.99 (0.986–0.993)	0.137 (0.020–0.250) **	2.326	6.446	0.994 (0.992–0.996)	0.075 (0.05–0.2) **	2.997	8.306	0.993 (0.991–0.995)	0.12 (0.017–0.21) **	2.565	7.109
MTT-VR	0.985 (0.979–0.989)	0.982 (0.975–0.988)	0.26 (0.183–0.334) **	2.314	6.413	0.947 (0.927–0.964)	0.15 (0.084–0.214)	2.476	6.864	0.928 (0.9–0.951)	0.042 (0.04–0.05) *	2.703	7.491
PPH-VL	0.964 (0.950–0.975)	0.99 (0.986–0.993)	0.392 (0.293–0.484) **	1.928	5.343	0.986 (0.981–0.99)	0.225 (0.131–0.315) **	2.022	5.605	0.969 (0.956–0.978)	0.12 (0.011–0.246) *	2.439	6.76
PPH-VR	0.992 (0.989–0.995)	0.987 (0.981–0.991)	0.53 (0.364–0.664) **	1.535	4.256	0.951 (0.931–0.966)	0.384 (0.265–0.491) **	1.577	4.37	0.981 (0.9–0.951)	0.472 (0.291–0.621)	1.656	4.591

Abbreviations: ICC = Intraclass Correlation Coefficient; CI = Confidence Interval; CCC = Concordance Correlation Coefficient; SEM = standard error of measurement; MDC = minimal detectable change; MTT ADD = first metatarsal bone in Adduction, I MTT toward to medial body line; MTT ABD = first metatarsal bone in Abduction, I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexion, I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexion, I MTT away to the floor; MTT VL = first metatarsal bone spins to inside, toward medial body line; MTT VR = first metatarsal bone spin to outside, away to medial body line; PPH AD = proximal phalanx of hallux in adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexion: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexion: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; ±SD = standard deviation. Movements and axes refer to the medial line of the body; all values are expressed in degrees;* p-value < 0.05;** p-value < 0.001.

Static mobility grades of IMTT (sensor 1) and PPH (sensor 2) bones in unshod and 3 cm, 6 cm and 9 cm heel height are summarized in Table 4. The dynamic mobility grades of both sensor 1 and sensor 2 in unshod and 3 cm, 6 cm and 9 cm heel heights in Table 5.

Under static conditions, the PPH abduction increases from $1.01 \pm 0.36°$ to $1.31 \pm 0.46°$ ($p < 0.05$) after wearing shoes with a 3 cm heel. There was a statistically significant valgus of the I MTT with increasing heel heights from $1.77 \pm 0.20°$ without high heels to $2.15 \pm 0.10°$ with 3 cm ($p < 0.001$) high heels. There was increased varus movement from 1.50 ± 0.23 to 3.87 ± 0.20 with 3 cm heels ($p < 0.001$).

In dynamic tests, wearing shoes with 6 cm high heels led to an increase in PPH valgus and abduction deviation from $3.15 \pm 0.10°$ to $3.46 \pm 0.05°$ ($p < 0.05$) and from $1.35 \pm 0.28°$ to $1.69 \pm 0.30°$ ($p < 0.001$), respectively. In addition, PPH had abduction that increased to $1.91 \pm$ with 9 cm heels ($p < 0.001$). On the other hand, I MTT had a valgus decrease with 6 cm heels (from 3.94 ± 0.28 to 3.70 ± 0.13 ($p < 0.001$)) but without any concordance correlation.

Table 4. STATIC mobility grades of first metatarsal (sensor 1) and proximal phalanx of hallux (sensor 2) bones in unshod with heels of 3 cm, 6 cm, and 9 cm high.

Variable	Unshod Mean (°) ± SD (95% CI)	Heel 3 cm Mean (°) ± SD (95% CI)	Heel 6 cm Mean (°) ± SD (95% CI)	Heel 9 cm Mean (°) ± SD (95% CI)	*p*-Value Unshod vs. Heel 3 cm	*p*-Value Unshod vs. Heel 6 cm	*p*-Value Unshod vs. Heel 9 cm
MTT AD	3.78 ± 0.36 (3.70–3.85)	2.14 ± 0.41 (2.05–2.22)	2.04 ± 0.48 (1.93–2.14)	1.20 ± 0.25 (1.14–1.25)	<0.001 **	<0.001 **	<0.001 **
MTT AB	1.32 ± 0.16 (1.28–1.35)	1.74 ± 0.12 (1.71–1.76)	1.36 ± 0.17 (1.32–1.39)	2.14 ± 0.17 (2.10–2.17)	1.115	<0.001 **	<0.05 *
MTT PF	23.88 ± 1.90 (23.46–24.29)	24.12 ± 1.00 (23.90–24.33)	26.69 ± 1.25 (26.41–26.96)	17.68 ± 0.61 (17.54–17.81)	<0.001 **	<0.05 *	<0.001 **
MTT DF	0.66 ± 0.11 (0.63–0.68)	0.57 ± 0.10 (0.54–0.59)	0.62 ± 0.10 (0.59–0.64)	1.45 ± 0.17 (1.41–1.48)	<0.001 **	<0.001 **	0.064
MTT VL	1.77 ± 0.20 (1.72–1.81)	2.15 ± 0.10 (2.12–2.17)	4.00 ± 0.24 (3.94–4.05)	3.46 ± 0.15 (3.42–3.49)	<0.001 **	0.501	0.062
MTT VR	1.50 ± 0.23 (1.44–1.55)	3.87 ± 0.20 (3.82–3.91)	0.93 ± 0.19 (0.88–0.97)	1.76 ± 0.18 (1.72–1.79)	<0.001 **	<0.001 **	<0.001 **
PPH ADD	1.34 ± 0.09 (1.32–1.35)	1.60 ± 0.13 (1.57–1.62)	1.05 ± 0.12 (1.02–1.07)	1.78 ± 0.15 (1.74–1.81)	0.239	0.268	0.216
PPH ABD	1.01 ± 0.36 (0.93–1.08)	2.32 ± 0.18 (2.28–2.35)	1.31 ± 0.46 (1.20–1.41)	1.17 ± 0.15 (1.13–1.20)	0.907	<0.05 *	0.121
PPH PF	3.32 ± 0.29 (3.25–3.38)	3.90 ± 0.35 (3.82–3.97)	4.37 ± 0.42 (4.27–4.46)	4.64 ± 0.43 (4.54–4.73)	<0.05 *	0.893	0.642
PPH DF	2.23 ± 0.37 (2.14–2.31)	2.40 ± 0.43 (2.30–2.49)	1.94 ± 0.55 (1.81–2.06)	5.79 ± 0.48 (5.68–5.89)	<0.001 **	<0.05 *	<0.001 **
PPH VL	2.57 ± 0.12 (2.54–2.59)	2.31 ± 0.12 (2.28–2.33)	3.07 ± 0.16 (3.03–3.10)	3.52 ± 0.09 (3.50–3.53)	0.555	0.102	0.419
PPH VR	1.89 ± 0.22 (1.84–1.93)	2.70 ± 0.19 (2.65–2.74)	1.29 ± 0.18 (1.25–1.32)	1.78 ± 0.15 (1.74–1.81)	<0.05 *	<0.05 *	0.169

Abbreviations: MTT ADD = first metatarsal bone in Adduction: I MTT toward to medial body line; MTT ABD = first metatarsal bone in abduction: I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexion: I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexion: I MTT away to the floor; MTT VL = first metatarsal bone spins to inside: I MTT toward medial body line; MTT VR = first metatarsal bone spin to outside: I MTT away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; PPH ADD = proximal phalanx of hallux in adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in Abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexion: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexion: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; ±SD = standard deviation; *p*-value < 0.05 * (within a 95% confidence interval) was considered statistically significant; *p*-value < 0.001 ** (within a 95% confidence interval) was considered statistically strong significant; Movements and axes refer to the medial line of the body; all values are expressed in degrees.

Table 5. DYNAMIC mobility grades of first metatarsal (sensor 1) and proximal phalanx of hallux (sensor 2) bones in unshod as well as heels 3 cm, 6 cm and 9 cm high.

Variable	Unshod Mean (°) ± SD (95% CI)	Heel 3 cm Mean (°) ± SD (95% CI)	Heel 6 cm Mean (°) ± SD (95% CI)	Heel 9 cm Mean (°) ± SD (95% CI)	*p*-Value Unshod vs. Heel 3 cm	*p*-Value Unshod vs. Heel 6 cm	*p*-Value Unshod vs. Heel 9 cm
MTT ADD	4.97 ± 0.36 (4.89–5.04)	1.42 ± 0.35 (1.34–1.49)	2.51 ± 0.30 (2.43–2.59)	2.25 ± 030 (2.17–2.32)	0.226	0.088	0.852
MTT ABD	1.89 ± 0.17 (1.85–1.92)	0.76 ± 0.10 (0.73–0.78)	1.04 ± 0.08 (1.02–1.06)	2.01 ± 0.14 (1.96–2.04)	0.211	<0.001 **	<0.05 *
MTT PF	38.30 ± 3.38 (37.55–39.04)	35.06 ± 1.99 (34.62–35.49)	36.13 ± 2.67 (35.54–36.71)	44.88 ± 1.40 (44.57–45.18)	<0.001 **	<0.001 **	<0.05 *
MTT DF	0.97 ± 0.15 (0.93–1.00)	0.83 ± 0.10 (0.80–0.85)	0.86 ± 0.13 (0.83–0.89)	0.81 ± 0.14 (0.76–0.86)	<0.001 **	<0.001 **	<0.001 **
MTT VL	3.94 ± 0.28 (3.87–4.00)	2.88 ± 0.10 (2.85–2.90)	3.70 ± 0.13 (3.67–3.73)	3.60 ± 0.08 (3.58–3.61)	0.724	<0.001 **	0.526
MTT VR	3.92 ± 0.19 (3.85–3.94)	1.11 ± 0.21 (1.06–1.16)	1.08 ± 0.15 (1.04–1.11)	2.27 ± 0.26 (2.21–2.32)	0.98	0.25	0.653
PPH ADD	2.39 ± 0.13 (2.36–2.41)	0.87 ± 0.15 (0.83–0.90)	1.31 ± 0.14 (1.26–1.33)	1.12 ± 0.19 (1.07–1.16)	<0.001 **	0.088	<0.001 **
PPH ABD	1.35 ± 0.28 (1.28–1.41)	4.56 ± 0.50 (4.45–4.66)	1.69 ± 0.30 (1.66–1.71)	1.91 ± 0.24 (1.85–1.96)	0.486	<0.001 **	<0.05 *
PPH PF	8.89 ± 0.48 (8.78–8.99)	4.66 ± 0.10 (4.63–4.69)	8.87 ± 0.48 (8.76–8.97)	2.25 ± 0.21 (2.20–2.29)	0.372	<0.001 **	0.605
PPH DF	1.97 ± 0.39 (1.88–2.05)	1.67 ± 0.20 (1.62–1.71)	1.97 ± 0.22 (1.92–2.01)	1.56 ± 0.27 (1.50–1.63)	0.051	<0.05 *	<0.05 *
PPH VL	3.15 ± 0.10 (3.12–3.17)	3 ± 0.11 (2.97–3.02)	3.46 ± 0.05 (3.45–3.46)	4.31 ± 0.04 (4.30–4.31)	0.411	<0.05 *	0.873
PPH VR	1.56 ± 0.33 (1.48–1.63)	1.26 ± 0.27 (1.20–1.31)	1.26 ± 0.22 (1.21–1.30)	1.68 ± 0.29 (1.61–1.74)	<0.001 **	<0.001 **	<0.05 *

Abbreviations: MTT ADD = first metatarsal bone in Adduction: I MTT toward to medial body line; MTT ABD = first metatarsal bone in Abduction: I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexión: I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexión: I MTT away to the floor; MTT VL = first metatarsal bone spins to inside: I MTT toward medial body line; MTT VR = first metatarsal bone spin to outside: I MTT away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; PPH ADD = proximal phalanx of hallux in Adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in Abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexión: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexión: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; ±SD = standard deviation; *p*-value < 0.05 * (within a 95% confidence interval) was considered statistically significant; *p*-value < 0.001 ** (within a 95% confidence interval) was considered statistically strong significant; Movements and axes refer to the medial line of the body; all values are expressed in degrees.

Finally, Tables 6 and 7 summarize the Spearman's Rho correlation coefficients of static and dynamic conditions between the height of the heels and the movements of PPH and I MTT, respectively.

In static tests, adduction of IMTT had a statistically significant positive correlation while wearing 6 cm heels versus unshod condition (0.5, $p < 0.001$) as well as abduction (0.347, $p < 0.001$). Valgus values had a statistically significant positive correlation too under 3 cm high heel (0.51, $p < 0.001$). Regarding PPH abduction, there was a statistical positive correlation in 6 cm heels (0.261, $p < 0.05$) as well as a statistically significant negative correlation to adduction under 3 cm and 6 cm of high heels ($p < 0.05$).

During dynamic tests, PPH abduction had a statistically significant positive correlation with heels of 6 cm (0.527, $p < 0.001$) combined with a statistically significant inverse correlation in adduction (−0.278, $p < 0.05$) with 6 cm; there was a significant positive correlation with valgus values (0.242, $p < 0.05$).

Table 6. STATIC Spearman-Rho correlation coefficients between variables.

Variable	Unshod vs. Heel 3 cm (ρ-Coefficient)	*p*-Value	Unshod vs. Heel 6 cm (ρ-Coefficient)	*p*-Value	Unshod vs. Heel 9 cm (ρ-Coefficient)	*p*-Value
MTT ADD	0.256	0.3	0.5	<0.001 **	0.521	<0.001 **
MTT ABD	−0.69	0.77	0.347	<0.001 **	0.228	<0.05 *
MTT PF	−0.189	−0.192	0.723	0.061	−0.125	0.75
MTT DF	0.173	0.124	−0.639	0.072	0.146	0.6
MTT VL	0.510	<0.001 **	−0.370	0.5	0.487	0.336
MTT VR	0.26	0.122	−0.401	<0.001 **	0.587	0.2
PPH ADD	0.391	0.46	0.150	−0.112	0.0984	0.226
PPH ABD	0.412	0.477	0.261	<0.05 *	0.58	0.398
PPH PF	0.184	0.103	−0.017	0.116	−0.189	0.69
PPH DF	0.054	0.111	−0.079	0.086	−0.141	0.333
PPH VL	0.179	−0.037	0.71	0.406	0.183	522
PPH VR	−0.306	<0.05 *	−0.222	<0.05 *	0.36	0.922

Abbreviations: ρ-coefficient = Spearman Rho coefficient; MTT ADD = first metatarsal bone in adduction: I MTT toward to medial body line; MTT ABD = first metatarsal bone in abduction: I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexion: I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexion: I MTT away to the floor; MTT VL = first metatarsal bone spins to inside: I MTT toward medial body line; MTT VR = first metatarsal bone spin to outside: I MTT away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; PPH ADD = proximal phalanx of hallux in adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexion: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexion: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; ±SD = standard deviation; *p*-value < 0.05 * (within a 95% confidence interval) was considered statistically significant; *p*-value < 0.001 ** (within a 95% confidence interval) was considered statistically strong significant; Movements and axes refer to the medial line of the body; all values are expressed in degrees.

Table 7. DINADYNAMIC Spearman-Rho correlation coefficients between variables.

Variable	Unshod vs. Heel 3 cm (ρ-Coefficient)	*p*-Value	Unshod vs. Heel 6 cm (ρ-Coefficient)	*p*-Value	Unshod vs. Heel 9 cm (ρ-Coefficient)	*p*-Value
MTT ADD	−0.276	<0.05 *	0.025	0.061	0.181	0.3
MTT ABD	0.306	0.189	0.606	0.138	0.258	0.224
MTT PF	0.086	−0.193	0.026	−0.139	−0.119	0.148
MTT DF	0.056	0.203	0.043	0.128	−0.142	0.956
MTT VL	0.167	0.128	−0.104	0.134	0.523	<0.001 **
MTT VR	−0.071	0.0733	−0.088	0.0621	0.666	0.459
PPH ADD	−0.364	<0.001 **	−0.278	<0.05 *	0.856	0.1
PPH ABD	0.71	0.25	0.527	<0.001 **	0.375	0.476
PPH PF	−0.128	−0.005	−0.028	0.148	−0.109	0.945
PPH DF	0.1	−0.116	0.148	0.058	0.167	0.8
PPH VL	−0.169	0.150	0.242	<0.05 *	0.154	0.36
PPH VR	−0.522	<0.001 **	0.108	−0.089	0.146	0.221

Abbreviations: ρ-coefficient = Spearman Rho coefficient; MTT ADD = first metatarsal bone in Adduction: I MTT toward to medial body line; MTT ABD = first metatarsal bone in abduction: I MTT away to medial body line; MTT PF = first metatarsal bone in plantarflexion: I MTT toward to the floor; MTT DF = first metatarsal bone in dorsiflexion: I MTT away to the floor; MTT VL = first metatarsal bone spins to inside: I MTT toward medial body line; MTT VR = first metatarsal bone spin to outside: I MTT away to medial body line; UNSHOD = barefoot participants without any height in shoe heel; PPH ADD = proximal phalanx of hallux in adduction: PPH toward to medial body line; PPH ABD = proximal phalanx of hallux in abduction: PPH away to medial body line; PPH PF = proximal phalanx of hallux in plantarflexion: PPH toward to the floor; PPH DF = proximal phalanx of hallux in dorsiflexion: PPH away to the floor; PPH VL = proximal phalanx of hallux spins to inside: PPH toward medial body line; PPH VR = proximal phalanx of hallux spin to outside: PPH away to medial body line; ±SD = standard deviation; *p*-value < 0.05 * (within a 95% confidence interval) was considered statistically significant; *p*-value < 0.001 ** (within a 95% confidence interval) was considered statistically strong significant; Movements and axes refer to the medial line of the body; all values are expressed in degrees.

4. Discussion

HAV is a forefoot pathology related to PPH in valgus and abduction deviations plus the IMTT bone in adduction deviation [1,2,51]. Previous studies [18,52–54] have shown the relationship between HAV development and the narrow toe tip footwear typical of the high heel shoes; other groups [1,27] speculated that high heels have an etiology factor in HAV development with only one prior review [12] considering the isolated high heels's effect on first MPJ deviation on its conclusions but without any further study. This research was the first study to use the Polhemus Fastrack® to assess the effects of high heels on static and dynamic conditions on PPH and IMTT. Thus, many statistically significant variables have been obtained, but of the 96 variables studied here across 80 subjects, only a few had statistically significant correlations. Due to the independent evaluation of the bones, we could not establish a direct discussion with the findings of other authors who interpreted HAV as a global forefoot disease [34]. In addition, this work has studied a healthy population without any limitation of mobility in their joints. Thus, a small difference in segment movements was expected.

Our results on the effect of high heels on PPH during dynamic testing showed that a high heel of 6 cm had a statistically significant increase in abduction in the transverse plane and valgus movement in the frontal plane through IMPJ. There was positive correlation of these values as well as a statistically significant reduction in the adduction; there was negative correlation to the barefoot condition. In addition, during static tests, there was a statistically significant abduction increase in PPH for high heels 6 cm or higher with a corresponding positive correlation; therefore, abduction is the only movement that appeared to have a positive correlation and statistical significance in both static and dynamic tests. Thus, we conclude that heels over 6 cm correlated with an abduction effect on PPH without the narrow shoe box interference. This agrees with arguments on the biomechanical development of HAV processes that claimed that the PPH was the first precursor bone segment to begin the HAV pathology [51,55–57] due to medial capsular tension ligaments that become hyper-elastic and let the PPH proceed to abduction deviation [7]. The PPH then has a strong push forward to the IMTT in the push off phase that can lead to adduction deviation.

More recently, Wang et al. [34] reported an increase in forefoot abduction while wearing 5 cm high-heeled shoes during walking vs. barefoot arguing that the squeezing effect of the high heels on the foot had a displacement toward the toe tip. This produced valgus and abduction of PPH. We agree with these conclusions, but we showed that the foot produces this "abductor effect" on PPH; it is not from the narrow box of the shoes. In addition, we obtained a large increase in the value of PPH abduction with a 3 cm heel. This result was not statistically significant.

The effects of high heels had contradictory effects on IMTT movements. There were no statistically significant results to justify its implication on HAV development in contrast with other groups that identified IMTT adduction and valgus [1,2] as well as inclination of the IMTT axis as risk factors of bunion [58] or IMCJ hypermobility [59]; both movements on transverse and frontal planes are under doubt because of a lack of objective data [60]. Surprisingly, it seemed that wearing any high heel might decrease valgus deviation of IMTT although this condition only had statistical significance with 6 cm heels; there was no positive correlation. The absence of concrete IMTT values related with typical HAV development suggests that PPH may be the principal bone to start the pathological process. This agrees with a study that identified the presence of HAV with greater reduction in size of the adductor hallucis muscle [61] as one of the most important muscles to balance the PPH.

In contrast to previous studies [62] that found no association between footwear characteristics (heel height and narrow box) and HAV development, our cohort had (18–38 years) had movement deviation of PPH in heels over 6 cm. This agrees with other studies where older women reported HAV. They wore shoes with heels over 5 cm [12]. This work showed data on IMTT and PPH from the two different static and dynamic conditions and 3 kinds of high heels. Thus, we selected and summarized heel heights and determined that high heels could develop HAV. 6 cm was the common height for both static and dynamic situations; these different variables converged to induce HAV development.

Most of the main limitations in other similar studies were equivocal results secondary to small sample size [19], differences in anthropometric characteristics of the subjects groups [63], or the inclusion of participants wearing their own high-heeled shoes [64]. This leads to a heterogeneous sample [19]. We studied a homogeneous sample that improved the measurement conditions of other groups that also failed to show ICC, CCC, SEM or MDC values [10,34]; our data had low to moderate correlation, and we considered these statistical parameters.

Coupling relationships between hindfoot inversion/eversion and forefoot abduction/adduction ($R^2 = 0.5$) and hallux dorsiflexion/plantarflexion ($R^2 = 0.7$) were the only prior references found on this topic [65]. It had similar Spearman values in that work but did show any contrast between data because the authors did not assess individual movements of any segment bone as PPH or I MTT, like in the present work. These are the only specific correlations done relative to PPH and IMTT in the literature and confirm the low Spearman correlations of our values.

5. Limitations

Bone segment measurements were performed with an electronic Polhemus Fastrack® goniometer and showed current instability due to inherent human gait fluctuations and sensor noise. The MDC values were higher than the grades obtained in the dynamic PPH valgus and abduction as well as in the static abduction PPH variables. Considering that the percent error set for the device was around 1.6%—and considering that the ±SD and SEM values obtained in the study were under this 1.6%—these results are considered to be statistically valid but with caution.

On the other hand, we were not able to study the effect of order on our sample because didn't write the different orders of each one neither the number of these selections. We assumed the possible "order effect" as "perfectly balanced" because all the study subgroups of each station have the same number of subjects and this can dilutes the "order effect". Future study design should include a section of studying the effect of order of experiments.

6. Conclusions

Wearing shoes with heels over 6 cm may produce a valgus and abduction increase in PPH movement. This abduction is specifically detected in the development of HAV pathology. Future cohort studies will be required to clarify the time period that is needed to develop HAV pathology related to high-heeled shoes.

Author Contributions: Conceptualization, R.S.-G. and R.B.d.B.-V.; Methodology, R.S.-G. and R.B.d.B.-V.; Software, E.M.M.-J.; Validation, C.C.-L. and C.R.-M.; Formal Analysis, C.C.-L.; Investigation, M.E.L.-I.; Resources, P.P.-L. and D.L.-L.; Data Curation, P.P.-L. and D.L.-L.; Writing-Original Draft Preparation, R.S.-G.; Writing-Review & Editing, R.S.-G., R.B.d.B.-V., C.C.-L., C.R.-M., E.M.M.-J., M.E.L.-I., P.P.-L. and D.L.-L.; Visualization, D.L.-L. and E.M.M.-J.; Supervision, P.P.-L. and M.E.L.-I.; Project Administration, C.C.-L. and C.R.-M.; Funding Acquisition, anyone.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mann, R.A.; Coughlin, M.J. Hallux valgus—Etiology, anatomy, treatment and surgical considerations. *Clin. Orthop. Relat. Res.* **1981**, *157*, 31–41. [CrossRef]
2. Piqué-Vidal, C.V.J. A geometric analysis of hallux valgus: Correlation with clinical assessment of severity. *J. Foot Ankle Res.* **2009**, *2*, 15. [CrossRef] [PubMed]
3. Menz, H.B.; Munteanu, S.E. Radiographic validation of the Manchester scale for the classification of hallux valgus deformity. *Rheumatology* **2005**, *44*, 1061–1066. [CrossRef] [PubMed]
4. Wolff, J. The classic: On the inner architecture of bones and its importance for bone growth. 1870. *Clin. Orthop. Relat. Res.* **2010**, *468*, 1056–1065. [CrossRef] [PubMed]

5. Root, M.L.; Orien, W.P.W.H. *Normal and Abnormal Function of the Foot*; Corp, C.B.: Los Angeles, CA, USA, 1977; Volume II.

6. McBride, I.D.; Wyss, U.P.; Cooke, T.D.; Murphy, L.; Phillips, J.; Olney, S.J. First metatarsophalangeal joint reaction forces during high-heel gait. *Foot Ankle* **1991**, *11*, 282–288. [CrossRef] [PubMed]

7. Hicks, J. The mechanics of the foot. II. The plantar aponeurosis and the arch. *J. Anat.* **1954**, *88*, 25–30. [PubMed]

8. Faber, F.W.; Kleinrensink, G.J.; Mulder, P.G.; Verhaar, J.A. Mobility of the first tarsometatarsal joint in hallux valgus patients: A radiographic analysis. *Foot Ankle Int.* **2001**, *22*, 965–969. [CrossRef] [PubMed]

9. Nix, S.; Smith, M.; Vicenzino, B. Prevalence of hallux valgus in the general population: A systematic review and meta-analysis. *J. Foot Ankle Res.* **2010**, *3*, 21. [CrossRef]

10. Kawakami, W.; Takahashi, M.; Iwamoto, Y.; Shinakoda, K. Coordination Among Shank, Rearfoot, Midfoot, and Forefoot Kinematic Movement During Gait in Individuals with Hallux Valgus. *J. Appl. Biomech.* **2018**, *35*, 44–51. [CrossRef]

11. Coughlin, M.J.; Jones, C.P. Hallux Valgus: Demographics, Etiology, and Radiographic Assessment. *Foot Ankle Int.* **2007**, *28*, 759–777. [CrossRef]

12. Nguyen, U.S.; Hillstrom, H.J.; Li, W.; Dufour, A.B.; Kiel, D.P.; Procter-Gray, E.; Gagnon, M.M.; Hannan, M.T. Factors associated with hallux valgus in a population-based study of older women and men: The MOBILIZE Boston Study. *Osteoarthr. Cartil.* **2010**, *18*, 41–46. [CrossRef] [PubMed]

13. Owoeye, B.A.; Akinbo, S.R.; Aiyegbusi, A.L.; Ogunsola, M.O. Prevalence of hallux valgus among youth population in Lagos, Nigeria. *Niger. Postgrad. Med. J.* **2011**, *18*, 51–55. [PubMed]

14. Dawson, J.; Thorogood, M.; Marks, S.A.; Juszczak, E.; Dodd, C.; Lavis, G.; Fitzpatrick, R. The prevalence of foot problems in older women: A cause for concern. *J. Public Health Med.* **2002**, *24*, 77–84. [CrossRef] [PubMed]

15. Menz, H.B.; Lord, S.R. Foot problems, functional impairment, and falls in older people. *J. Am. Podiatr. Med. Assoc.* **1999**, *89*, 458–467. [CrossRef] [PubMed]

16. Glassy, C.M.; Glassy, M.S.; Guggenheim, C. Relationship between self-reported high-heeled shoe use and bone mineral density using quantitative ultrasound at a community health fair. *Clin. Rheumatol.* **2013**, *32*, 37–41. [CrossRef] [PubMed]

17. Gerber, S.B.; Costa, R.V.; Grecco, L.A.C.; Pasini, H.; Marconi, N.F.; Oliveira, C.S. Interference of high-heeled shoes in static balance among young women. *Hum. Mov. Sci.* **2012**, *31*, 1247–1252. [CrossRef] [PubMed]

18. Snow, R.E.; Williams, K.R. High heeled shoes: Their effect on center of mass position, posture, three-dimensional kinematics, rearfoot motion, and ground reaction forces. *Arch. Phys. Med. Rehabil.* **1994**, *75*, 568–576.

19. Cronin, N.J. The effects of high heeled shoes on female gait: A review. *J. Electromyogr. Kinesiol.* **2014**, *24*, 258–263. [CrossRef]

20. Mika, A.; Oleksy, Ł.; Kielnar, R.; Świerczek, M. The influence of high- and low-heeled shoes on balance in young women. *Acta Bioeng. Biomech.* **2016**, *18*, 97–103.

21. Barnish, M.S.; Barnish, J. High-heeled shoes and musculoskeletal injuries: A narrative systematic review. *BMJ Open* **2016**, *6*, e010053. [CrossRef]

22. Hong, W.-H.; Lee, Y.-H.; Lin, Y.-H.; Tang, S.F.T.; Chen, H.-C. Effect of shoe heel height and total-contact insert on muscle loading and foot stability while walking. *Foot Ankle Int.* **2013**, *34*, 273–281. [CrossRef] [PubMed]

23. Nyska, M.; McCabe, C.; Linge, K.; Klenerman, L. Plantar foot pressures during treadmill walking with high-heel and low-heel shoes. *Foot Ankle Int.* **1996**, *17*, 662–666. [CrossRef] [PubMed]

24. Yung-Hui, L.; Wei-Hsien, H. Effects of shoe inserts and heel height on foot pressure, impact force, and perceived comfort during walking. *Appl. Ergon.* **2005**, *36*, 355–362. [CrossRef]

25. Penny, J.Ø.; Speedtsberg, M.B.; Kallemose, T.; Bencke, J. Can an off-the-rack orthotic stiletto alter pressure and comfort scores in the forefoot, arch and heel? *Ergonomics* **2018**, *61*, 1130–1138. [CrossRef] [PubMed]

26. Dufour, A.B.; Casey, V.A.; Golightly, Y.M.; Hannan, M.T. Characteristics associated with hallux valgus in a population-based foot study of older adults. *Arthritis Care Res.* **2014**, *66*, 1880–1886. [CrossRef] [PubMed]

27. Wülker, N.; Mittag, F. The treatment of hallux valgus. *Dtsch. Arztebl. Int.* **2012**, *109*, 857–867. [CrossRef] [PubMed]

28. Mann, R.A.; Coughlin, M.J. *Surgery of the Foot and Ankle*, 6th ed.; Mosby: St Louis, MO, USA, 1992.

29. Swanson, J.E.; Stoltman, M.G.; Oyen, C.R.; Mohrbacher, J.A.; Orandi, A.; Olson, J.M.; Glasoe, W.M. Comparison of 2D-3D Measurements of Hallux and First Ray Sagittal Motion in Patients with and Without Hallux Valgus. *Foot Ankle Int.* **2016**, *37*, 227–232. [CrossRef] [PubMed]

30. Deschamps, K.; Birch, I.; Desloovere, K.; Matricali, G.A. The impact of hallux valgus on foot kinematics: A cross-sectional, comparative study. *Gait Posture* **2010**, *32*, 102–106. [CrossRef]

31. Gur, G.; Ozkal, O.; Dilek, B.; Aksoy, S.; Bek, N.; Yakut, Y. Effects of Corrective Taping on Balance and Gait in Patients with Hallux Valgus. *Foot Ankle Int.* **2017**, *38*, 532–540. [CrossRef]

32. Klugarova, J.; Janura, M.; Svoboda, Z.; Sos, Z.; Stergiou, N.; Klugar, M. Hallux valgus surgery affects kinematic parameters during gait. *Clin. Biomech.* **2016**, *40*, 20–26. [CrossRef]

33. Nawoczenski, D.A.; Ludewig, P.M. The Effect of Forefoot and Arch Posting Orthotic Designs on First Metatarsophalangeal Joint Kinematics During Gait. *J. Orthop. Sport Phys. Ther.* **2004**, *34*, 317–327. [CrossRef]

34. Wang, M.; Gu, Y.; Baker, J.S. Analysis of foot kinematics wearing high heels using the Oxford foot model. *Technol. Health Care* **2018**, *26*, 815–823. [CrossRef]

35. Becerro de Bengoa Vallejo, R.; Gomez, R.S.; Losa Iglesias, M.E. Clinical improvement in functional hallux limitus using a cut-out orthosis. *Prosthet. Orthot. Int.* **2016**, *40*, 215–223. [CrossRef]

36. Hong, W.-H.; Lee, Y.-H.; Chen, H.-C.; Pei, Y.-C.; Wu, C.-Y. Influence of heel height and shoe insert on comfort perception and biomechanical performance of young female adults during walking. *Foot Ankle Int.* **2005**, *26*, 1042–1048. [CrossRef]

37. Buell, T.; Green, D.R.; Risser, J. Measurement of the first metatarsophalangeal joint range of motion. *J. Am. Podiatr. Med. Assoc.* **1988**, *78*, 439–448. [CrossRef]

38. Umberger, B.R.; Nawoczenski, D.A.; Baumhauer, J.F. Reliability and validity of first metatarsophalangeal joint orientation measured with an electromagnetic tracking device. *Clin. Biomech.* **1999**, *14*, 74–76. [CrossRef]

39. Halstead, J.; Redmond, A.C. Weight-bearing passive dorsiflexion of the hallux in standing is not related to hallux dorsiflexion during walking. *J. Orthop. Sports Phys. Ther.* **2006**, *36*, 550–556. [CrossRef]

40. Du, W.-Y.; Huang, T.-S.; Hsu, K.-C.; Lin, J.-J. Measurement of scapular medial border and inferior angle prominence using a novel scapulometer: A reliability and validity study. *Musculoskelet. Sci. Pract.* **2017**, *32*, 120–126. [CrossRef]

41. Nawoczenski, D.A.; Cook, T.M.; Saltzman, C.L. The Effect of Foot Orthotics on Three-Dimensional Kinematics of the Leg and Rearfoot During Running. *J. Orthop. Sport Phys. Ther.* **1995**, *21*, 317–327. [CrossRef]

42. Milne, A.D.; Chess, D.G.; Johnson, J.A.; King, G.J. Accuracy of an electromagnetic tracking device: A study of the optimal range and metal interference. *J. Biomech.* **1996**, *29*, 791–793. [CrossRef]

43. Welsh, B.J.; Redmond, A.C.; Chockalingam, N.; Keenan, A.-M. A case-series study to explore the efficacy of foot orthoses in treating first metatarsophalangeal joint pain. *J. Foot Ankle Res.* **2010**, *3*, 17. [CrossRef] [PubMed]

44. Shereff, M.J.; Bejjani, F.J.; Kummer, F.J. Kinematics of the first metatarsophalangeal joint. *J. Bone Jt. Surg. Am.* **1986**, *68*, 392–398. [CrossRef]

45. Nawoczenski, D.A.; Baumhauer, J.F.; Umberger, B.R. Relationship between clinical measurements and motion of the first metatarsophalangeal joint during gait. *J. Bone Jt. Surg. Am.* **1999**, *81*, 370–376. [CrossRef]

46. Ramanathan, A.K.; John, M.C.; Arnold, G.P.; Cochrane, L.; Abboud, R.J. The effects of off-the-shelf in-shoe heel inserts on forefoot plantar pressure. *Gait Posture* **2008**, *28*, 533–537. [CrossRef] [PubMed]

47. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]

48. Li, L. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, *45*, 255–268.

49. Jacobson, N.; Truax, P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* **1991**, *59*, 12–19. [CrossRef] [PubMed]

50. Portney, L.; Watkins, M. *Foundations of Clinical Research: Applications to Practice*, 3rd ed.; Hall PP: Upper Saddle River, NJ, USA, 2009.

51. Eustace, S.; O'Byrne, J.; Stack, J.; Stephens, M.M. Radiographic features that enable assessment of first metatarsal rotation: The role of pronation in hallux valgus. *Skelet. Radiol.* **1993**, *22*, 153–156. [CrossRef]

52. Frey, C. Foot health and shoewear for women. *Clin. Orthop. Relat. Res.* **2000**, *372*, 32–44. [CrossRef]

53. Ashizawa, K.; Kumakura, C.; Kusumoto, A.; Narasaki, S. Relative foot size and shape to general body size in Javanese, Filipinas and Japanese with special reference to habitual footwear types. *Ann. Hum. Biol.* **1997**, *24*, 117–129. [CrossRef]

54. Menz, H.B.; Morris, M.E. Footwear Characteristics and Foot Problems in Older People. *Gerontology* **2005**, *51*, 346–351. [CrossRef] [PubMed]

55. Saltzman, C.L.; Aper, R.L.; Brown, T.D. Anatomic determinants of first metatarsophalangeal flexion moments in hallux valgus. *Clin. Orthop. Relat. Res.* **1997**, *339*, 261–269. [CrossRef]

56. Arinci İncel, N.; Genç, H.; Erdem, H.R.; Yorgancioglu, Z.R. Muscle Imbalance in Hallux Valgus. *Am. J. Phys. Med. Rehabil.* **2003**, *82*, 345–349. [CrossRef]

57. Shereff, M.J. Pathophysiology, anatomy, and biomechanics of hallux valgus. *Orthopedics* **1990**, *13*, 939–945. [PubMed]

58. Glasoe, W.M.; Jensen, D.D.; Kampa, B.B.; Karg, L.K.; Krych, A.R.; Pena, F.A.; Ludewig, P.M. First Ray Kinematics in Women with Rheumatoid Arthritis and Bunion Deformity: A Gait Simulation Imaging Study. *Arthritis Care Res.* **2014**, *66*, 837–843. [CrossRef] [PubMed]

59. Lapidus, P.W. A quarter of a century of experience with the operative correction of the metatarsus varus primus in hallux valgus. *Bull. Hosp. Jt. Dis.* **1956**, *17*, 404–421.

60. Doty, J.F.; Coughlin, M.J. Hallux valgus and hypermobility of the first ray: Facts and fiction. *Int. Orthop.* **2013**, *37*, 1655–1660. [CrossRef]

61. Aiyer, A.; Stewart, S.; Rome, K. The effect of age on muscle characteristics of the abductor hallucis in people with hallux valgus: A cross-sectional observational study. *J. Foot Ankle Res.* **2015**, *8*, 19. [CrossRef]

62. Menz, H.B.; Roddy, E.; Marshall, M.; Thomas, M.J.; Rathod, T.; Peat, G.M.; Croft, P.R. Epidemiology of Shoe Wearing Patterns Over Time in Older Women: Associations with Foot Pain and Hallux Valgus. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2016**, *71*, 1682–1687. [CrossRef]

63. Cronin, N.J.; Barrett, R.S.; Carty, C.P. Long-term use of high-heeled shoes alters the neuromechanics of human walking. *J. Appl. Physiol.* **2012**, *112*, 1054–1058. [CrossRef]

64. Csapo, R.; Maganaris, C.N.; Seynnes, O.R.; Narici, M.V. On muscle, tendon and high heels. *J. Exp. Biol.* **2010**, *213 Pt 15*, 2582–2588. [CrossRef]

65. Dubbeldam, R.; Nester, C.; Nene, A.V.; Hermens, H.J.; Buurke, J.H. Kinematic coupling relationships exist between non-adjacent segments of the foot and ankle of healthy subjects. *Gait Posture* **2013**, *37*, 159–164. [CrossRef] [PubMed]

Ultrasonography Features of the Plantar Fascia Complex in Patients with Chronic Non-Insertional Achilles Tendinopathy: A Case-Control Study

Carlos Romero-Morales [1], Pedro Javier Martín-Llantino [1], César Calvo-Lobo [2], Daniel López-López [3,*], Rubén Sánchez-Gómez [1,4], Blanca De-La-Cruz-Torres [5] and David Rodríguez-Sanz [1,4]

[1] Faculty of Sport Sciences, Universidad Europea de Madrid, Villaviciosa de Odón, 28670 Madrid, Spain; carlos.romero@universidadeuropea.es (C.R.-M.); pejamalla@gmail.com (P.J.M.-L.); rusanc02@ucm.es (R.S.-G.); davidrodriguezsanz@ucm.es (D.R.-S.)

[2] Faculty of Health Sciences, Institute of Biomedicine (IBIOMED), Universidad de León, 24401 Ponferrada, Spain; ccall@unileon.es

[3] Research, Health and Podiatry Unit, Department of Health Sciences, Faculty of Nursing and Podiatry, Universidade da Coruña, 15403 Ferrol, Spain

[4] Facultad de Enfermería, Fisioterapia y Podología, Universidad Complutense de Madrid, 28040 Madrid, Spain

[5] Department of Physiotherapy, University of Seville, 41009 Seville, Spain; bcruz@us.es

* Correspondence: daniel.lopez.lopez@udc.es; Tel.: +34-981-337-400 (ext. 3546)

Abstract: Purpose: The goal of the present study was to assess, by ultrasound imaging (USI), the thickness of the plantar fascia (PF) at the insertion of the calcaneus, mid and forefoot fascial locations, and the calcaneal fat pad (CFP) in patients with Achilles tendinopathy (AT). Methods: An observational case-control study. A total sample of 143 individuals from 18 to 55 years was evaluated by USI in the study. The sample was divided into two groups: A group composed of the chronic non-insertional AT (n = 71) and B group comprised by healthy subjects (n = 72). The PF thicknesses at insertion on the calcaneus, midfoot, rearfoot and CFP were evaluated by USI. Results: the CFP and PF at the calcaneus thickness showed statistically significant differences ($P < 0.01$) with a decrease for the tendinopathy group with respect to the control group. For the PF midfoot and forefoot thickness, no significant differences ($P > 0.05$) were observed between groups. Conclusion: The thickness of the PF at the insertion and the CPF is reduced in patients with AT measured by USI.

Keywords: ultrasonography; Achilles tendon; diagnostic; imaging; tendinopathy

1. Introduction

Chronic Achilles tendinopathy (AT) is one of the most common conditions of the ankle and foot, characterized by the combination of pain, swelling and a deficit of functionality in the lower limb [1]. The reported incidence rate was estimated at about 2.35–2.16 per 1000 adults [2]. Recent studies revealed that the presence of disorganization and degeneration in collagen fibers, especially associated with obesity and diabetes disorders, as well as changes in vascularity were risk factors for AT [3]. In addition, changes in tendon thickness and cross-sectional area (CSA) in response to load adaptations were related in patients with AT [4]. Currently, two types of AT were described according to the following locations: Non-insertional (2 to 6 cm from the calcaneus insertion of the tendon) and insertional (at the calcaneus insertion of the tendon).

The Achilles tendon complex may be considered the largest tendon in the body. Its architecture is designed to support tensile loads during its contraction or elongation, and it is susceptible to

overuse injuries [5]. In addition, its location makes it a fundamental structure for ankle mobility and locomotion [6]. The Achilles tendon works in a coordinated manner with the gastrocnemius and soleus muscles [7], and the extrinsic and intrinsic foot muscles [8]. The mechanism of AT remains unclear; Galloway et al. [9] reported that changes in the architecture of the tendon and surrounding structures in response to the mechanical load were related in patients with AT. Those adaptations were associated with changes in plantar fascia (PF) morphology, such as the thickness.

The PF is a soft tissue structure across the plantar foot and provides an important role in ankle and foot biomechanics, being a structure capable of working with greater loads by modifying its thickness and stiffness [10]. For example, Huang et al. [11] reported changes in the thickness of the PF in individuals with pes planus, related to mechanical load changes.

Several authors have studied PF and CPF thickness in different conditions and populations by ultrasound imaging (USI) [12–14]. However, there is a lack of evidence about PF thickness in individuals with AT in comparison with healthy subjects.

Ultrasound imaging (USI) has been employed to assess the thickness and CSA of diverse muscles associated with fascial and musculoskeletal conditions. Regarding the lower limb, Lobo et al. [15] showed a reduction in the thickness and CSA of the flexor hallucis brevis and abductor hallucis in subjects with hallux valgus. The peroneus longus muscle CSA was examined by USI and a reduction was observed in the CSA in individuals with ankle sprains [16]. Angin et al. [10] reported a greater CSA and thickness in PF in subjects with pes planus. In addition, Taniguchi et al. [17] showed a reduced vastus medialis thickness in individuals with knee osteoarthritis. Considering other locations, USI has been useful to evaluate the temporalis, sternocleidomastoid and masseter muscles, showing changes in patients with temporomandibular disorders [18]. CSA of the intrinsic hand muscles has been evaluated by Mohseny et al. [19] with USI in subjects with nerve injuries. In the cervical region, the morphology of the deep cervical muscles were observed by USI while they developed an exercise program in subjects with neck pain [20]. Considering the trunk region, Whittaker et al. [21] related the architecture of the abdominal wall muscles with lumbopelvic pain. Several authors argued that USI is a non-invasive, safe and valid method to examine soft tissues and musculoskeletal disorders [22,23].

For the Achilles tendon complex, prior studies were focused on the tendon thickness and CSA in subjects with AT [13]. However, limited evidence of the surrounding soft tissues has been observed, such as PF and calcaneal fat pad (CFP) by USI [24].

The goal of the present study was to assess, using USI, the thickness of the PF at the insertion of the calcaneus, mid and forefoot fascial locations, and the CFP in patients with AT. We hypothesized that these selected soft tissue structures would demonstrate changes in individuals with AT.

2. Methods

2.1. Design

An observational study was carried out following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [25] guidelines from January to December 2017.

2.2. Participants

A total sample of 143 individuals from 18 to 55 years was involved in the study. The sample was divided in two groups: A group composed of the chronic non-insertional AT (n = 71) and B group comprised of healthy subjects (n = 72). For the AT group, subjects were included if they showed the following features: pain in the mid-portion of the Achilles tendon measuring at least 3 out of 10 points in the pain visual analogue scale (VAS), a decrease in the function and tenderness in the body of the tendon for at least 3 months, and undergoing no physical therapy, analgesic or corticosteroid interventions during the study. Exclusion criteria were as follows: subjects with any systemic disease or infection [26], fracture [27] surgeries, plantar orthoses and lower limb disturbances within the last year [28].

2.3. Calculation of the Sample Size

G*Power software was employed for the sample size calculation in order to measure the difference between the chronic non-insertional AT group and control group using the CFP thickness (mm) variable of a pilot study ($n = 20$) that was divided into two groups (mean ± SD): 10 subjects with chronic non-insertional AT (7.27 ± 1.97), and 10 subjects for the healthy group (8.51 ± 0.51). For the sample size calculation, a power of 0.80, an α error of 0.05 and an effect size of 0.86 with a one-tailed hypothesis were employed. In conclusion, a sample of 36 was calculated. Nevertheless, we could include a sample of 143 individuals.

2.4. Ethical Statement

The study was approved by the La Princesa Hospital ethics committee (Madrid, Spain). All the subjects included in the study signed the informed consent form. The research is in accordance with the Declaration of Helsinki for human experimentation (Project identification code: 2828A).

2.5. Ultrasonography Measurements

The USI assessments were carried out using a high-quality system LogiQ P7 (GE Healthcare; UK) with a 4 to 13 MHz linear transducer (L6-12 RS type, 38 mm footprint). All ultrasound evaluations of the PF were carried out in a supine position with the transducer placed in direct contact with the skin. The PF was scanned in a longitudinal view in all locations. Firstly, for the enthesis at the calcaneus location, the transducer was placed on the line between the medial calcaneal tubercle and the second toe over the PF (Figure 1, A). Secondly, for the midfoot PF assessment, the transducer was located on the same scanning line at the navicular tubercle (Figure 1, B). Thirdly, for the metatarsal region of the PF, the transducer was located on the same longitudinal line near to the second metatarsal head (Figure 1, C) [29]. For the ultrasound examination of the CFP, the subjects were placed in a prone position and the transducer was placed longitudinally at the midpoint of the heel (Figure 2) [24]. According to López et al. [30], the evaluator both flexed and extended the big toe while palpating the subject´s PF to allow accurate identification of the calcaneal tuberosity. All the examinations were carried out by a physiotherapist (P.M.L.) with more than 5 years of experience in ultrasonography. The final scores were collected by the mean of 3 repeated values for each measurement with the ImageJ software (Bethesda, MD, USA).
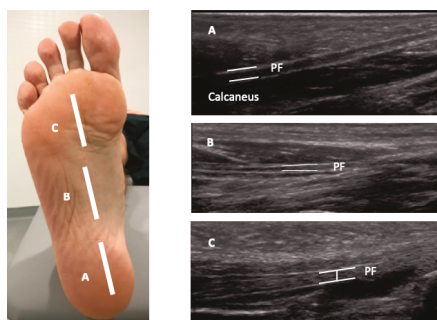

Figure 1. Ultrasound imaging thickness for the plantar fascia at the insertion, midfoot and forefoot locations in an individual with Achilles tendinopathy (AT). Abbreviations: PF, plantar fascia.
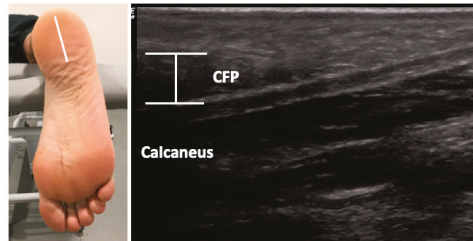
Figure 2. Ultrasound imaging thickness for the calcaneal fat pad in an individual with AT. Abbreviations: CFP, calcaneal fat pad.

2.6. Data Analysis

Statistical analysis was performed using the SPSS software for Windows (version 22, IBM Corp., Armonk, NY, USA). Firstly, the Kolmogorov–Smirnov test was utilized in order to assess normality distribution. Secondly, the descriptive analysis for the total sample was performed. Finally, a comparative analysis for both groups was developed. The mean, standard deviation (SD) and Student´s *t*-test for independent samples were employed for the parametric data. For the non-parametric data, median, interquartile range (IR) and Mann–Whitney *U* test were realized. Moreover, Levene´s test was performed to assess the equality of variances. The Fisher exact test was used to compare differences between sex and group. In addition, bar graphs completed with the 95 confidence interval (CI) error bars were added in order to illustrate the differences between both groups. For all statistical tests, an α error of 0.05 (95% CI) and a desired power of 80% (β error of 0.2) were employed.

In addition, a multivariate analysis was carried out using a linear regression (stepwise selection method; P_{in} = 0.05; P_{out} = 0.10) in order to predict the influence of the descriptive data and group (presence of Achilles tendinopathy) on the statistically significant outcome measurements (showed in the prior described analyses). The dependent variables were CFP thickness and PF thickness at calcaneus insertion. The independent variables were group, sex, weight, height, BMI and age.

3. Results

The sociodemographic data did not show statistically significant differences ($P > 0.05$) for the sex, age, weight and height between groups but it did show statistically significant differences ($P < 0.05$) for the body mass index (BMI) between groups (Table 1). Regarding Table 2 and Figure 3, ultrasound evaluations of the CFP and PF at the calcaneus thickness showed statistically significant differences ($P < 0.01$) with a decrease observed for the tendinopathy group with respect to the control group. For the PF midfoot and forefoot thickness, no significant differences ($P > 0.05$) were observed between groups.

Table 1. Sociodemographic features, pain scores and VISA-A scale of the sample.

Data	Achilles Tendinopathy (*n* = 71)	Controls (*n* = 70)	*P*-Value Cases vs. Controls
Weight, kg	76.00 ± 12.00 [†]	75.00 ± 18.50 [†]	0.412 [‡]
Age, year	45.11 ± 12.75 [*]	37.61 ± 11.91 [*]	0.200 [**]
Height, m	1.76 ± 0.11 [†]	1.76 ± 0.12 [†]	0.566 [‡]
BMI, kg/m^2	24.81 ± 2.13 [†]	23.88 ± 3.67 [†]	0.012 [‡]
VAS	2.00 ± 3.00 [†]	N/A	N/A
VISA-A	56.00 ± 14.00 [†]	N/A	N/A
Sex, men/women	62/9	54/18	0.086 [‡‡]

Abbreviations: VAS, visual analogue scale. [*] Mean ± standard deviation (SD) was applied. [**] Student´s *t*-test for independent samples was performed. [†] Median ± interquartile range (IR) was used. [‡] Mann–Whitney *U* test was utilized. [‡‡] Fisher exact test was used.

Table 2. Ultrasonography measurements.

Measurement	Tendinopathy (*n* = 71)	Controls (*n* = 72)	*P*-Value
Distance			
Calcaneal fat pad	7.21 ± 1.59 (4.77–11.21) [†]	8.87 ± 1.59 (6.68–12.74) [†]	0.000 [‡]
PF at the calcaneus insertion thickness	6.99 ± 1.84 (4.30–10.93) [†]	8.94 ± 1.65 (6.11–12.62) [†]	0.000 [‡]
PF at midfoot thickness	7.55 ± 1.66 (4.72–12.45) [*]	7.41 ± 1.24 (4.85–10.43) [*]	0.579 [**]
PF at forefoot thickness	6.19 ± 1.55 (4.20–10.84) [†]	6.32 ± 1.64 (3.83–8.55) [†]	0.607 [‡]

Abbreviations: PF, plantar fascia. [*] Mean ± standard deviation (SD) (minimum–maximum) was applied. [**] Student´s *t*-test for independent samples was performed. [†] Median ± interquartile range (IR) (minimum–maximum) was used. [‡] Mann–Whitney *U* test was utilized.
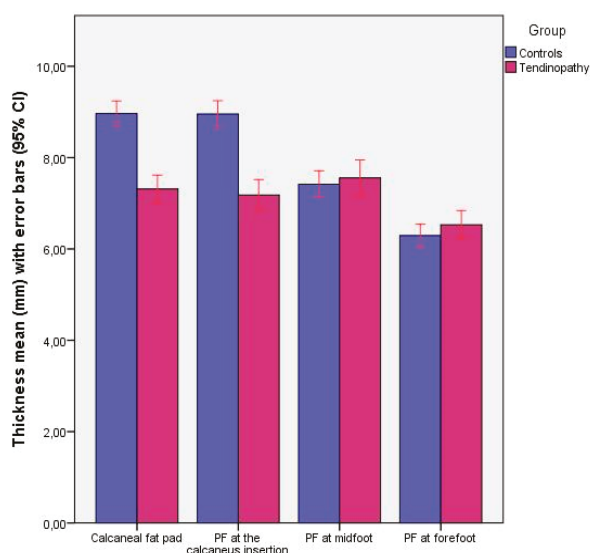

Figure 3. Bar graphs completed with the 95% CI to illustrate the thickness differences of calcaneus fat pad and plantar fascia at the calcaneus insertion, midfoot and forefoot between patients with chronic non-insertional Achilles tendinopathy and controls.

According to the linear regression analysis (Table 3), the prediction model for CFP thickness (R^2 = 0.382) was determined by group (presence of Achilles tendinopathy) and sex, and also the prediction model for PF thickness at calcaneus insertion (R^2 = 0.323) was determined by group and weight. The rest of the independent variables did not predict these statistically significant differences between the case and control groups.

Table 3. Multivariate predictive analysis for CFP and PF thicknesses for patients with Achilles tendinopathy and controls.

Parameter	Model	R^2 Change	Model R^2
CFP thickness (mm)	8.226 −1.774 * Group +0.985 * Sex	0.315 [‡] 0.067 [‡]	0.382
PF thickness at calcaneus insertion (mm)	7.371 −1.818 * Group + 0.021 * Weight (kg)	0.309 [‡] 0.023 [†]	0.323

Abbreviations: CFP, calcaneus fat pad; PF, plantar fascia. [*] Multiplay: Group (control = 0; Tendinopathy = 1); Sex (women = 0; men = 1). [†] *P*-value < 0.05 for a 95% confidence interval was shown. [‡] *P*-value < 0.001 for a 95% confidence interval was shown.

4. Discussion

USI was considered a valid and reliable imaging method to assess soft tissue structures, architectures and sizes. Considering the strong relationship between the Achilles tendon and PF by their locations and insertions, this study may provide a new approach to the assessment and treatment of individuals with AT. This is the first study where the thickness of the PF and CPF were evaluated with USI in patients with AT, variables of interest due to ankle and foot biomechanics [13].

Foot overpronation was considered a risk factor to the predisposition of AT [31] through disturbances in ankle biomechanics producing an extra mechanical stress on the soft tissue structures, such as the PF. Those findings were related to a thickness decrease in the PF at the calcaneus, a structure that serves as a link between the PF and the Achilles tendon. In addition, Cornwall et al. [32] argued that an excessive foot pronation may increase the foot mobility and the level of stress applied through PF. Many studies have also reported an altered foot arch in patients with disturbances in PF modifies the capability of absorbing ground reaction forces [32,33]. These structural alterations were related to the Achilles tendon capacity to store and release energy during the gait [34]. Our findings showed an altered thickness in the PF at the calcaneus insertion, which could be explained by the changes in the ankle biomechanics, the extrinsic and intrinsic foot muscles observed in patients with AT.

CFP protects the rear-foot from the stress produced during the initial phase of locomotion and the heel-strike actions. CFP thickness was described as a valid and reliable method to quantify and assess the heel fat pad by USI [24]. Several authors argued that CFP has been implicated in foot and plantar disturbances, such as diabetes [35], fractures [36] and plantar heel pain [37]. To our knowledge, this is the first study to evaluate the CPF thickness in patients with AT. The results of our study showed a decrease in CPF thickness in the tendinopathy group, so it could be considered a relevant variable for the diagnosis and management of individuals with AT.

The findings of the present study did not intend to provide an explanation about the etiology of AT. In addition, several studies argued that the etiology of the tendinopathy was caused by multiple factors. The authors try to offer a novel approach to evaluate and quantify soft tissue structures that usually present disturbances in AT by USI. Thus, USI assessment of the reduction in PF in subjects with AT could be useful to carry out a follow-up for the interventions prescribed to treat AT.

5. Limitations

This study presented several limitations. Firstly, ultrasound M-mode was not used, which could have been convenient for the evaluation of muscle and tissue features. Secondly, a pressure platform was not employed which may have been useful to contemplate load variables, such as plantar pressures.

6. Conclusions

The thickness of the PF at the insertion and the CPF is reduced in patients with AT measured by USI. The present study did not intend to provide an explanation of the cause or management of AT. Consequently, USI abnormalities in the PF complex should be interpreted within the clinical context in patients with AT.

References

1. Jarvinen, T.A.; Kannus, P.; Paavola, M.; Jarvinen, T.L.; Jozsa, L.; Jarvinen, M. Achilles tendon injuries. *Curr. Opin. Rheumatol.* **2001**, *13*, 150–155. [CrossRef]

2. Albers, I.S.; Zwerver, J.; Diercks, R.L.; Dekker, J.H.; Van den Akker-Scheek, I. Incidence and prevalence of lower extremity tendinopathy in a Dutch general practice population: A cross sectional study. *BMC Musculoskelet. Disord.* **2016**, *17*, 16. [CrossRef]

3. Cook, J.L.; Purdam, C.R. Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *Br. J. Sports Med.* **2009**, *43*, 409–416. [CrossRef] [PubMed]

4. Romero-Morales, C.; Martin-Llantino, P.J.; Calvo-Lobo, C.; Palomo-Lopez, P.; Lopez-Lopez, D.; Pareja-Galeano, H.; Rodriguez-Sanz, D. Comparison of the sonographic features of the Achilles Tendon complex in patients with and without achilles tendinopathy: A case-control study. *Phys. Ther. Sport* **2019**, *35*, 122–126. [CrossRef]

5. Nadeau, M.-J.; Desrochers, A.; Lamontagne, M.; Larivière, C.; Gagnon, D.H. Quantitative ultrasound imaging of Achilles tendon integrity in symptomatic and asymptomatic individuals: Reliability and minimal detectable change. *J. Foot Ankle Res.* **2016**, *9*, 30. [CrossRef]

6. Lorimer, A.V.; Hume, P.A. Achilles tendon injury risk factors associated with running. *Sports Med.* **2014**, *44*, 1459–1472. [CrossRef] [PubMed]

7. Maganaris, C.N.; Narici, M.V.; Maffulli, N. Biomechanics of the Achilles tendon. *Disabil. Rehabil.* **2008**, *30*, 1542–1547. [CrossRef] [PubMed]

8. Hintermann, B.; Nigg, B.M.; Sommer, C. Foot movement and tendon excursion: an in vitro study. *Foot Ankle Int.* **1994**, *15*, 386–395. [CrossRef] [PubMed]

9. Galloway, M.T.; Lalley, A.L.; Shearn, J.T. The role of mechanical loading in tendon development, maintenance, injury, and repair. *J. Bone Jt. Surg. Am.* **2013**, *95*, 1620–1628. [CrossRef] [PubMed]

10. Angin, S.; Crofts, G.; Mickle, K.J.; Nester, C.J. Ultrasound evaluation of foot muscles and plantar fascia in pes planus. *Gait Posture* **2014**, *40*, 48–52. [CrossRef] [PubMed]

11. Huang, Y.-C.; Wang, L.-Y.; Wang, H.-C.; Chang, K.-L.; Leong, C.-P. The relationship between the flexible flatfoot and plantar fasciitis: Ultrasonographic evaluation. *Chang Gung Med. J.* **2004**, *27*, 443–448.

12. Boussouar, A.; Meziane, F.; Crofts, G. Plantar fascia segmentation and thickness estimation in ultrasound images. *Comput. Med. Imaging Graph.* **2017**, *56*, 60–73. [CrossRef] [PubMed]

13. Crofts, G.; Angin, S.; Mickle, K.J.; Hill, S.; Nester, C.J. Reliability of ultrasound for measurement of selected foot structures. *Gait Posture* **2014**, *39*, 35–39. [CrossRef]

14. Draghi, F.; Gitto, S.; Bortolotto, C.; Draghi, A.G.; Ori Belometti, G. Imaging of plantar fascia disorders: Findings on plain radiography, ultrasound and magnetic resonance imaging. *Insights Imaging* **2017**, *8*, 69–78. [CrossRef]

15. Lobo, C.C.; Marin, A.G.; Sanz, D.R.; Lopez, D.L.; Lopez, P.P.; Morales, C.R.; Corbalan, I.S. Ultrasound evaluation of intrinsic plantar muscles and fascia in hallux valgus: A case-control study. *Medicine* **2016**, *95*, e5243. [CrossRef]

16. Lobo, C.C.; Morales, C.R.; Sanz, D.R.; Corbalan, I.S.; Marin, A.G.; Lopez, D.L. Ultrasonography Comparison of Peroneus Muscle Cross-sectional Area in Subjects With or Without Lateral Ankle Sprains. *J. Manipulative Physiol. Ther.* **2016**, *39*, 635–644. [CrossRef]

17. Taniguchi, M.; Fukumoto, Y.; Kobayashi, M.; Kawasaki, T.; Maegawa, S.; Ibuki, S.; Ichihashi, N. Quantity and Quality of the Lower Extremity Muscles in Women with Knee Osteoarthritis. *Ultrasound Med. Biol.* **2015**, *41*, 2567–2574. [CrossRef]

18. Strini, P.J.S.A.; Strini, P.J.S.A.; de Souza Barbosa, T.; Gavião, M.B.D. Assessment of thickness and function of masticatory and cervical muscles in adults with and without temporomandibular disorders. *Arch. Oral Biol.* **2013**, *58*, 1100–1108. [CrossRef]

19. Mohseny, B.; Nijhuis, T.H.; Hundepool, C.A.; Janssen, W.G.; Selles, R.W.; Coert, J.H. Ultrasonographic quantification of intrinsic hand muscle cross-sectional area; reliability and validity for predicting muscle strength. *Arch. Phys. Med. Rehabil.* **2015**, *96*, 845–853. [CrossRef] [PubMed]

20. Javanshir, K.; Amiri, M.; Mohseni Bandpei, M.A.; Penas, C.F.D.L.; Rezasoltani, A. The effect of different exercise programs on cervical flexor muscles dimensions in patients with chronic neck pain. *J. Back Musculoskelet. Rehabil.* **2015**, *28*, 833–840. [CrossRef]

21. Teyhen, D.S.; Gill, N.W.; Whittaker, J.L.; Henry, S.M.; Hides, J.A.; Hodges, P. Rehabilitative ultrasound imaging of the abdominal muscles. *J. Orthop. Sports Phys. Ther.* **2007**, *37*, 450–466. [CrossRef]

22. Whittaker, J.L.; Warner, M.B.; Stokes, M. Comparison of the Sonographic Features of the Abdominal Wall Muscles and Connective Tissues in Individuals with and without Lumbopelvic Pain. *J. Orthop. Sports Phys. Ther.* **2013**, *43*, 11–19. [CrossRef]

23. Romero-Morales, C.; Almazán-Polo, J.; Rodríguez-Sanz, D.; Palomo-López, P.; López-López, D.; Vázquez-González, S.; Calvo-Lobo, C. Rehabilitative Ultrasound Imaging Features of the Abdominal Wall Muscles in Elite and Amateur Basketball Players. *Appl. Sci.* **2018**, *8*, 809. [CrossRef]

24. Hall, M.M.; Finnoff, J.T.; Sayeed, Y.A.; Smith, J. Sonographic Evaluation of the Plantar Heel in Asymptomatic Endurance Runners. *J. Ultrasound Med.* **2015**, *34*, 1861–1871. [CrossRef]

25. Welch, V.; Jull, J.; Petkovic, J.; Armstrong, R.; Boyer, Y.; Cuervo, L.G.; Edwards, S.; Lydiatt, A.; Gough, D.; Grimshaw, J.; et al. Protocol for the development of a CONSORT-equity guideline to improve reporting of health equity in randomized trials. *Implement. Sci.* **2015**, *10*, 146. [CrossRef]

26. Alfredson, H.; Pietila, T.; Jonsson, P.; Lorentzon, R. Heavy-load eccentric calf muscle training for the treatment of chronic Achilles tendinosis. *Am. J. Sports Med.* **1998**, *26*, 360–366. [CrossRef]

27. Alfredson, H.; Cook, J. A treatment algorithm for managing Achilles tendinopathy: New treatment options. *Br. J. Sports Med.* **2007**, *41*, 211–216. [CrossRef] [PubMed]

28. Romero-Morales, C.; Martín-Llantino, P.J.; Calvo-Lobo, C.; Beltran-Alacreu, H.; López-López, D.; Sánchez-Gómez, R.; Rodríguez-Sanz, D. Effectiveness of Eccentric Exercise and a Vibration or Cryotherapy Program in Enhancing Rectus Abdominis Muscle Thickness and Inter-Rectus Distance in Patients with Chronic Mid-Portion Achilles Tendinopathy: A Randomized Clinical Trial. *Int. J. Med. Sci.* **2018**, *15*, 1764–1770. [CrossRef] [PubMed]

29. Rathleff, M.S.; Moelgaard, C.; Lykkegaard Olesen, J. Intra- and interobserver reliability of quantitative ultrasound measurement of the plantar fascia. *J. Clin. Ultrasound* **2011**, *39*, 128–134. [CrossRef] [PubMed]

30. Lopez-Lopez, D.; Becerro-de-Bengoa-Vallejo, R.; Losa-Iglesias, M.E.; Soriano-Medrano, A.; Palomo-Lopez, P.; Morales-Ponce, A.; Rodriguez-Sanz, D.; Calvo-Lobo, C. Relationship Between Decreased Subcalcaneal Fat Pad Thickness and Plantar Heel Pain. A Case Control Study. *Pain Physician* **2019**, *22*, 109–116.

31. Maffulli, N.; Kader, D. Tendinopathy of tendo achillis. *J. Bone Jt. Surg. Br.* **2002**, *84*, 1–8. [CrossRef] [PubMed]

32. Cornwall, M.W.; McPoil, T.G. Plantar fasciitis: Etiology and treatment. *J. Orthop. Sports Phys. Ther.* **1999**, *29*, 756–760. [CrossRef]

33. Karr, S.D. Subcalcaneal heel pain. *Orthop. Clin. N. Am.* **1994**, *25*, 161–175.

34. Lai, A.; Schache, A.G.; Lin, Y.-C.; Pandy, M.G. Tendon elastic strain energy in the human ankle plantar-flexors and its role with increased running speed. *J. Exp. Biol.* **2014**, *217*, 3159–3168. [CrossRef]

35. Hsu, C.-C.; Tsai, W.-C.; Hsiao, T.-Y.; Tseng, F.-Y.; Shau, Y.-W.; Wang, C.-L.; Lin, S.-C. Diabetic effects on microchambers and macrochambers tissue properties in human heel pads. *Clin. Biomech.* **2009**, *24*, 682–686. [CrossRef]

36. Silver, D.A.; Kerr, P.S.; Andrews, H.S.; Atkins, R.M. Heel pad thickness following calcaneal fractures: ultrasound findings. *Injury* **1994**, *25*, 39–40. [CrossRef]

37. Prichasuk, S. The heel pad in plantar heel pain. *J. Bone Jt. Surg. Br.* **1994**, *76*, 140–142. [CrossRef]

Electromyographic Evaluation of the Impacts of Different Insoles in the Activity Patterns of the Lower Limb Muscles during Sport Motorcycling: A Cross-Over Trial

Israel Casado-Hernández [1], Ricardo Becerro-de-Bengoa-Vallejo [1], Marta Elena Losa-Iglesias [2], Daniel López-López [3,*], David Rodríguez-Sanz [1], Eva María Martínez-Jiménez [1] and César Calvo-Lobo [4]

1   Facultad de Enfermería, Fisioterapia y Podología, Universidad Complutense de Madrid, 28040 Madrid, Spain; israelcasado@yahoo.es (I.C.-H.); ribebeva@ucm.es (R.B.-d.-B.-V.); davidrodriguezsanz@gmail.com (D.R.-S.); eva.hache2@hotmail.com (E.M.M.-J.)
2   Faculty of Health Sciences, Universidad Rey Juan Carlos, 28922 Alcorcón, Spain; marta.losa@urjc.es
3   Research, Health and Podiatry Unit, Department of Health Sciences, Faculty of Nursing and Podiatry, Universidade da Coruña, 15403 Ferrol, Spain
4   Nursing and Physical Therapy Department, Institute of Biomedicine (IBIOMED), Faculty of Health Sciences, Universidad de León, 24401 Ponferrada, Spain; ccall@unileon.es
*   Correspondence: daniel.lopez.lopez@udc.es; Tel.: +34-981-337-400 (ext. 3546)

Abstract: Customized foot insoles (CFI) have been recognized to reduce the prevalence of foot disorders in sport. The aim of this study was to evaluate the effect of four types of CFI on the activity patterns of the lower limb muscles (LLM) in healthy people during sport motorcycling. Methods: This was a cross-over trial (NCT03734133. Participants were recruited from an outpatient foot specialist clinic. Their mean age was 33 ± 5.14 years. While participants were sport motorcycling in a simulator, the electromyography (EMG) function was registered for LLM via surface electrodes. Participants completed separate tests while wearing one of four types of CFI: (1) only polypropylene (58° Shore D), (2) selective aluminum (60 HB Brinell hardness) in metatarsal and first hallux areas and polypropylene elsewhere (58° Shore D), (3) ethylene vinyl acetate (EVA) (52° Shore A), and (4) standard EVA (25° Shore A) as the control. Results: The activity patterns of the LLM while sport motorcycling showed significantly lower peak amplitude for the selective aluminum CFI than the other types of CFI. Conclusion: EMG amplitude peaks for several LLM were smaller for the hardest CFI (selective aluminum 60 HB Brinell hardness) than the other CFIs (polypropylene 58° Shore D, EVA 52° Shore A, and standard EVA 25° Shore A), except for the fibularis longus in right curves that is increased when the knee touches the road increasing the stability.

Keywords: foot insoles; electromyography; joint instability; muscle contractions; motorcycling; plantar pressure

---

1. Introduction

Motorcycle racers ride on speed tracks with multiples curves, moving their bodies by taking different postures as they attempt to minimize their lap times. Motorcycle riders take aerodynamic positions, flexing their hips and knees, and also changing to a braking posture. In different phases of cornering, a rider shifts his mass around the center of mass of the motorcycle from one footpeg to another footpeg. Leaning the motorcycle into corners improves the rider's performance. To achieve

this, centrifugal forces have to be balanced and the rider must make rapid changes of lateral, medial, and anteroposterior movements [1].

Lower limb muscles play a relevant role in motorcycling performance through the forces produced by the feet and footpeg. Studies show the importance of motorcycle riders being in good physical shape to reduce muscle fatigue from the large loads to which they are exposed and thereby improve their performance [2]. Thus, motorcycle riders show an increase in the reactive forces of the joints during riding and the associated activity patterns of the LLM related to footwear and the footpeg. These forces and activity patterns may produce various musculoskeletal conditions, such as (1) Achilles tendinopathy, (2) alterations in plantar pressures, (3) muscle strains, (4) knee pain, (5) ligament sprains, (6) metatarsalgia, (7) patellofemoral syndrome, (8) plantar fasciitis, (9) tendon injuries, and (10) paresthesia [3–6].

These injuries suggest a variety of potential causes: (1) abnormality of LLM, (2) biomechanical dysfunction, (3) characteristics of the riding environment, (4) male gender, (5) type of boot or shoes, and (6) foot pedal force. These factors are complex and related, and their precise roles in causing injury are not clear [7–12].

No prior investigation has evaluated electromyography (EMG) activity patterns of LLM or the effects of customized foot insoles (CFI) on such patterns during sport motorcycling. Previous reports have recorded significant variations in the EMG activity patterns of LLM during cycling [13–15].

In this study, we evaluated the effects of four types of CFI in the activity patterns of the LLM in healthy people during sport motorcycling. We hypothesized that participants wearing different types of CFI might vary in their EMG function patterns of the LLM. Our main goal was to determine which material of CFI limits the EMG patterns the most.

2. Materials and Methods

2.1. Design and Sample

Participants were nine healthy motorcycle riders recruited from an outpatient foot specialist clinic in Madrid (Spain) from November to December 2018 using a non-random consecutive sampling method. The inclusion criteria included the following: (1) at least eighteen years old, (2) no medical problems in the clinical record nor a family history with relevant illnesses, and (3) written informed consent to participate in our research. The exclusion criteria were as follows: (1) medical trauma or a history of leg problems, (2) musculoskeletal disturbances, (3) vascular disease, (4) refusal to give written informed consent, and (5) inability to understand the protocol to participate in our research.

We conducted a cross-over trial study (prospectively registered in ClinicalTrials.gov as NCT03734133). Our report follows the guidelines and checklist in the Template for Intervention Description and Replication (TIDieR) [16].

2.2. Procedure

The baseline assessment comprised an interview that focused on (1) overall medical health, (2) anthropometric and socio-demographic characteristics (age, sex, height, weight, BMI, and Spanish foot size), (3) chronic disorders and illnesses (e.g., arthritis, diabetes mellitus, and musculoskeletal disturbances), and (4) recreational activities. A senior podiatrist also assessed each rider with a complete physical clinical evaluation and measured anthropometric variables such as height, weight, and body mass index.

Before testing commenced, the participants were given a two-minute acclimation period to drive on the motorcycle simulator. During this period, participants drove on the motorcycle simulator to adjust to the surroundings and ensure the shoes were comfortable. Each motorider wore their own shoes with a flat thin rubber sole.

In the simulator CKU28® (Mecanitzats Muntada SL, Manlleu, Spain), the riders drove the Circuit of the Americas race (Austin, TX, USA) in which professional sport motorcyclists compete [17]. During

the simulation, riders drove the race course, making the same movements as real circuit motorcycle riders would need to make. Riders had to accelerate to maximum speed and then, when taking a curve, decelerate. They needed to raise and lower their bodies on the motorcycle simulator to the degrees needed to complete the course. The simulator had a maximum inclination of 60° and riders rested their knees on the ground just as they do on real motorcycles. Riders completed the race at approximately 250 km/h in straightaways and at decreased speeds in curves over a period of 20 min of Racing for each CFI test.

We evaluated four types of CFI: (1) only polypropylene (58° Shore D), (2) selective aluminum (60 HB Brinell hardness) in the metatarsal and first hallux areas and polypropylene elsewhere (58° Shore D), (3) ethylene vinyl acetate (EVA) (52° Shore A), and (4) standard EVA (25° Shore A) as control. We tested each rider with each type of CFI and the order of CFIs was randomly assigned using a random number table. The order of CFIs was the same across all riders.

Before testing, each motorcyclist was situated in a calm place to fix surface electrodes (positive, negative, and earth interfaces) with a built-in preamplifier on the skin over six muscle areas on the right leg: (1) fibularis brevis, (2) fibularis longus, (3) tibialis anterior, (4) gastrocnemius lateralis, (5) gastrocnemius medialis, and (6) soleus. The belly of each muscle was marked on the skin with a permanent marker, swabbed with a 70% alcohol solution and then shaved. The skin was then abraded with skin sandpaper to reduce the electrical impedance to less than 5 kX. All active electrodes were placed 2 cm apart, parallel with the alignment of the muscle fibers, and secured with hypoallergenic tape. A stretch "underwrap" bandage was placed over the electrodes to prevent electrode movement [18].

The precise locations of the EMG electrodes were as follows:

- Fibularis brevis: The electrode was placed anterior to the tendon of the muscle peroneus longus at 25% of the line beginning from the tip of the lateral malleolus to the fibula head, and oriented along this line.
- Fibularis longus: The electrode was placed at 25% of the line beginning from the tip of the head of the fibula to the tip of the lateral malleolus, and oriented along this line.
- Tibialis anterior: The electrode was placed at 1/3 of the line beginning from the tip of the fibula to the tip of the medial malleolus, and oriented along this line. We placed the tibialis anterior electrodes carefully in line and parallel to the fibula because cross-talk can exist between the tibialis anterior and peroneus longus, particularly when the peroneus longus electrodes are placed too anteriorly on the leg. Manual muscle testing was performed for each muscle area to confirm correct placement.
- Gastrocnemius lateralis: The electrode was placed at 1/3 of the line beginning from the head of the fibula to the heel, and oriented along this line.
- Gastrocnemius medialis: The electrode was placed on the most prominent bulge of the muscle and oriented along the length of the leg.
- Soleus: The electrode was placed at 2/3 of the line beginning from the medial condyle of the femur to the medial malleolus, and oriented along this line.

For normalization purposes, before performing each trial, the participant was instructed on how to perform a maximum voluntary isometric contraction (MVIC) for each muscle based on a technique described by Yang and Winter [19]. Three MVICs were recorded for each muscle within a recording time of eight seconds. This allowed time for the participant to build up the contraction and for the tester to stabilize the foot. Thus, the MVIC was measured over a four-second window only. The participants were instructed to exert a maximum effort against the resistance of the tester and were given verbal encouragement while doing so. The rationale for this test is to normalize the maximal amplitude of a submaximal task, such as walking or riding a motorcycle, to the amplitude of the MVIC produced by the participant. One of the known limitations related to the reliability of the MVIC is the participant's ability to produce an MVIC sincerely [19]. Although the participant's effort can vary between each contraction, using the MVIC is better than not using a normalization technique at all [20]. To reduce

the effect of muscle fatigue between succeeding contractions, participants were given one-minute rest periods [21].

The EMG electrodes were inserted into a lightweight backpack EMG interface unit (model MWX8). The interface unit was wired to the filtering hardware by an overhead pulley system. This was connected to a personal computer where EMG information was recorded using DataLINK (v.5.0) (Biometrics Ltd., Ladysmith, VA, USA).

We measured the signal activity patterns with surface electromyography (EMG) sensors (SX230) (Biometrics Ltd., Ladysmith, VA, USA), using the protocol design of SENIAM to comply with a non-invasive evaluation of muscles [12]. Electrode specifications in this study were as follows: amplification, single differential; inter-electrode distance, 2 cm; contact sensors, two $0.37 \times 0.20 \times 0.06$ cm silver bars; preamplifier gain, 1000; input impedance, >10,000,000 MΩ; and common mode rejection ratio, >96 dB. The main amplifier unit feature was a gain of 1000-fold and frequency response of 20–460 Hz. Surface electromyography data were sampled at 1000 Hz and recorded by a computerized data logger (MWX8 Biometrics Ltd., Ladysmith, VA, USA).

We chose the 1 kHz data sampling rate because raw EMG signals were collected, converted with DataLINK, and recorded prior to producing the filtered/rectified data. This ensured that the peaks and onsets of activity were reproduced with subsequent data, and enabled the verification that movement artifacts were minimal [18]. Electromyography signals recorded during the selected times were filtered with a band pass filter (bandwidth of 20–460 Hz) [21,22]. EMG parameters such as frequency content and waveform shape may require sampling rates more than twice the Nyquist rate [23] for accurate interpretation. However, for the analysis of amplitude and timing parameters, over-sampling is unnecessary [24].

The average rectified value (ARV) was calculated. The total ARV and ARV during riding for each muscle and each participant were calculated and normalized by the peak ARV for each participant. To identify the muscular activation phase, the onset and offset of surface electromyography were assessed. Phases of muscular activation were defined by the period when the signal was above a threshold of 25% of the peak ARV for each muscle and participant [25,26]. The signal activity patterns related to the LLM were registered for each CFI test during the last five minutes of the simulated race.

For each riding, "raw" recordings (without bandpass or RMS/envelope filtering) were taken to compare with filtered recordings. The signals were then filtered during testing for each participant.

We calculated the root mean square (RMS) of each signal for each subject, and used a filter bank of 10 wavelets. We also computed center frequencies that characterize the wavelets [27]. These operations yielded a time series of EMG intensities for each wavelet (wavelet domain). These time series were re-sampled at 200 Hz and divided by the square root of the sum of the intensities of the data points in all wavelet domains. This yielded intensity patterns, which we used in blocks for further analysis.

During the collection and analysis of EMG tracings, the operator was blinded to the condition being evaluated. Muscle onset was determined by visual observation as the time when a rapid and continuous increase in voltage occurred that was unrelated to movement artifacts or noise interference.

Lastly, all EMG electrodes were placed on the muscles of the right lower extremity of each motorider, so the right curve means that the right lower extremity is touching the floor with the knee and applying force against the footpeg, while the left curve means that the same lower limb must move towards the motorcycle body part to maintain stability.

2.3. Ethical Considerations

The study protocol was reviewed and approved by the commission of the Clinical Research Ethics Committee of Hospital Clínico San Carlos, Madrid, Spain (number C.I. 18/506-E). All riders gave written informed consent to participate in our research. Furthermore, our study procedures complied with the ethical principles and human research guidelines of the World Medical Association in the Declaration of Helsinki.

Sample Size Calculation

We calculated the sample size necessary for our research goals with a program developed by the Clinical Epidemiology and Biostatistics Unit of the Universidade da Coruña (A Coruña, Spain) (http://www.fisterra.com/mbe/investiga/9muestras/9muestras2.asp) [28]. We based our calculations on prior results obtained by McCulloch on the effect of pedal type on EMG activity patterns in muscle recruitment (16.9 ± 7.2% for a lateral translation pedal and 30.9 ± 8.5% for a standard pedal) [29]. With these results, a sample of 8 riders would give 80% analysis power (beta = 20%), to detect a significant effect (alpha of 0.05) in a two-tailed test for this cross-over experimental study.

2.4. Statistical Analysis

We assessed all variables for normality with the Shapiro–Wilk test and considered a variable to have a normal distribution if the probability value for a test was greater than or equal to 0.05. To describe riders' demographic characteristics, we computed the means, standard deviations, and 95% confidence intervals of rider age, BMI, Spanish foot size, height, and weight.

We measured intra-trial reliability with the three records for each CFI for each rider. To evaluate reliability within trials in each rider, we computed intra-class correlation coefficients. For interpreting ICC values, we considered values less than 0.40 as poor, values between 0.40 and 0.59 as fair, values between 0.60 and 0.74 as good, and values 0.75 or greater as excellent [30,31]. Portney and Watkins [32] proposed that reliability coefficients greater than 0.90 were sufficient for clinical measurement. We also calculated the mean scores and the standard error of measurement (SEM) [30]. We used Brand and Altman's [33] formula for the SEM as follow: SEM = SD × sqrt (1 − ICC).

The minimal detectable change (MDC) can be used to assess the minimal magnitude of change required to be 95% confident that the observed change between the 2 tests reflects the true change and not measurement error [34]. The MDC was calculated as: $1.96 \times SEM \times \sqrt{2}$.

For demonstrating the effect size of the comparisons, the Cohen d coefficient was calculated. Cohen's d effect size can be interpreted as follows: values ≤0.20 indicate slight effects, values between 0.20 and 0.49 indicate fair effects, values between 0.50 and 0.79 indicate moderate effects, and values larger than 0.79 indicate large effects [35].

We computed nonparametric Mann–Whitney U tests to assess differences among CFIs groups. For all tests, we considered *p* values < 0.05 to be statistically significant. We used SPSS 19.0 for Windows (IBM, Chicago, IL, USA) to perform all analyses.

3. Results

3.1. Characteristics of the Riders

Table 1 shows the anthropometric and demographic characteristics of the nine healthy riders who participated in the study.

Table 1. Anthropometric and demographic characteristics of the sample of riders.

Variables	Mean ± SD (n = 9)	Range (Min–Max) (n = 9)	95% CI (n = 9)
Age (years)	33 ± 5.14	26–40	29.04–36.96
Height (cm)	175 ± 44	164–190	169.77–181.12
Weight (Kg)	71 ± 44	63–78	67.50–75.38
BMI (Kg/m^2)	23.45 ± 1.04	22.22–24.82	22.65–24.26
Spanish foot size	42.22 ± 2.86	37–47	40.02–44.42

Abbreviations: SD: standard deviation; Min: minimum; Max: maximum; 95% CI: 95% confidence interval; cm: centimeters; Kg: kilograms; Kg/m^2: kilograms/meter^2

3.2. Normality of EMG Variables of the LLM by CFI

Table 2 shows the Shapiro–Wilk test results on the normality of the EMG variables of the LLM by CFI types for left and right curves in the simulator.

Table 2. Normality of the electromyography variables of the lower limb muscles for particular types of customized foot insoles for left and right curves in the simulator.

Variable LLM and Curve	*p*-Value	Variable LLM and Curve	*p*-Value
Tibialis Anterior standard EVA left curve	0.232	Soleus standard EVA left curve	0.628
Tibialis Anterior standard EVA right curve	0.934	Soleus standard EVA right curve	0.791
Tibialis Anterior EVA left curve	0.080	Soleus EVA left curve	0.840
Tibialis Anterior EVA right curve	0.086	Soleus EVA right curve	0.507
Tibialis Anterior polypropylene left curve	0.656	Soleus polypropylene left curve	0.863
Tibialis Anterior polypropylene right curve	0.016	Soleus polypropylene right curve	0.375
Tibialis Anterior selective aluminum left curve	0.115	Soleus selective aluminum left curve	0.787
Tibialis Anterior selective aluminum right curve	0.006	Soleus selective aluminum right curve	0.042
Gastrocnemius lateralis standard EVA left curve	0.319	Fibularis brevis standard EVA left curve	0.972
Gastrocnemius lateral standard EVA right curve	0.716	Fibularis brevis standard EVA right curve	0.264
Gastrocnemius lateralis EVA left curve	0.636	Fibularis brevis EVA left curve	0.084
Gastrocnemius lateralis EVA right curve	0.095	Fibularis brevis EVA right curve	0.495
Gastrocnemius lateralis polypropylene left curve	0.579	Fibularis brevis polypropylene left curve	0.059
Gastrocnemius lateralis polypropylene right curve	0.089	Fibularis brevis polypropylene right curve	0.753
Gastrocnemius lateralis selective aluminum left curve	0.338	Fibularis brevis selective aluminum left curve	0.081
Gastrocnemius lateralis selective aluminum right curve	0.769	Fibularis brevis selective aluminum right curve	0.852
Gastrocnemius medialis standard EVA left curve	0.810	Fibularis longus standard EVA left curve	0.468
Gastrocnemius medialis standard EVA right curve	0.019	Fibularis longus standard EVA right curve	0.460
Gastrocnemius medialis EVA left curve	0.250	Fibularis longus EVA left curve	0.017
Gastrocnemius medialis EVA right curve	0.233	Fibularis longus EVA right curve	0.799
Gastrocnemius medialis polypropylene left curve	0.644	Fibularis longus polypropylene left curve	0.144
Gastrocnemius medialis polypropylene right curve	0.812	Fibularis longus polypropylene right curve	0.728
Gastrocnemius medialis selective aluminum left curve	0.084	Fibularis longus selective aluminum left curve	0.276
Gastrocnemius medialis selective aluminum right curve	0.810	Fibularis longus selective aluminum right curve	0.291

Abbreviations: EVA, ethylene vinyl acetate; Right curve, Knee touching the road; Left curve, Lower extremity towards the motorcycle body part. *p* Values are from Shapiro-Wilk tests

Intraclass Correlation Coefficient for EMG in LLM

Table 3 shows the ICC values for EMG activity in the LLM by CFI types for left and right turns in the simulator.

Table 4 shows the EMG activity patterns of the LLM by type of CFI for left and right curves in the simulator. Table 5 shows comparisons between pairs of CFIs on these activity patterns.

The selective aluminum (60 HB Brinell hardness) CFI generated statistically significant smaller amplitude peaks than Eva 52% and Polypropylene 58% CFIs compared with Standard EVA 25% CFI for left and right curves in the simulator ($p < 0.001$) in all muscles studied except for the fibularis longus right curve that is increased ($p < 0.001$).

In another hand, polypropylene 58° CFI produced statistically significant smaller amplitude peaks than Eva 52% CFI compared with Standard EVA 25% for left and right curves in the simulator for all muscles ($p < 0.001$) except also, for the fibularis longus right curve which increased ($p < 0.001$).

The results using Eva 52% are lower compared with Standard EVA 25% for left and right curves in the simulator for all muscles ($p < 0.001$) except for the fibularis longus right curve which increased ($p < 0.001$).

Table 3. Intraclass correlation coefficients, standard error of measurement and minimal detectable change for electromyography (EMG) activity patterns of the lower limb muscles by types of customized foot insoles for left and right curves in the simulator.

Variable LLM and Curve	ICC (95% CI)	SEM	MDC	*p*-Value
Tibial Anterior standard EVA left curve	0.980 (0.940–0.995)	0.499	1.384	<0.001
Tibialis Anterior standard EVA right curve	0.835 (0.481–0.960)	0.995	2.759	0.002
Tibialis Anterior EVA left curve	0.942 (0.824–0.986)	0.352	0.975	<0.001
Tibial Anterior EVA curve right	0.785 (0.295–0.948)	1.433	3.971	0.006
Tibialis Anterior polypropylene left curve	0.935 (0.803–0.984)	1.050	2.912	<0.001
Tibialis Anterior polypropylene right curve	0.981 (0.940–0.995)	0.416	1.154	<0.001
Tibialis Anterior selective aluminum left curve	0.969 (0.904–0.992)	0.868	2.406	<0.001
Tibialis Anterior selective aluminum right curve	0.933 (0.792–0.983)	0.774	2.145	<0.001
Gastrocnemius lateralis standard EVA left curve	0.932 (0.797–0.983)	0.386	1.070	<0.001
Gastrocnemius lateral standard EVA right curve	0.995 (0.866–0.989)	0.330	0.915	<0.001
Gastrocnemius lateralis EVA left curve	0.974 (0.916–0.994)	0.221	0.612	<0.001
Gastrocnemius lateralis EVA right curve	0.956 (0.868–0.989)	0.501	1.390	<0.001
Gastrocnemius lateralis polypropylene left curve	0.950 (0.707–0.989)	0.101	0.279	<0.001
Gastrocnemius lateralis polypropylene right curve	0.974 (0.917–0.994)	0.585	1.622	<0.001
Gastrocnemius lateralis selective aluminum left curve	0.618 (−0.213–0.907)	0.655	1.816	0.053
Gastrocnemius lateralis selective aluminum right curve	0.986 (0.959–0.997)	0.380	1.053	<0.001
Gastrocnemius medialis standard EVA left curve	0.972 (0.913–0.993)	0.321	0.891	<0.001
Gastrocnemius medialis standard EVA right curve	0.980 (0.939–0.995)	0.856	2.372	<0.001
Gastrocnemius medialis EVA left curve	0.506 (−0.412–0.875)	2.488	6.897	0.101
Gastrocnemius medialis EVA right curve	0.995 (0.984–0.999)	0.324	0.898	<0.001
Gastrocnemius medialis polypropylene left curve	0.868 (0.604–0.967)	0.516	1.430	<0.001
Gastrocnemius medialis polypropylene right curve	0.985 (0.955–0.996)	0.311	0.862	<0.001
Gastrocnemius medialis selective aluminum left curve	0.994 (0.981–0.998)	0.225	0.625	<0.001
Gastrocnemius medialis selective aluminum right curve	0.980 (0.935–0.995)	0.296	0.819	<0.001
Soleus standard EVA left curve	0.981 (0.942–0.995)	1.083	3.003	<0.001
Soleus standard EVA right curve	0.943 (0.829–0.986)	2.151	5.963	<0.001
Soleus EVA left curve	0.823 (0.474–0.956)	3.315	9.189	0.001
Soleus EVA right curve	0.989 (0.968–0.997)	0.915	2.535	0.001
Soleus polypropylene left curve	0.939 (0.815–0.985)	0.620	1.718	<0.001
Soleus polypropylene right curve	0.849 (0.531–0.963)	2.304	6.387	<0.001
Soleus selective aluminum left curve	0.972 (0.917–0.993)	0.540	1.498	<0.001
Soleus selective aluminum right curve	0.950 (0.836–0.988)	1.346	3.731	0.001
Variable LLM and curve	**ICC (95% CI)**	**SEM**	**MDC**	***p*-Value**
Fibularis brevis standard EVA left curve	0.801 (0.421–0.950)	2.360	6.541	<0.001
Fibularis brevis standard EVA right curve	0.862 (0.587–0.966)	1.612	4.469	<0.001
Fibularis brevis EVA left curve	0.951 (0.850–0.988)	1.260	3.491	<0.001
Fibularis brevis EVA right curve	0.988 (0.962–0.997)	0.620	1.719	<0.001
Fibularis brevis polypropylene left curve	0.950 (0.851–0.988)	0.762	2.114	<0.001
Fibularis brevis polypropylene right curve	0.956 (0.864–0.989)	0.990	2.744	<0.001
Fibularis brevis selective aluminum left curve	0.981 (0.941–0.995)	0.804	2.227	<0.001
Fibularis brevis selective aluminum right curve	0.950 (0.843–0.988)	0.892	2.473	<0.001
Fibularis longus standard EVA left curve	0.202 (−0.892–0.780)	3.716	10.301	0.308
Fibularis longus standard EVA right curve	0.888 (0.654–0.972)	2.349	6.512	<0.001
Fibularis longus EVA left curve	0.992 (0.975–0.998)	1.042	2.888	<0.001
Fibularis longus EVA right curve	0.965 (0.892–0.991)	1.478	4.097	<0.001
Fibularis longus polypropylene left curve	0.996 (0.989–0.999)	0.844	2.339	<0.001
Fibularis longus polypropylene right curve	0.990 (0.971–0.998)	0.934	2.589	<0.001
Fibularis longus selective aluminum left curve	0.974 (0.921–0.994)	2.628	7.285	<0.001
Fibularis longus selective aluminum right curve	0.989 (0.967–0.997)	0.824	2.285	<0.001

Abbreviations: 95% CI, 95% confidence interval; ICC, intraclass correlation coefficient; SEM: standard error of the mean; MDC: minimal detectable change; EVA, ethylene vinyl acetate; Right curve, Knee touching the road; Left curve, Lower extremity towards the motorcycle body part

Table 4. Maximum muscular activity of the lower limb muscles (LLM) as measured by EMG for different customized foot insoles (CFI) for left and right curves in the simulator.

Variables	Standard EVA 25% (25° Shore A)		EVA 52% (52° Shore A)		Polypropylene 58% (58° Shore D)		Selective Aluminum 60 HD (60 HB Brinell Hardness)	
LLM and Curve	Mean ± SD (95% CI)	Median (95% CI)	Mean ± SD (95% CI)	Median (95% CI)	Mean ± SD (95% CI)	Median (95% CI)	Mean ± SD (95% CI)	Median (95% CI)
Tibialis anterior left curve	40.65 ± 2.45 (38.76 to 42.53)	41.36 (38.97 to 41.58)	28.45 ± 3.09 (26.08 to 30.83)	27.57 (26.60 to 29.42)	24.88 ± 3.02 (22.56 to 27.19)	22.91 (22.48 to 28.33)	25.88 ± 2.99 (23.58 to 28.17)	27.14 (25.73 to 27.58)
Tibialis anterior right curve	49.39 ± 3.53 (46.67 to 52.11)	48.1 (47.23 to 52.66)	40.20 ± 1.46 (39.07 to 41.32)	40.39 (40.17 to 41.48)	32.48 ± 4.12 (29.31 to 35.64)	30.73 (30.18 to 34.85)	29.70 ± 4.93 (25.91 to 33.49)	29.53 (24.21 to 34.53)
Gastrocnemius lateralis left curve	95.38 ± 4.67 (91.79 to 98.97)	96.17 (94.17 to 97.32)	58.69 ± 2.39 (56.85 to 60.52)	57.70 (57.27 to 60.19)	43.86 ± 3.63 (41.06 to 46.65)	43.22 (42.46 to 44.69)	27.15 ± 3.21 (24.68 to 29.62)	27.04 (24.62 to 30.08)
Gastrocnemius lateralis right curve	26.28 ± 1.48 (25.14 to 27.41)	26.87 (24.79 to 27.52)	21.82 ± 1.37 (20.77 to 22.88)	21.78 (21.00 to 22.58)	18.06 ± 0.45 (17.71 to 18.41)	17.99 (17.85 to 18.15)	16.75 ± 1.06 (15.93 to 17.56)	16.81 (16.58 to 17.37)
Gastrocnemius medialis left curve	40.43 ± 6.05 (35.78 to 45.09)	36.56 (36.00 to 46.15)	37.39 ± 4.58 (33.87 to 40.91)	35.96 (34.92 to 42.17)	26.21 ± 2.54 (24.26 to 28.16)	26.83 (25.06 to 27.10)	17.32 ± 2.09 (15.71 to 18.93)	17.31 (15.64 to 18.49)
Gastrocnemius medialis right curve	37.73 ± 1.92 (36.25 to 39.20)	37.56 (37.12 to 39.12)	31.87 ± 3.54 (29.15 to 34.59)	32.28 (31.56 to 33.88)	19.40 ± 1.42 (18.32 to 20.49)	19.71 (18.74 to 19.93)	8.49 ± 2.91 (6.25 to 10.73)	6.91 (6.39 to 9.54)
Soleus left curve	77.61 ± 9.01 (70.69 to 84.54)	76.19 (69.60 to 84.13)	63.54 ± 8.72 (56.84 to 70.24)	65.08 (58.24 to 67.94)	55.28 ± 5.93 (50.73 to 59.84)	53.98 (52.29 to 61.41)	51.84 ± 6.02 (47.21 to 56.47)	54.36 (49.14 to 56.47)
Soleus right curve	58.05 ± 7.86 (52.00 to 64.09)	60.58 (51.04 to 63.49)	57.60 ± 7.88 (51.54 to 63.65)	57.06 (53.14 to 62.54)	39.07 ± 2.51 (37.13 to 41.00)	39.77 (37.65 to 40.27)	33.65 ± 3.23 (31.17 to 36.13)	33.33 (31.43 to 34.97)
Fibularis brevis left curve	75.13 ± 5.29 (71.06 to 79.20)	75.33 (71.59 to 78.36)	66.97 ± 5.69 (62.59 to 71.35)	68.31 (60.52 to 71.56)	54.87 ± 3.41 (52.25 to 57.49)	56.20 (53.62 to 57.31)	57.78 ± 5.83 (53.29 to 62.26)	56.17 (54.55 to 57.89)
Fibularis brevis right curve	51.19 ± 4.34 (47.85 to 54.53)	50.00 (47.51 to 52.95)	48.31 ± 5.66 (43.96 to 52.67)	47.76 (44.44 to 51.80)	31.52 ± 4.72 (27.90 to 35.15)	31.34 (27.96 to 35.53)	31.82 ± 3.99 (28.75 to 34.88)	31.72 (30.93 to 33.28)
Fibularis longus left curve	62.47 ± 7.02 (57.07 to 67.87)	60.77 (58.68 to 68.92)	52.26 ± 7.90 (46.19 to 58.34)	51.61 (45.63 to 55.96)	52.00 ± 9.34 (44.82 to 59.18)	50.29 (45.77 to 55.54)	43.88 ± 7.86 (37.83 to 49.92)	45.18 (37.20 to 49.91)
Fibularis longus right curve	40.75 ± 4.16 (37.54 to 43.95)	41.67 (38.35 to 44.05)	54.01 ± 11.65 (45.06 to 62.96)	49.00 (45.34 to 67.09)	59.88 ± 13.34 (49.63 to 70.14)	63.85 (47.19 to 69.52)	75.64 ± 16.30 (63.11 to 88.18)	76.86 (62.52 to 89.20)

Abbreviations: SD: standard deviation; 95% CI: 95% confidence interval; EVA, ethylene vinyl acetate; Right curve, Knee touching the road; Left curve, Lower extremity towards the motorcycle body part

Table 5. Comparisons between types of customized foot insoles in terms of EMG activity of the lower limb muscles for left and right curves in the simulator.

LLM and Curve	EVA 25° Vs EVA 52° *p* Value	EVA 25° Vs Polypropylene 58° *p* Value	EVA 25° Vs Aluminum 60° *p* Value
Tibialis Anterior left curve	<0.001	<0.001	<0.001
Tibialis Anterior right curve	0.001	<0.001	<0.001
Gastrocnemius lateralis left curve	<0.001	<0.001	<0.001
Gastrocnemius lateralis right curve	0.001	<0.001	<0.001
Gastrocnemius medialis left curve	0.194	<0.001	<0.001
Gastrocnemius medialis right curve	0.012	<0.001	<0.001
Soleus left curve	<0.001	<0.001	<0.001
Soleus right curve	0.486	<0.001	<0.001
Fibularis brevis left curve	0.062	0.011	0.001
Fibularis brevis right curve	0.038	0.017	<0.001
Fibularis longus left curve	0.003	0.001	0.001
Fibularis longus right curve	1.000	<0.001	<0.001

Abbreviations: SD: standard deviation; 95% CI: 95% confidence interval; EVA, ethylene vinyl acetate; Right curve, Knee touching the road; Left curve, Lower extremity towards the motorcycle body part

4. Discussion

Researchers have carried out several investigations on EMG activity patterns of LLM during cycling [36–38]. Notably, Bousie et al. [36] found that contoured insoles increased the contact area of the foot in stationary cycling. In addition, Casado et al. reported variations in motorcycle riders' plantar pressures with different CFIs when riding in a motorcycle simulator, concluding that metatarsal and hallux pressures decreased with the hardest CFI [17].

Riding on a speed circuit with multiples curves and bends requires precise control of the dynamics of the supersport motorcycle. Stability on the motorcycle is achieved with the hands on the handlebars, buttocks on the seat, legs, and feet on the footpegs. Motorcycle riders perform load transfers during changes in speed, lateral tilts on curves, and changes in direction, demanding repeated physical effort [39,40].

To our knowledge, our investigation is the first to use EMG for recording the muscular activity of the LLM in motorcycling with different CFI. We evaluated the effects of four types of CFI: (1) only polypropylene (58° Shore D), (2) selective aluminum (60 HB Brinell hardness) in the metatarsal and first hallux areas and polypropylene elsewhere (58° Shore D), (3) ethylene vinyl acetate (EVA) (52° Shore A), and (4) standard EVA (25° Shore A; used as control).

In our research, we studied the variation of EMG activity peak amplitudes in the right lower leg of each rider taking right and left curves as they used different CFIs. To maintain the stability of the motorcycle when taking a curve, a rider touches the road on the inner part of the curve with his knee and applies force against the footpeg with the plantar aspect of his first metatarsal bone. When taking the curve, the rider also moves his other leg toward the body of the motorcycle. The fibularis longus is the main muscle that keeps the first metatarsal bone in plantarflexion against the footpeg and therefore is one of the most important muscles for maintaining stability as the knee touches the road during the curve.

The select aluminum CFI had significantly smaller amplitude peaks than the other CFIs for both left and right curves in the simulator for the tibialis anterior, gastrocnemius lateralis, gastrocnemius medialis, soleus, and fibularis brevis. For the fibularis longus, the select aluminum CFI had significantly smaller amplitude peaks than other CFIs in left curves, but significantly higher amplitude peaks than other CFIs in right curves. We suspect that the high activity of the fibularis longus in right curves, when the knee touches the road, increases stability while riding. Also, decreased activity of the other LLM in left and right curves would likely contribute to decreased fatigue when riding.

More studies are needed on muscular activity during sport motorcycling because there are currently speed circuits with a greater number of right and left curves than in the circuit our riders rode. The data we obtained would help estimate the overloading of LLM in such contexts.

A limitation of our study is the lack of inertial forces and steering control challenges in our simulator that occur when riding a motorcycle on a real speed circuit. Further research is needed on the physical effects of motorcycling during actual sport races.

5. Conclusions

EMG amplitude peaks for several LLM were smaller for the hardest CFI (selective aluminum 60 HB Brinell hardness) than the other CFIs (polypropylene 58° Shore D, EVA 52° Shore A, and standard EVA 25° Shore A), except for the fibularis longus in right curves that is increased when the knee touches the road increasing the stability.

References

1. Ibbot, A. *MotoGP Performance Riding Techniques*; Haynes: Somerset, UK, 2013.
2. Hill, E.C.; Housh, T.J.; Smith, C.M.; Cochrane, K.C.; Jenkins, N.D.M.; Cramer, J.T.; Schmidt, R.J.; Johnson, G.O. Effect of sex on torque, recovery, EMG, and MMG responses to fatigue. *J. Musculoskelet. Neuronal Interact.* **2016**, *16*, 310–317. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5259572/ (accessed on 10 September 2018). [PubMed]
3. Sanner, W.; O'Halloran, W. The biomechanics, etiology, and treatment of cycling injuries. *J. Am. Podiatr. Med. Assoc.* **2000**, *90*, 354–376. Available online: http://www.ncbi.nlm.nih.gov/pubmed/10933005 (accessed on 10 September 2018). [CrossRef]
4. Ramos-Ortega, J.; Domínguez, G.; Castillo, J.M.; Fernández-Seguín, L.; Munuera, P.V. Angular Position of the Cleat According to Torsional Parameters of the Cyclist's Lower Limb. *Clin. J. Sport Med.* **2014**, *24*, 251–255. Available online: http://www.ncbi.nlm.nih.gov/pubmed/24451688 (accessed on 10 September 2018). [PubMed]
5. Mellion, M.B. Common Cycling Injuries. *Sport Med.* **1991**, *11*, 52–70. Available online: http://www.ncbi.nlm. nih.gov/pubmed/2011683 (accessed on 10 September 2018). [CrossRef] [PubMed]
6. Thompson, M.J.; Rivara, F.P. Bicycle-related injuries. *Am. Fam. Physician* **2001**, *63*, 2007–2014. Available online: http://www.ncbi.nlm.nih.gov/pubmed/11388717 (accessed on 10 September 2018). [PubMed]
7. Horner, C.H.; O'Brien, A.A. Motorcycle racing injuries on track and road circuits in Ireland. *Br. J. Sports Med.* **1986**, *20*, 157–158. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1478319/ (accessed on 10 September 2018). [CrossRef] [PubMed]
8. Hinds, J.D.; Allen, G.; Morris, C.G. Trauma and motorcyclists; born to be wild, bound to be injured? *Injury* **2007**, *38*, 1131–1138. Available online: https://www.ncbi.nlm.nih.gov/pubmed/17880974 (accessed on 10 September). [CrossRef]
9. Aare, M.; von Holst, H. Injuries from motorcycle and moped crashes in Sweden from 1987 to 1999. *Inj. Control Saf. Promot.* **2003**, *10*, 131–138. Available online: https://www.ncbi.nlm.nih.gov/pubmed/12861911 (accessed on 10 September 2018). [CrossRef]
10. Tomida, Y.; Hirata, H.; Fukuda, A.; Tsujii, M.; Kato, K.; Fujisawa, K.; Uchida, A. Injuries in elite motorcycle racing in Japan. *Br. J. Sports Med.* **2005**, *39*, 508–511. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1725294/ (accessed on 10 September 2018). [CrossRef]
11. Bedolla, J.; Santelli, J.; Sabra, J.; Cabanas, J.G.; Ziebell, C.; Olvey, S. Elite motorcycle racing: Crash types and injury patterns in the MotoGP class. *Am. J. Emerg. Med.* **2016**, *34*, 1872–1875. Available online: https://www.ncbi.nlm.nih.gov/pubmed/27431735 (accessed on 10 September 2018).

12. Varley, G.W.; Spencer-Jones, R.; Thomas, P.; Andrews, D.; Green, A.D.; Stevens, D.B. Injury patterns in motorcycle road racers: Experience on the Isle of Man 1989–1991. *Injury* **1993**, *24*, 443–446. Available online: https://www.ncbi.nlm.nih.gov/pubmed/8406760 (accessed on 10 September 2018). [CrossRef]

13. Momeni, K.; Faghri, P.D.; Evans, M. Lower-extremity joint kinematics and muscle activations during semi-reclined cycling at different workloads in healthy individuals. *J. Neuroeng. Rehabil.* **2014**, *11*, 146. Available online: http://www.ncbi.nlm.nih.gov/pubmed/25325920 (accessed on 10 September 2018). [CrossRef]

14. Baum, B.S.; Li, L. Lower extremity muscle activities during cycling are influenced by load and frequency. *J. Electromyogr. Kinesiol.* **2003**, *13*, 181–190. Available online: http://www.ncbi.nlm.nih.gov/pubmed/12586523 (accessed on 10 September 2018). [CrossRef]

15. Hug, F.; Laplaud, D.; Lucia, A.; Grelot, L. EMG Threshold Determination in Eight Lower Limb Muscles During Cycling Exercise: A Pilot Study. *Int. J. Sports Med.* **2006**, *27*, 456–462. Available online: http://www.ncbi.nlm.nih.gov/pubmed/16767610 (accessed on 10 September 2018). [CrossRef]

16. Hoffmann, T.C.; Glasziou, P.P.; Boutron, I.; Milne, R.; Perera, R.; Moher, D.; Altman, D.G.; Barbour, V.; Macdonald, H.; Johnston, M.; et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ* **2014**, *348*, g1687. Available online: http://www.ncbi.nlm.nih.gov/pubmed/24609605 (accessed on 10 September 2018). [CrossRef]

17. Casado-Hernández, I.; Becerro-de-Bengoa-Vallejo, R.; López-López, D.; Gómez-Bernal, A.; Losa-Iglesias, M.E. Aluminum foot insoles reduce plantar forefoot pressure and increase foot comfort for motorcyclists. *Prosthet. Orthot. Int.* **2018**, *42*, 606–611. [CrossRef]

18. Hermens, H.J.; Freriks, B.; Merletti, R.; Stegeman, D.; Blok, J.; Rau, G.; Disselhorst-Klug, C.; Hägg, G. European Recommendations for Surface ElectroMyoGraphy Results of the SENIAM Project. Available online: https://pdfs.semanticscholar.org/1ab2/8b8afcb1216cab1b2f8da0de246c3d5ed6e8.pdf (accessed on 24 January 2018).

19. Yang, J.F.; Winder, D.A. Electromyographic amplitude normalization methods: Improving their sensitivity as diagnostic tools in gait analysis. *Arch. Phys. Med. Rehabil.* **1984**, *65*, 517–521. Available online: https://www.ncbi.nlm.nih.gov/pubmed/6477083 (accessed on 10 September 2018).

20. Knutson, L.M.; Soderberg, G.L.; Ballantyne, B.T.; Clarke, W.R. A study of various normalization procedures for within day electromyographic data. *J. Electromyogr. Kinesiol.* **1994**, *4*, 47–59. Available online: https://www.ncbi.nlm.nih.gov/pubmed/20870546 (accessed on 10 September 2018). [CrossRef]

21. Cram, J.R.; Kasman, G.S.; Holtz, J. *Introduction to Surface Electromyography*; Aspen: Gaithersburg, MD, USA, 1998.

22. Clancy, E.A.; Morin, E.L.; Merletti, R. Sampling, noise-reduction and amplitude estimation issues in Surface electromyography. *J. Electromyogr. Kinesiol.* **2002**, *12*, 1–16. Available online: https://www.ncbi.nlm.nih.gov/pubmed/11804807 (accessed on 10 September 2018). [CrossRef]

23. Hunter, A.M.; St Clair Gibson, A.; Lambert, M.; Dennis, S.; Mullany, H.; O'Malley, M.J.; Vaughan, C.L.; Kay, D.; Noakes, T.D. EMG amplitude in maximal and submaximal exercise is dependent on signal capture rate. *Int. J. Sports Med.* **2003**, *24*, 83–89. Available online: https://www.ncbi.nlm.nih.gov/pubmed/12669251 (accessed on 10 September 2018). [CrossRef]

24. Ives, J.C.; Wigglesworth, J.K. Sampling rate effects on surface EMG timing and amplitude measures. *Clin. Biomech. (Bristol, Avon)* **2003**, *18*, 543–552. Available online: https://www.ncbi.nlm.nih.gov/pubmed/12828904 (accessed on 10 September 2018). [CrossRef]

25. Duc, S.; Bertucci, W.; Pernin, J.N.; Grappe, F. Muscular activity during uphill cycling: Effect of slope, posture, hand grip position and constrained bicycle lateral sways. *J. Electromyogr. Kinesiol.* **2008**, *18*, 116–127. Available online: https://www.ncbi.nlm.nih.gov/pubmed/17123833 (accessed on 10 September 2018). [CrossRef]

26. Li, L.; Caldwell, G.E. Muscle coordination in cycling: Effect of surface incline and posture. *J. Appl. Physiol. (1985)* **1998**, *85*, 927–934. Available online: https://www.ncbi.nlm.nih.gov/pubmed/9729566 (accessed on 10 September 2018). [CrossRef]

27. von Tscharner, V.; Goepfert, B. Gender dependent EMGs of runners resolved by time/frequency and principal pattern analysis. *J. Electromyogr. Kinesiol.* **2003**, *13*, 253–272. Available online: https://www.ncbi.nlm.nih.gov/pubmed/12706605 (accessed on 10 September 2018). [CrossRef]

28. Fernández, P. Investigación: Determinación del tamaño muestral Determinación del tamaño muestral. *Cad Aten Primaria Actual.* **1996**, *303*, 138–140.

29. McCulloch, R.S. Influence of lateral pedal translation on muscle recruitment and kinematics in cyclists. *J. Exerc. Sci. Fit.* **2018**, *16*, 62–67. Available online: https://www.sciencedirect.com/science/article/pii/S1728869X18300121 (accessed on 10 September 2018). [CrossRef] [PubMed]

30. Enderlein, G.; Fleiss, J.L. The Design and Analysis of Clinical Experiments. Wiley, New York - Chichester - Brislane - Toronto - Singapore 1986, 432 S., £38.35. *Biom. J.* **2007**, *30*, 304. [CrossRef]

31. Hallgren, K.A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor. Quant. Methods Psychol.* **2012**, *8*, 23–34. Available online: https://www.ncbi.nlm.nih.gov/pubmed/22833776 (accessed on 10 September 2018). [CrossRef] [PubMed]

32. Portney, L.; Watkins, M. *Foundations of Clinical Research: Applications to Practice*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2009.

33. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (Lond. Engl.)* **1986**, *327*, 307–310. Available online: https://www.ncbi.nlm.nih.gov/pubmed/2868172 (accessed on 10 September 2018). [CrossRef]

34. Stratford, P.W.; Binkley, J.; Solomon, P.; Finch, E.; Gill, C.; Moreland, J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys. Ther.* **1996**, *76*, 359–365. Available online: https://www.ncbi.nlm.nih.gov/pubmed/8606899 (accessed on 10 September 2018). [CrossRef]

35. Cohen, J. A power primer. *Psychol. Bull.* **1992**, *112*, 155–159. Available online: https://www.ncbi.nlm.nih.gov/pubmed/19565683 (accessed on 10 September 2018). [CrossRef]

36. Bousie, J.A.; Blanch, P.; McPoil, T.G.; Vicenzino, B. Contoured in-shoe foot orthoses increase mid-foot plantar contact area when compared with a flat insert during cycling. *J. Sci. Med. Sport* **2013**, *16*, 60–64. Available online: http://www.ncbi.nlm.nih.gov/pubmed/22658845 (accessed on 10 September 2018). [CrossRef]

37. Yeo, B.K.; Bonanno, D.R. The effect of foot orthoses and in-shoe wedges during cycling: A systematic review. *J. Foot Ankle Res.* **2014**, *7*, 31. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4064526/ (accessed on 10 September 2018). [CrossRef]

38. McCormick, C.J.; Bonanno, D.R.; Landorf, K.B. The effect of customised and sham foot orthoses on plantar pressures. *J. Foot Ankle Res.* **2013**, *6*, 19. Available online: http://www.ncbi.nlm.nih.gov/pubmed/23680496 (accessed on 10 September 2018). [CrossRef]

39. D'Artibale, E.; Laursen, P.B.; Cronin, J.B. Profiling the physical load on riders of top-level motorcycle circuit racing. *J. Sports Sci.* **2017**. [CrossRef]

40. D'Artibale, E.; Tessitore, A.; Capranica, L. Heart rate and blood lactate concentration of male road-race motorcyclists. *J Sports Sci.* **2008**, *26*, 683–689. Available online: https://www.ncbi.nlm.nih.gov/pubmed/18409099 (accessed on 10 September 2018). [CrossRef]

Validity of the Walked Distance Estimated by Wearable Devices in Stroke Individuals

Maxence Compagnat [1,2,*], Charles Sebiyo Batcho [3,4], Romain David [2], Nicolas Vuillerme [5], Jean Yves Salle [1,2], Jean Christophe Daviet [1,2] and Stéphane Mandigout [1]

[1]   (Handicap, Aging, Autonomy, Environment) HAVAE EA6310, University of Limoges, 87042 Limoges Cedex, France; JYVES.SALLE@chu-limoges.fr (J.Y.S.); jean-christophe.daviet@unilim.fr (J.C.D.); mandigout@unilim.fr (S.M.)

[2]   Department of Physical Medicine and Rehabilitation, University Hospital Center of Limoges, 87042 Limoges Cedex, France; Romain-david@hotmail.fr

[3]   Center for Interdisciplinary Research in Rehabilitation and Social Integration (CIRRIS), Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale Nationale (CIUSSS-CN), G1M 2S8 Quebec, QC, Canada; charles.batcho@fmed.ulaval.ca

[4]   Department of Rehabilitation, Faculty of Medicine, Université Laval, G1M 2S8 Quebec, QC, Canada

[5]   AGEIS, University Grenoble Alpes, Grenoble, 38 041 Grenoble CEDEX 9, France; nicolas.vuillerme@univ-grenoble-alpes.fr

*   Correspondence: maxence.compagnat@icloud.com; Tel.: +33-555-055-555; Fax: +33-555-056-513

Abstract: Background: Health professionals need valid devices to assess a stroke individual's ability to walk. The aim was to evaluate the validity of the estimation of the walked distance by wearable devices and the impact of the sensor's position in stroke individuals. Methods: Post-stroke patients able to walk without human assistance were equipped with several wearable devices: pedometers, Actigraph, and Sensewear Armband placed according to the manufacturers' recommendations. Participants walked for 6 min at a comfortable speed wearing all sensors at the same time. We analyzed the validity of sensor-estimated distances according to their position using Bland–Altman analysis, root-mean-square error, and coefficient of correlation. Results: In total, 35 individuals were included (mean age = 65 ± 15 years). The best estimations were given by the Actigraph worn on the unaffected ankle (mean bias (MB) = 22.6 ± 32.4 m; $p = 0.37$) and by the pedometer worn on the unaffected hip (MB = 20.5 ± 24.6 m; $p = 0.46$). The other sensors and positions provided large estimation errors over 95 m ($p < 0.05$). Conclusion: This study led to a recommendation of a pedometer worn on the unaffected hip or an Actigraph worn on the unaffected ankle to get a valid estimation of the distance walked by stroke individuals.

1. Introduction

Walking is a key activity to improve the quality of life and social participation in individuals with stroke sequelae [1,2]. However, studies showed that 30% of patients do not recover the ability to walk independently six months after the stroke, and nearly 65% consider that their social participation is limited [3]. The most often reported parameters in the literature to estimate an individual's capacity to walk in a social environment are the speed—commonly evaluated with a 10-m test—and the distance (d), assessed with a 6-min walk test [4]. Perry et al. demonstrated that a speed over 0.4 ms^{-1} was the threshold defining the ability to walk in a social environment [5]. However, there may be a gap between a subject's performance in these tests and the actual capacity to walk freely due to obstacles in the environment [6,7]. This is especially true since the distances walked in social situations are

significantly longer than in tests [8]. For example, Salbach et al. showed in a study including 24 cities around the world that an individual has to travel a mean distance between 57 and 98 m to go to a post office, between 77 m and 114 m to a place of worship, 260 m for a doctor appointment (SD: 78 m), and 129–381 m to go to a shopping mall [4].

To assess the individual's real walking ability outside of institutions, health professionals need reliable and practical tools that can be used in daily life situations [7,9]. The most commonly used tools to measure the walked capacity in terms of distance or number of steps are global positioning systems (GPSs), pedometers, and accelerometers [10]. On the one hand, the number of steps tends to be an unreliable parameter for medical practitioners as the step length is variable depending on the type and level of motor deficiency, which causes a lower comparability between individuals [11,12]. On the other hand, distance is more reliable, with better comparability between individuals and readily accessible data pertaining to social participation in the literature [2,4,8,13]. It can be measured accurately using a GPS, but its use is limited to outdoor situations due to the necessity of satellite synchronization [14].

Accelerometers and pedometers can be used indoors and were validated for estimations of the walked distance in healthy populations [15,16]. These systems generally use the product of the step count and the step length. Step length is evaluated by measuring the distance traveled over 20 steps [15]. However, this method remains debatable in post-stroke subjects because of the variability of a given individual's step length [17]. Furthermore, Crouter et al. reported that the accuracy of these devices sharply drops when individuals walk slower than $0.8~\mathrm{ms}^{-1}$ [15], and post-stroke patients typically tend to walk slower than healthy individuals [18]. Caroll et al. indeed demonstrated in a population of 51 post-stroke individuals that pedometers placed on the hip or around the neck were unable to detect steps when walking slower than $0.5~\mathrm{ms}^{-1}$ [19]. Other types of sensors are, however, able to make accurate estimations of the number of steps even in slow-walking individuals. For instance, Fulk et al. demonstrated that the accelerometer-based step activity monitor had an estimation error of five steps in a 2-min walk test with an intraclass coefficient correlation (ICC) of 0.97, even in subjects walking slower than $0.5~\mathrm{ms}^{-1}$ [20]. Additionally, Klassen et al. demonstrated that the location of the sensor had a direct impact on the results; a triaxial accelerometer reported a standard error lower than 15% when placed on the ankle, while its standard error was greater than 80% when placed on the hip in a study including 43 post-stroke subjects. The impact of the sensor's position was even greater in individuals who walked slowly ($<0.5~\mathrm{ms}^{-1}$) [21].

Naturally, when confronted with a large number of possible devices and positions recommended by the manufacturer, medical practitioners and users are at a loss when it comes to choosing the right sensor and position to obtain the best estimation of walked distance. Therefore, it seems essential to check the validity of each option to point toward a global recommendation in post-stroke populations.

The objective of this study was to evaluate the validity of the walked distance estimations by pedometers and accelerometers and to evaluate the impact of their position (ankle, hip, wrist/affected, unaffected side) in a post-stroke population walking at a comfortable speed.

2. Materials and Methods

2.1. Participants Selection

Participants were recruited in the Physical Medicine and Rehabilitation Department. The inclusion criteria were (1) a single stroke confirmed with brain imaging, and (2) the ability to walk continuously for 6 min without human assistance. Participants with acute cardiac or respiratory pathologies or decompensated chronic pathologies were excluded.

The health professional responsible for the protocol informed the patients of the details of the protocol and registered their consent. The research protocol was accepted by an ethics committee, notice number CERNI 2015-01-13-57.

2.2. Assessment of the Hemiplegia

Motor function was evaluated using the Demeurisse motricity index [22]. Spasticity was evaluated using the modified Ashworth scale (MAS) [23]. Walking autonomy was assessed using the functional ambulation classification modified (FACm) [24]. Autonomy related to activities of daily living was evaluated using the Barthel index [25]. All these evaluations were performed by the same examiner for all participating subjects.

2.3. Instrumentation

We selected several accelerometers with various technological features which were previously used in several studies on physical activity in post-stroke populations [26–28]. This selection was based on the type of device (pedometer [19], accelerometer [29], multisensor [30]) and the position options [31] (additional information is provided in Table S1).

2.3.1. Actigraph GT3x

The Actigraph GT3x (Actigraph LLC, Pensacola, FL, USA) is the most widely used accelerometer for physical activity evaluations in clinical research [32]. The Actigraph GT3x is a small (4.6 cm × 3.3 cm × 1.5 cm) and lightweight (42.5 g) triaxial accelerometer designed to measure accelerations in the range of 0.05–2 g with a band-limited frequency of 0.25–2.50 Hz. This corresponds to the range in which most human activities are performed [33]. We chose a standard configuration with a standard sampling frequency of 30 Hz and no specific filtering. The device was initialized using 1-s epochs. The Actigraph data can be downloaded to a personal computer via a reader interface unit. In our work, we used the step count estimate provided by the ActigraphGT3x. It can be worn on the wrist, hip, or ankle [34]. We chose to place the devices on the wrists and ankles on the affected and unaffected sides, and on the hip on the unaffected side. The sensor was only placed on the unaffected hip because Rand et al. [35] reported that there was no difference in the estimation of the number of steps whether the sensor was on the affected or unaffected hip. This was, therefore, done to simplify the protocol. The sensor was attached to the different placements by an elastic band provided by the manufacturer. This choice was made to judge the impact of the impairments and the positioning of the estimations of the sensor.

2.3.2. Sensewear Armband

A multisensor array (Sensewear Armband, Body Media, Pittsburgh, PA, USA) was positioned on the back of the participant's arm, midway between the shoulder and elbow. It measures 7 cm × 7 cm × 2 cm and weighs 55 g. It gathers raw physiological data of movements using a biaxial accelerometer and records heat flux, skin temperature, and galvanic skin response. Cycles of 1 s were used in this study, i.e., accelerations were integrated over 1-s periods to produce a number of counts for each second, which was written into the internal memory. No specific filtering was used. At the end of each trial, the stored Sensewear Armband data were downloaded to a personal computer for analysis using the reader interface unit supplied by the manufacturer. We used the step count estimate provided by Sensewear Armband for our work. We chose to place the devices on the affected and unaffected arm to judge the impact of the impairments of the estimations of the sensor. The sensor was attached to the arm by an elastic band provided by the manufacturer.

2.3.3. Pedometer (ONStep 400, Geonaute)

The ONStep 400 (Decathlon France S.A.S, Villeneuve d'Ascq, France) is an inexpensive pedometer available to the general public. Its size is 6.5 cm × 4.1 cm × 1.5 cm, and its weight is 40 g. It can be worn on the hip, in the pockets of trousers, or around the neck using a strap [36]. The data recorded by the pedometer can be observed directly on the device, which does not require specific software. The recording is triggered when the subject initiates the first step and ends when the subject is no longer

in motion. This device uses a piezoelectric mechanism to count the number of steps. We used the step count estimate provided by the pedometer for our work. We chose to place the devices on the unaffected hip and around the neck to judge the impact of the placement on the estimations of the sensor. The sensor was attached to the limb by an elastic band for hip placement and a fabric strap provided by the manufacturer for placement around the neck.

2.4. Walked Distance

The participants were required to walk at a comfortable speed on a circuit with no half-turn or obstacle, graduated every 5 m. The distance walked by each patient was measured by the examiner using the scales marked on the ground. The estimated walked distance was calculated by multiplying the number of steps reported by the wearable devices by the average step length. The average step length was calculated using the method described by Bassett and Crouter [15,16]. This method consists of asking the participant to walk 20 steps at a comfortable speed and then measuring the distance walked. The average step length was then obtained by dividing this distance by 20. To weigh the variability of this measurement, we asked the subject to repeat the operation three times, after which we calculated the average step length from the three trials.

2.5. Test Protocol

The test protocol involved the following steps:

1. Measurement of the average step length over three trials of 20 steps.
2. Installation of the sensors. Actigraph GT3x devices were placed on the wrists and ankles on both the affected and unaffected sides, as well as at the unaffected hip. Sensewear Armbands were placed on both the affected and unaffected arms. Pedometers were placed at the unaffected hip and around the neck. The device placements are illustrated in Figure 1.
3. The participants performed a six-minute walk test at a comfortable walking speed. During this walking period, the distance walked was measured by the examiner with the graduations marked on the floor of the corridor.
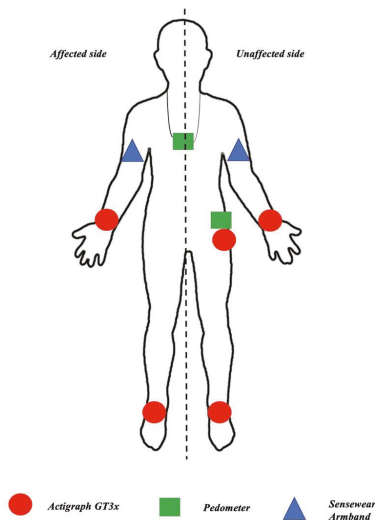4. Download of the data from all devices.

Figure 1. Placement of the devices.

2.6. Statistical Analysis

Based on the work of Carroll et al., we calculated the necessary sample size by predicting a mean difference of 10% and a standard deviation of 20% between the estimated and measured distances [19]. Thus, from a Bland–Altman analysis, we planned to include a total of 32 subjects in order to reach 80% power with an alpha-risk of 0.05 [37]. The validity of the walked distance estimated by wearable devices was determined by analyzing the accuracy, and the agreement for each device in comparison to the criterion measure. These were defined by the mean bias (MB) and 95% limits of agreement (95% LoA) on the Bland–Altman analysis, root-mean-square error (RMSE), Pearson's correlation coefficient (*r*), and coefficient of determination. We performed a paired-sample Wilcoxon test to analyze the significance of the difference between the measured distance and the distance estimated by each sensor according to their position. The threshold of significance was 0.05. All calculations were performed using the RealStats 2011 software (Real Statistics Using Excel© 2012–2019, Charles Zaiontz).

3. Results

3.1. Population Characteristics

We included 35 subjects with a mean age of 64.6 ± 14.8 years. The participants showed heterogeneous levels of deficiency and autonomy (see Table 1). The mean step length was 0.46 ± 0.11 m.

Table 1. Characteristics of the population. SD: standard deviation; BMI: body mass index; FACm: functional ambulation classification modified; MAS: modified Ashworth scale.

	MEAN / MEDIAN	SD	MIN	MAX
AGE (YEAR)	64.60	14.80	34	88
BMI (KG·M^{-2})	26.70	5.50	20	43
TIME AFTER STROKE (DAYS)	781	1492	9	5110
DEMEURISSE UPPER LIMB SCORE (/100)	68		1	100
DEMEURISSE LOWER LIMB SCORE (/100)	77		43	100
MAS (/5)	1		0	4
BARTHEL INDEX (/100)	74		40	100
FACM (/8)	5		4	8
SPEED (MS^{-1})	0.56	0.30	0.06	1.22

3.2. Validity of the Analysis

Across all recording devices, the Actigraph data of one patient were lost and six armband devices (17% of armbands) had a recording failure. The reasons for these recording issues were not found.

The mean walked distance measured by the examiner at the end of the test was 208.2 ± 109.2 m. No significant differences were found between the distance measured by the examiner and the following devices: the Actigraph worn on the unaffected ankle (d = 185.7 ± 100.6 m; *p* = 0.37), the Actigraph worn on the affected ankle (d = 175.7 m ± 109.3 m; *p* = 0.21), the pedometer worn around the neck (d = 183.4 m ± 147.2 m; *p* = 0.32), and the pedometer worn at the non-affected hip (d = 231.1 m ± 121.2 m; *p* = 0.46). For all other combinations of sensors and locations, there was a significant measurement error with *p* < 0.001 (Figure 2). The step count of each device is illustrated in Table 2.
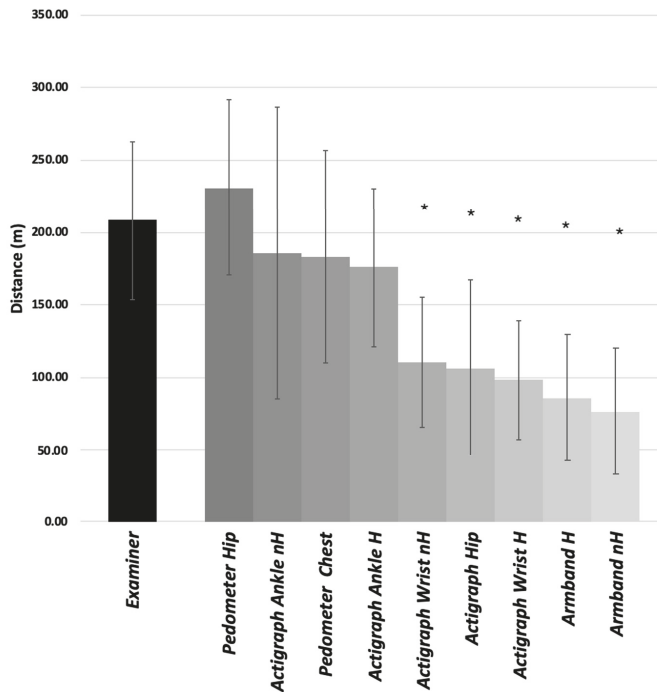
Figure 2. Distance measured by the practitioner and estimated from step counts reported by all devices according to type and placement. nH: non hemiparetic side; H: hemiparetic side. * $p < 0.05$ at the end of the Wilcoxon test comparing the distance measured by the examiner and that estimated by the device from the number of steps.

Table 2. Step count reported by the devices. nH: non hemiparetic side; H: hemiparetic side.

	Pedometer Hip	Actigraph Ankle nH	Pedometer Chest	Actigraph Ankle H	Actigraph Wrist nH	Actigraph Hip	Actigraph Wrist H	Armband H	Armband nH
Mean step count (step)	514	410	406	387	237	221	212	195	170
SD step count (step)	251	188	295	216	166	235	161	249	196

3.3. Validity Parameters

The parameters of validity of each sensor are summarized in Table 3. The most accurate estimations were obtained using the pedometer worn at the hip on the non-affected side (MB = 9.7%, RMSE = 10.9%) and the Actigraph placed at the ankle on the non-affected side (MB = 10.7%, RMSE = 14.6%). On the other hand, the pedometer worn on the hip and the Actigraph worn on the ankle on the affected side showed the best coefficients of correlation ($r > 0.90$) and the lowest limits of agreement. The Actigraph worn on the ankle on the unaffected side had lower correlation ($r = 0.93$) and higher 95% LoA (111.4; −46.4 m) compared to the same device located on the affected limb, even though no statistical difference was observed between the two estimations.

Table 3. Validity parameters of distance estimated by wearable devices versus distance measured by examiner. Unit in meters; nH: non hemiparetic side; H: hemiparetic side; percentage difference: mean bias expressed in percentage of distance measured and estimated by device; 95% LoA: limits of agreement of Bland–Altman analysis; r = Pearson correlation coefficient; p = statistical significance of Pearson correlation coefficient; RMSE: root-mean-square error; percentage RMSE: RMSE expressed in percentage of distance measured by examiner.

	Mean Bias (m)	Percentage Difference (%)	95% LoA Up (m)	95% LoA Down (m)	Percentage 95%LoA (%)	r	p	RMSE (m)	Percentage RMSE (%)
Distance Actigraph Ankle nH	22.58	10.70%	87.45	−42.29	30.80%	0.95	<0.001	30.79	14.60%
Distance Actigraph Ankle H	32.50	15.40%	111.39	−46.38	37.40%	0.93	<0.001	40.20	19.00%
Distance Actigraph Hip	101.78	48.30%	222.37	−18.81	57.20%	0.86	<0.001	62.28	29.50%
Distance Actigraph Wrist nH	97.55	46.30%	228.04	−32.93	61.90%	0.79	<0.001	55.08	26.10%
Distance Actigraph Wrist H	110.04	52.20%	237.47	−17.39	60.50%	0.81	<0.001	49.24	23.30%
Distance Armband nH	127.26	60.40%	286.80	−32.28	75.70%	0.68	<0.001	65.92	31.30%
Distance Armband H	120.62	57.20%	288.25	−47.01	79.60%	0.72	<0.001	83.01	39.40%
Distance Pedometer Chest	27.20	12.90%	156.42	−102.02	61.30%	0.91	<0.001	61.67	29.20%
Distance Pedometer Hip	−20.51	−9.70%	28.68	−69.70	23.30%	0.98	<0.001	23.12	10.90%

4. Discussion

The objective of this work was to evaluate the validity of the estimations of walked distance by wearable devices in individuals with stroke sequelae. We observed that the best estimations over a 6-min walk at comfortable speed were provided by the pedometer worn at the hip on the non-affected side (MB = 9.7%, RMSE = 10.9%) and by the Actigraph worn at the ankle on the non-affected side (MB = 10.7%, RMSE = 14.6%).

We observed significant differences between the combinations of sensor type and position, which demonstrates the impact of these parameters. For instance, despite being placed in the same location (hip on the non-affected side), the pedometer provided a better estimation of the walked distance than the Actigraph, even though the first is a piezoelectric device and the second is a triaxial accelerometer. On the one hand, the Actigraph's measurement error was considerable (MB = 101.8 ± 60.1 m; RMSE = 60.3 m) compared to the pedometer (MB = 20.5 ± 24.6 m; RMSE = 23.1 m). This could potentially be explained by an issue in the settings of the Actigraph or in its algorithm. The chosen settings were standard, i.e., a standard sampling frequency (30 Hz) and no specific filtering. It is possible that the use of other algorithms or other settings may alter the accuracy of the Actigraph GT3X. The manufacturer of this device recently published an add-on called "low-frequency filter extension" which can be enabled for healthy individuals with low amounts of physical activity [34]. This add-on lowers the detection threshold of the Actigraph to improve the data acquisition sensitivity [34,38]. It would be relevant to evaluate the impact of these settings on the estimation of the subject's walked distance, especially in individuals with limited walking capabilities.

On the other hand, the pedometer might have a lower detection threshold, as it is specifically designed to count the number of steps of an individual. This would, therefore, induce a better sensitivity in this population. However, we are unable to confirm these hypotheses since we could not analyze the raw data used by the sensors. Like the Actigraph, the Sensewear Armband had a

consequential measurement error despite being a multisensor validated for the assessment of energy expenditure in post-stroke populations, with an MB of 127.3 ± 79.8 m and RMSE = 65 m when placed on the affected side, and an MB of 120.6 m ± 83.8 m and RMSE = 80 m when placed on the unaffected side. These results are consistent with those reported by Manns et al., who observed an error of 193.1 ± 168.1 steps in two six-minute walk tests in a sample of 12 post-stroke subjects [30]. Similar results were also reported by Vanroy et al., whose study reported an estimation error between 110 and 190 steps using the Sensewear Armband after walking 120 m at a comfortable speed in a group of 14 post-stroke subjects [39]. Thus, this device does not seem reliable to estimate the number of steps in post-stroke subjects. The cause may be the device's algorithm, which might fail to correctly count the number of steps, but the impossibility of accessing the raw data prevented us from confirming this hypothesis. In post-stroke populations, we would recommend limiting the use of this sensor to its main function, i.e., assessing the energy expenditure.

The sensor's position on the body also had a significant impact, as proven by the differences in the results of the Actigraph between the ankle, hip, and wrist. The estimation of the Actigraph worn at the ankle was the closest to the measured walked distance (MB = 22.6 ± 32.4 m), while the same device worn at the hip or on the wrist had MB values over 95 m (Table 3). It is possible that placing the Actigraph on the ankle provides a better exposure to the accelerations of the limb, which would enable a better acquisition of the number of steps. This hypothesis is supported by Klassen et al., who demonstrated that placing a Fitbit triaxial accelerometer on the ankle provided a lower estimation error, from 84.6% ± 30.5% when placed at the waist to 15.8% ± 22.3% when placed on the ankle in a study including 43 post-stroke subjects [21].

4.1. Strengths of the Study

This study brings light on the precision of several types of sensors and the impact of their position on the estimation of the distance walked by individuals with stroke sequelae. Overall, the devices had an estimation error of about 100 m after walking for 6 min, with the exception of the pedometer and the Actigraphs worn on the ankles, which provided better results. Knowing that stroke survivors walk at an average speed of 0.5 ms^{-1} (1.8 kmh^{-1}), this estimation error would cause an underestimation of about 1 km per hour of walking, which would amount to more than 50% of the measured distance. The most accurate estimation was obtained using the Actigraph on the ankle on the unaffected side and the pedometer on the hip on the unaffected side with mean standard error of about 20 m after a 6-min walk, which corresponds to 200 m per hour of walking. Therefore, according to these results, medical practitioners should use either the Actigraph placed on the unaffected ankle or the pedometer placed at the hip. Between these two choices, the pedometer appears to be an attractive solution due to its low price. Additionally, our results bring to both the practitioner's and the end-user's attention the accuracy issues of these devices and the necessity to wisely choose the position and type of the sensor to avoid unreliable estimations.

4.2. Limitations

We are aware of two limitations of our study. The first limitation is the method used to estimate the walked distance. We used the average step length, which can be variable in one individual during a single activity or along the day, especially in the stroke population [11,17]. The sensor was set on a constant step length, which might have induced an overestimation of the distance if it counted one full step while the subject was limping. It may be necessary to analyze the accuracy of these devices in a real-life situation over several days at home to get more precise data. Even so, the technique we selected to estimate the walked distance using the subject's average step length enabled us to make an initial configuration of these devices for a population of post-stroke subjects, which seems critical to improve the use of these tools in a clinical setting. The second limitation of our study pertains to the external validity of the results. A plethora of physical activity trackers exist, and many of them could have been included in the study (e.g., step activity monitor, smartphones, smartwatches, etc.).

We intended to use the three most commonly used technologies: piezoelectric devices (pedometers), triaxial accelerometers, and multisensors [32]. However, our results may be difficult to expand to other untested devices due to the possible impact of their specific algorithms on the estimated number of steps.

5. Conclusions

The validity of the estimations of the walked distance by wearable sensors varied significantly according to the device's type and location. Following our results, we recommend using a pedometer (piezoelectric device) worn on the hip on the unaffected side, or the Actigraph activity monitor (triaxial accelerometer) worn on the ankle on the unaffected side to estimate the walked distance in individuals with neurological sequelae of stroke.

- The sensor type and its location on the body strongly impact the estimation of the walked distance in individuals with stroke sequelae.
- The pedometer (piezoelectric device) placed on the hip and the Actigraph activity monitor (triaxial accelerometer) worn on the hip on the non-affected side provided the closest estimations of the walked distance.
- Placing an Actigraph on the upper limbs caused a significant underestimation of the walked distance in individuals with stroke sequelae.
- The Sensewear Armband strongly underestimated the walking distance regardless of its placement on the affected or unaffected upper limb of the stroke individuals.

References

1. Barclay, R.E.; Stevenson, T.J.; Poluha, W.; Ripat, J.; Nett, C.; Srikesavan, C.S. Interventions for improving community ambulation in individuals with stroke. *Cochrane Database Syst Rev.* **2015**, *13*, CD010200. [CrossRef]
2. Lord, S.E.; McPherson, K.; McNaughton, H.K.; Rochester, L.; Weatherall, M. Community ambulation after stroke: How important and obtainable is it and what measures appear predictive? *Arch. Phys. Med. Rehabil.* **2004**, *85*, 234–239. [CrossRef]
3. Mayo, N.E.; Wood-Dauphinee, S.; Côté, R.; Durcan, L.; Carlton, J. Activity, participation, and quality of life 6 months poststroke. *Arch. Phys. Med. Rehabil.* **2002**, *83*, 1035–1042. [CrossRef] [PubMed]
4. Salbach, N.M.; O'Brien, K.; Brooks, D.; Irvin, E.; Martino, R.; Takhar, P.; Chan, S.; Howe, J.-A. Speed and Distance Requirements for Community Ambulation: A Systematic Review. *Arch. Phys. Med. Rehabil.* **2014**, *95*, 117–128. [CrossRef]
5. Perry, J.; Garrett, M.; Gronley, J.K.; Mulroy, S.J. Classification of Walking Handicap in the Stroke Population. *Stroke* **1995**, *26*, 982–989. [CrossRef]

6. Patla, A.E.; Shumway-Cook, A. Dimensions of Mobility: Defining the Complexity and Difficulty Associated with Community Mobility. *J. Aging Phys. Act.* **1999**, *7*, 7–19. [CrossRef]

7. Corrigan, R.; McBurney, H. Community ambulation: Environmental impacts and assessment inadequacies. *Disabil. Rehabil.* **2008**, *30*, 1411–1419. [CrossRef]

8. Andrews, A.W.; Chinworth, S.A.; Bourassa, M.; Garvin, M.; Benton, D.; Tanner, S. Update on distance and velocity requirements for community ambulation. *J. Geriatr. Phys. Ther.* **2010**, *33*, 128–134.

9. Billinger, S.A.; Arena, R.; Bernhardt, J.; Eng, J.J.; Franklin, B.A.; Johnson, C.M.; MacKay-Lyons, M.; Macko, R.F.; Mead, G.E.; Roth, E.J.; et al. Physical Activity and Exercise Recommendations for Stroke Survivors: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **2014**, *45*, 2532–2553. [CrossRef] [PubMed]

10. English, C.; Manns, P.J.; Tucak, C.; Bernhardt, J. Physical Activity and Sedentary Behaviors in People With Stroke Living in the Community: A Systematic Review. *Phys. Ther.* **2014**, *94*, 185–196. [CrossRef]

11. Balasubramanian, C.K.; Neptune, R.R.; Kautz, S.A. Variability in spatiotemporal step characteristics and its relationship to walking performance post-stroke. *Gait Posture* **2009**, *29*, 408–414. [CrossRef]

12. Chen, G.; Patten, C.; Kothari, D.H.; Zajac, F.E. Gait differences between individuals with post-stroke hemiparesis and non-disabled controls at matched speeds. *Gait Posture* **2005**, *22*, 51–56. [CrossRef]

13. Lee, K.B.; Lim, S.H.; Ko, E.H.; Kim, Y.S.; Lee, K.S.; Hwang, B.Y. Factors related to community ambulation in patients with chronic stroke. *Top. Stroke Rehabil.* **2015**, *22*, 63–71. [CrossRef] [PubMed]

14. Mahendran, N.; Kuys, S.S.; Downie, E.; Ng, P.; Brauer, S.G. Are Accelerometers and GPS Devices Valid, Reliable and Feasible Tools for Measurement of Community Ambulation After Stroke? *Brain Impair.* **2016**, *17*, 151–161. [CrossRef]

15. Crouter, S.E.; Schneider, P.L.; Karabulut, M.; Bassett, D.R. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med. Sci. Sports Exerc.* **2003**, *35*, 1455–1460. [CrossRef] [PubMed]

16. Schneider, P.L.; Crouter, S.E.; Lukajic, O.; Bassett, D.R. Accuracy and Reliability of 10 Pedometers for Measuring Steps over a 400-m Walk. *Med. Sci. Sports Exerc.* **2003**, *35*, 1779–1784. [CrossRef]

17. Chisholm, A.E.; Makepeace, S.; Inness, E.L.; Perry, S.D.; McIlroy, W.E.; Mansfield, A. Spatial-temporal gait variability poststroke: Variations in measurement and implications for measuring change. *Arch. Phys. Med. Rehabil.* **2014**, *95*, 1335–1341. [CrossRef] [PubMed]

18. Reisman, D.S.; Rudolph, K.S.; Farquhar, W.B. Influence of Speed on Walking Economy Poststroke. *Neurorehabil. Neural Repair* **2009**, *23*, 529–534. [CrossRef]

19. Carroll, S.L.; Greig, C.A.; Lewis, S.J.; McMurdo, M.E.; Sniehotta, F.F.; Johnston, M.; Johnston, D.W.; Scopes, J.; Mead, G.E. The Use of Pedometers in Stroke Survivors: Are They Feasible and How Well Do They Detect Steps? *Arch. Phys. Med. Rehabil.* **2012**, *93*, 466–470. [CrossRef]

20. Fulk, G.D.; Combs, S.A.; Danks, K.A.; Nirider, C.D.; Raja, B.; Reisman, D.S. Accuracy of 2 activity monitors in detecting steps in people with stroke and traumatic brain injury. *Phys. Ther.* **2014**, *94*, 222–229. [CrossRef]

21. Klassen, T.D.; Simpson, L.A.; Lim, S.B.; Louie, D.R.; Parappilly, B.; Sakakibara, B.M.; Zbogar, D.; Eng, J.J. "Stepping Up" activity poststroke: Ankle-positioned accelerometer can accurately record steps during slow walking. *Phys. Ther.* **2016**, *96*, 355–360. [CrossRef]

22. Demeurisse, G.; Demol, O.; Robaye, E. Motor evaluation in vascular hemiplegia. *Eur. Neurol.* **1980**, *19*, 382–389. [CrossRef] [PubMed]

23. Bohannon, R.W.; Smith, M.B. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys. Ther.* **1987**, *67*, 206–207. [CrossRef] [PubMed]

24. Brun, V.; Mousbeh, Z.; Jouet-Pastre, B.; Benaim, C.; Kunnert, J.E.; Dhoms, G.; d'Angeli-Chevassut, M.; Torres, B.; Pélissier, J. Évaluation clinique de la marche de l'hémiplégique vasculaire: Proposition d'une modification de la functional ambulation classification. *Ann. Phys. Rehabil. Med.* **2000**, *43*, 14–20. [CrossRef]

25. Mahoney, F.I.; Barthel, D.W. Functional evaluation: The Barthel Index. *Md. State Med. J.* **1965**, *14*, 61–65. [PubMed]

26. Bakken, L.N.; Kim, H.S.; Finset, A.; Lerdal, A. Stroke patients' functions in personal activities of daily living in relation to sleep and socio-demographic and clinical variables in the acute phase after first-time stroke and at six months of follow-up. *J. Clin. Nurs.* **2012**, *21*, 1886–1895. [CrossRef] [PubMed]

27. English, C.; Healy, G.N.; Coates, A.; Lewis, L.K.; Olds, T.; Bernhardt, J. Sitting time and physical activity after stroke: Physical ability is only part of the story. *Top. Stroke Rehabil.* **2016**, *23*, 36–42. [CrossRef]

28. van der Ploeg, H.P.; Streppel, K.R.M.; van der Beek, A.J.; van der Woude, L.H.V.; Vollenbroek-Hutten, M.; van Mechelen, W. The Physical Activity Scale for Individuals with Physical Disabilities: Test-retest reliability and comparison with an accelerometer. *J. Phys. Act. Health* **2007**, *4*, 96–100. [CrossRef]

29. Gebruers, N.; Vanroy, C.; Truijen, S.; Engelborghs, S.; De Deyn, P.P. Monitoring of Physical Activity after Stroke: A Systematic Review of Accelerometry-Based Measures. *Arch. Phys. Med. Rehabil.* **2010**, *91*, 288–297. [CrossRef] [PubMed]

30. Manns, P.J.; Haennel, R.G. SenseWear Armband and Stroke: Validity of Energy Expenditure and Step Count Measurement during Walking. *Stroke Res. Treat.* **2012**, *2012*, 247165. [CrossRef]

31. Mandigout, S.; Lacroix, J.; Ferry, B.; Vuillerme, N.; Compagnat, M.; Daviet, J.-C. Can energy expenditure be accurately assessed using accelerometry-based wearable motion detectors for physical activity monitoring in post-stroke patients in the subacute phase? *Eur. J. Prev. Cardiol.* **2017**, *24*, 2009–2016. [CrossRef]

32. Bassett, D.R.; Troiano, R.P.; McClain, J.J.; Wolff, D.L. Accelerometer-based physical activity: Total volume per day and standardized measures. *Med. Sci. Sports Exerc.* **2015**, *47*, 833–838. [CrossRef] [PubMed]

33. Farrahi, V.; Niemelä, M.; Kangas, M.; Korpelainen, R.; Jämsä, T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait Posture* **2019**, *68*, 285–299. [CrossRef]

34. ActiGraph GT3X+. Available online: http://actigraphcorp.com/support/activity-monitors/gt3xplus/ (accessed on 6 March 2018).

35. Rand, D.; Eng, J.J.; Tang, P.-F.; Jeng, J.-S.; Hung, C. How Active Are People With Stroke? Use of Accelerometers to Assess Physical Activity. *Stroke* **2009**, *40*, 163–168. [CrossRef]

36. Podomètre ONstep 400—Geonaute. Available online: https://customercare.geonaute.com/hc/fr/sections/201652082-Podom%C3%A8tre-ONstep-400 (accessed on 12 June 2018).

37. Lu, M.-J.; Zhong, W.-H.; Liu, Y.-X.; Miao, H.-Z.; Li, Y.-C.; Ji, M.-H. Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method. *Int. J. Biostat.* **2016**, *12*. [CrossRef]

38. Korpan, S.M.; Schafer, J.L.; Wilson, K.C.; Webber, S.C. Effect of ActiGraph GT3X+ Position and Algorithm Choice on Step Count Accuracy in Older Adults. *J. Aging Phys. Act.* **2015**, *23*, 377–382. [CrossRef]

39. Vanroy, C.; Vissers, D.; Cras, P.; Beyne, S.; Feys, H.; Vanlandewijck, Y.; Truijen, S. Physical activity monitoring in stroke: SenseWear Pro2 Activity accelerometer versus Yamax Digi-Walker SW-200 Pedometer. *Disabil. Rehabil.* **2013**, *20*, 1695–1703. [CrossRef] [PubMed]

40. Moore, S.A.; Hallsworth, K.; Bluck, L.J.C.; Ford, G.A.; Rochester, L.; Trenell, M.I. Measuring Energy Expenditure After Stroke Validation of a Portable Device. *Stroke* **2012**, *43*, 1660–1662. [CrossRef] [PubMed]

Article

A Sensor Array for the Measurement of Relative Motion in Lower Limb Prosthetic Sockets

Veronika Noll [1],*, Sigrid Whitmore [1,2], Philipp Beckerle [1,3] and Stephan Rinderknecht [1]

[1]   Institute for Mechatronic Systems in Mechanical Engineering, Technische Universität Darmstadt,
      Otto-Berndt-Str. 2, 64287 Darmstadt, Germany; beckerle@ims.tu-darmstadt.de (P.B.);
      rinderknecht@ims.tu-darmstadt.de (S.R.)
[2]   Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[3]   Elastic Lightweight Robotics Group, Robotics Research Institute, TU Dortmund University,
      44227 Dortmund, Germany
*   Correspondence: noll@ims.tu-darmstadt.de

Abstract: The relative motion between residual limb and prosthetic socket could be a relevant factor in quantifying socket fit. The measurement of these movements, particularly in dynamic gait situations, poses a challenging task. This paper presents the realization of a measurement concept based on multiple optical 2D-motion sensors. The performance of the system was evaluated on a test rig considering accuracy and precision as well as accomplished measurement frequency and reliability of the system. Additionally, results of a pilot study measuring the relative motion between residual limb and prosthetic socket at seven specific locations of one individual with transtibial amputation during straight level walking are presented. The sensor functionality of the array was confirmed and the test rig experiments were comparable to the previously tested functional model ($err_{\text{rel}} = 0.52 \pm 1.87\%$). With a sampling frequency of 1.3 kHz to be distributed among the number of sensor units, the developed system is suitable for investigating the relative movement between residual limb and prosthetic socket in dynamic gait situations. Results of the pilot study show the majority of relative motion occurring during the second half of the gait cycle. The measured relative motions show the residual limb sinking deeper into the socket, extending in the Sagittal plane and rotating internally in the Transverse plane during stance phase. Data captured during swing phase indicate a lower limb extension in the Sagittal plane as well as an external rotation in the Transverse plane.

1. Introduction

The socket is the mechanical interface between prosthesis and residual limb. It accounts for stability, ensures control over the prosthetic device, and determines its level of comfort. Thus, stump–socket interaction strongly influences the well-being and mobility of amputees [1]. To ensure the quality of socket fit, quantitative measures are advantageous. Pistoning or relative movement between residual lower limb and prosthesis is considered to be one indicative parameter of socket fit quality [2].

The most common techniques that are used to acquire data on the relative movement between residual limb surface and prosthetic socket are: motion capture [3–6], other optical means [7,8], inductive sensors [9–12], and vacuum pressure fluctuations of elevated vacuum suspension systems [11,13]. These approaches are subject to different drawbacks. For instance, vacuum pressure fluctuations cannot be used to evaluate the relative motion at specific or problematic locations within the interface. Additionally,

despite the number of different measurement approaches, studies presenting gait cycle dependent data of relative motion during ambulation are scarce [3,4,9,14].

A concept for measuring the relative motion between residual limb surface and prosthetic socket at specific locations in dynamic gait situations is presented in [15]. The proposed concept is based on optical 2D-motion sensor units, whose applicability has been tested experimentally using a functional model consisting of one sensor unit on a test rig as well as in biomechanical substitute studies [16].

This paper presents the subsequently realized low-cost measurement system capable of interfacing with an array of sensors. The implementation of the optimized measurement chain is presented in Section 2. In Section 3, the measurement system is evaluated experimentally on a test rig regarding accuracy, precision, achievable sampling frequency, and overall reliability. A pilot study using seven sensors to measure the relative motion between residual limb and prosthetic socket is reported in Section 4. Based on a discussion of the results, the paper is concluded in Section 5 where an outlook to future work is also given.

2. Measurement System

A schematic of the system components and measurement chain is shown in Figure 1. The system consists of up to eight sensor units and an electronics box containing a microcontroller connected to a PC via USB.
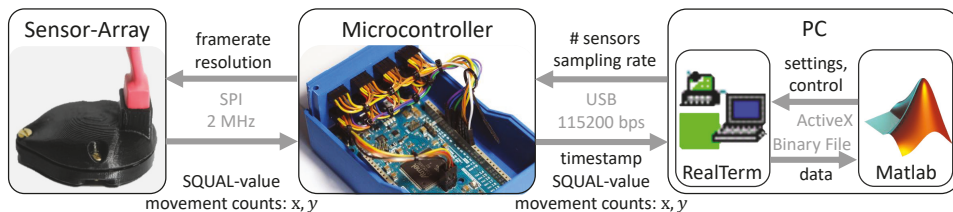

Figure 1. Realized measurement chain with indication of communication between system components: sensor array (**left**); microcontroller (**center**); and software running on PC (**right**).

A breakout version (https://www.tindie.com/products/jkicklighter/adns-9800-laser-motion-sensor/) of the ADNS-9800 optical sensor and accompanying ADNS-6190 lens (Avago Technologies, Broadcom Ltd., San Jose, CA, USA) was chosen due to its high-end specifications (30 g, 3.8 m/s, programmable maximum 12,000 fps and 8200 cpi). The sensor compares sequentially acquired images to mathematically derive resolution-dependent movement counts for the relative displacement along its two main axes x and y. Each sensor unit is protected by a case, which also serves as a mounting base for attachment to a prepared location on a prosthetic socket or to a testing bench. Both the protective case and mounting base are fabricated in polylactide (PLA) with a Fused-Deposition-Modeling 3D-printer. The sensors interface with the microcontroller over a Serial-Peripheral-Interface (SPI) connection.

The Arduino Due (Arduino AG, Ivrea, Italy) was chosen as a replacement for the Arduino Uno (Arduino AG, Ivrea, Italy) used in the functional model [16] due to its higher processor and SPI frequencies (64 MHz and up to 42 MHz) as well as compatibility with the existing firmware. It utilizes the sensor's burst mode register reading functionality to continuously acquire movement counts in the sensor's main axes x and y along with the surface quality (SQUAL-) value from the sensor. The SQUAL-value is a dimensionless value equal to 1/4 of the features observed by the sensor and is an indicator of surface texture. The microcontroller transmits these quantities and a timestamp in μs to a PC via a USB connection at a specified sampling frequency.

Incoming data are received and saved to a binary file by the serial terminal program RealTerm (https://realterm.sourceforge.io/). The measurement process is controlled from a Matlab (MathWorks, Natick, MA, USA) GUI, from which the number of sensors, sampling frequency per sensor and calibration factor are set. The calibration factor is calculated from uncalibrated measurements over

a known distance. After completion of a measurement, the GUI converts movement counts into displacement distance with Δd as the displacement in mm, c_x the movement counts, and k_x a dimensionless, surface-dependent calibration factor:

$$\Delta d \;=\; k_x c_x \frac{25.4 \text{ mm/inch}}{8200 \text{ cpi}}. \tag{1}$$

Optimization of the measurement chain with respect to sampling rate is achieved by minimizing register reading times, on-board processing, and the number of bytes transmitted to the PC. The sensor resolution and frame rate are set to their maximum values of 8200 cpi and 12,000 fps. The SPI frequency was set to 2 MHz and the baud rate to 115,200 bps. These modifications increase the system's top frequency from about 62.5 Hz for one sensor unit [16] to approximately 1.3 kHz, to be distributed among all attached sensors.

3. Test Rig Evaluation

The following section presents the methodology and results of the test rig evaluation of an array consisting of four sensor units.

3.1. Methodology

The system's performance was evaluated on a bi-axial test rig. Two linear drives (Indradyn, Bosch Rexroth, Lohr am Main, Germany) were mounted perpendicular to one another. Motion was controlled by a Rexroth MTX 13V programmable open-loop control system. The rig moved a cantilever in the horizontal plane over a base plate. A bracket with four attached sensors was mounted to the cantilever and a patch of liner material placed beneath. In this testing phase, a wiring junction printed circuit board (pcb) was used to connect the sensor units to the Arduino. The described measurement setup is shown in Figure 2.
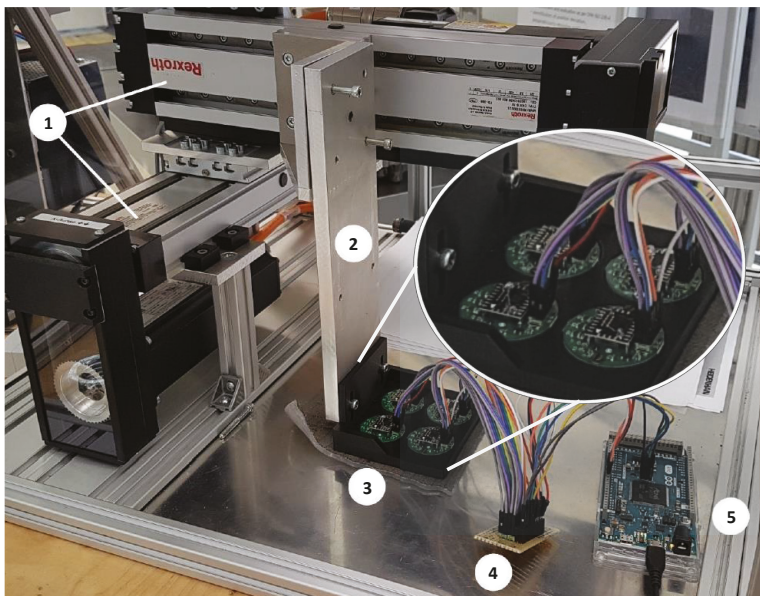

Figure 2. The test rig with linear drives (**1**); cantilever (**2**); test bracket, sensors and liner (**3**); wiring junction pcb (**4**); and Arduino Due (**5**) with USB cable leading to laptop; adapted from [17].

The sampling rate was set to 200 Hz per sensor, which meets the capabilities of commercially-available products for gait analysis (≥50 Hz). Measurements of uni-axial and diagonal motion, each consisting of fifteen motion steps, were conducted for each combination of distance (1, 5, 10, and 40 mm) and velocity (1, 10, and 100 mm/s). Complete measurement data (x and y movement counts, SQUAL-value, and timestamp) were continuously recorded. Movement counts were converted to displacements and scaled with the chosen calibration factor. The sections of each measurement containing motion were automatically extracted and the total displacement recorded for each step determined. The relative error with respect to the known displacement of the test rig Δx was determined with Δd, as determined in Equation (1):

$$err_{\text{rel}} = \frac{\Delta d - \Delta x}{\Delta x}.$$ (2)

3.2. Performance Results

Results are presented using data of a single sensor of the four-sensor array as a proxy for the behavior of the overall system. The calculated relative errors are shown in Figure 3. The bias of the mean (cross marker) indicates accuracy while the standard deviation (error bar) represent precision. Because of the limited influence of test rig velocity on sensor performance [16], velocities were combined for each distance. A bias of less than 0.9% was achieved for the x-direction at 40 mm. The y-direction achieved biases within 1.43% for uni-axial motion and 4.76% for diagonal motion at 40 mm. The standard deviation at 40 mm was less than 0.9% for uniaxial as well as diagonal movements.
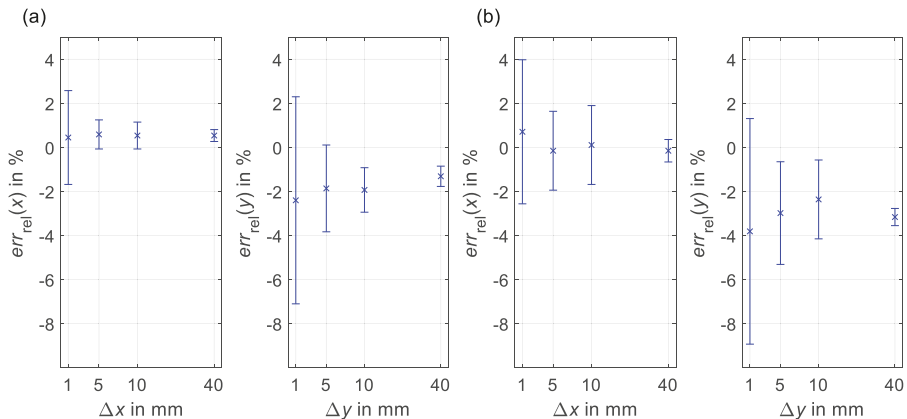
Figure 3. Mean (cross marker) and standard deviation (error bar) of the relative error for one sensor from (**a**) uni-axial and (**b**) diagonal movement. Velocities were combined at each distance. Calibration factors were $k_x = 2.018$ and $k_y = 1.987$.

The calculated relative errors demonstrate comparable results in both accuracy and precision to the functional model [16]. Although the behavior among sensors was comparable, differences were noted in the required calibration factors with values for the four sensors ranged between 1.976 and 2.018 for k_x and between 1.915 and 1.987 for k_y. As the thickness of the liner patch varied by approximately 1 mm along the diagonal, these differences likely stemmed from offset variations between the individual sensors and the liner. The system's effective resolution depends upon the necessary calibration factor and is in the low micrometer range (the chosen sensor resolution of 8200 cpi corresponds to 3.1 μm per count).

Sampling frequencies per sensor were confirmed for four sensors between 25 Hz and 275 Hz in 25 Hz increments. Across all frequencies, the difference between set sampling period and the average sampling period calculated from the recorded timestamps was 0.01 ± 0.15 μs.

Errors were found to occur only rarely. Out of 839 measurements, seven cases of transient behavior in both *x* and *y* displacements were recorded. Since this error affected only individual sensors and not all four, the source was unlikely to have been due to external factors, e.g., a bump to the test-rig. These errors were easily identifiable and localized, i.e., they did not affect subsequently recorded data. In cases where measurement duration approached or exceeded ten minutes, data recording at the PC-end was delayed, resulting in a loss of data if the measurement was ended too soon after the period of interest. Given that the duration of measurements is intended to be around two minutes for different dynamic walking tasks, this problem is unlikely to be relevant in practice.

3.3. Indication

The developed measurement system accommodates up to eight sensors and has a maximum single-sensor sampling frequency of 1299 Hz. Considering all four tested sensor units, experimental evaluations show the system's functionality to be $err_{\text{rel}} = -0.34 \pm 1.28\%$ in uniaxial and $err_{\text{rel}} = -0.90 \pm 0.98\%$ in diagonal test-rig motion. Measurement errors were found to be uncommon, of known types, and not critical to the measurement.

4. Pilot Study

This section presents the methodology and results of the conducted pilot study.

4.1. Methodology

The pilot study was conducted with a single participant, an active male (K4, 69 years, 1.88 m, 90 kg) with transtibial amputation on the right side (stump length 11 cm, amputated for five years, Pro Flex XC foot (Össur hf, Reykjavik, Iceland)) who is wearing a custom-built measurement socket based on a hydrostatic plaster impression with pin lock mechanism and size 30 silicone Relax™ Locking liner (Össur hf, Reykjavik, Iceland). Prior to the experiment, the participant provided his written informed consent. The study was conducted with a positive vote by the ethics committee of the Technische Universität Darmstadt (reference number: EK40/2017) and is in accordance with the Declaration of Helsinki in its current version.

The developed measurement system was used to record the relative motion at seven specific socket locations. The sampling frequency was set to 100 Hz per sensor unit. Measurement sites were identified via semi-structured interviews with the participant and the prosthetist. Four of the chosen sites have previously been experienced as problematic and can mostly be assigned to anatomical landmarks (e.g., distal end of residual bones). In addition, three unproblematic socket locations were chosen for comparison. Table 1 summarizes the positions of the individual sensor units.

Table 1. Identified measurement sites.

Location	Group	Description	Abbreviation
anterior proximal	comparison	medial tibial flair	ant prox
anterior distal	problem	distal end of tibia bone	ant dist
lateral proximal	comparison	lateral support	lat prox
lateral middle	problem	fibula	lat mid
lateral distal	problem	distal end of fibula bone	lat dist
posterior proximal	comparison	center of posterior compartment	post prox
posterior distal	problem	distal end of tibia bone	post dist

For a seamless sensor unit integration into the socket, the identified sites were already taken into account during the socket manufacturing process: The positive residual limb plaster model with marked locations was digitized with the 3D-scanner ATOS III (GOM GmbH, Braunschweig, Germany), individual sensor base plates corresponding to the plaster model surface at the measurement locations were designed, and 3D-printed in PLA using the Ultimaker 2+ (Ultimaker B.V., Geldermalsen,

Netherlands). These sensor base plates were attached to the plaster model, on which the carbon socket was then fabricated.

In addition to the presented measurement system, a custom-built measurement adapter [18] integrated into the prosthetic structure was used to measure the loads at the distal end of the socket with a sampling frequency of 500 Hz. The two systems were synchronized by a triggering signal provided by the adapter. Figure 4 shows the participant wearing the custom-built measurement socket with integrated sensor array as well as the adapter for measuring the occurring loads.
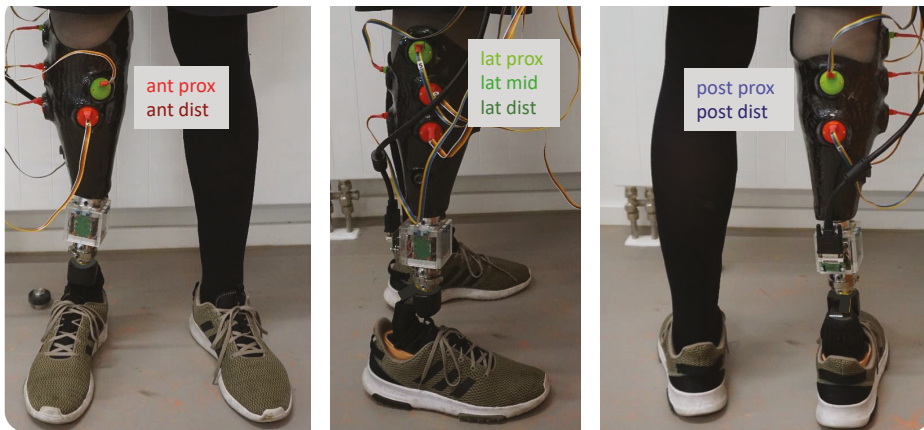

Figure 4. Participant equipped with custom-built measurement socket and load adapter.

For this pilot study, the participant completed a 5 m walkway of straight level walking 15 times. Data of the two measurement systems were interpolated to 200 Hz. The measured force in proximodistal (pd) direction was used to automatically identify and isolate completed gait cycles (GC). Of these gait cycles, the first and last steps of each measurement were excluded. In total, 54 gait cycles remained for data analysis.

4.2. Results and Discussion

For reasons of clarity, Figure 5 shows the mean of all 54 gait cycles of the measurement data. The first two subplots summarize the external loads measured by the adapter, while the other subplots show data captured by the sensor array. Due to the different measurement sites, all sensors captured the relative motion in pd-direction, while anteroposterior (ap) and mediolateral (ml) movement was only detected by sensor units affixed in the Sagittal and Coronal plane, respectively. The bottom subplot shows the SQUAL-values of each sensor unit.

The movement results show almost no motion during stance phase. This changes with the decrease of F_{pd} at around 55% of the gait cycle: With less weight bearing on the prosthesis, the relative motion between residual limb and prosthetic socket increased. Most of the relative motion was detected during swing phase. Movements in ml-direction were of smaller magnitude compared to motion in the Sagittal plane.

Except for three sensor units in the first 5–10% of the gait cycle, SQUAL-values were relatively constant during stance phase. More changes could be observed during swing phase: Three sensor units (lat dist, ant prox, and ant dist) showed a SQUAL-value of 0 for the majority of swing phase. The SQUAL-values of two sensor units (post dist and lat mid) showed a brief break-in during the first and second half of swing phase, respectively.
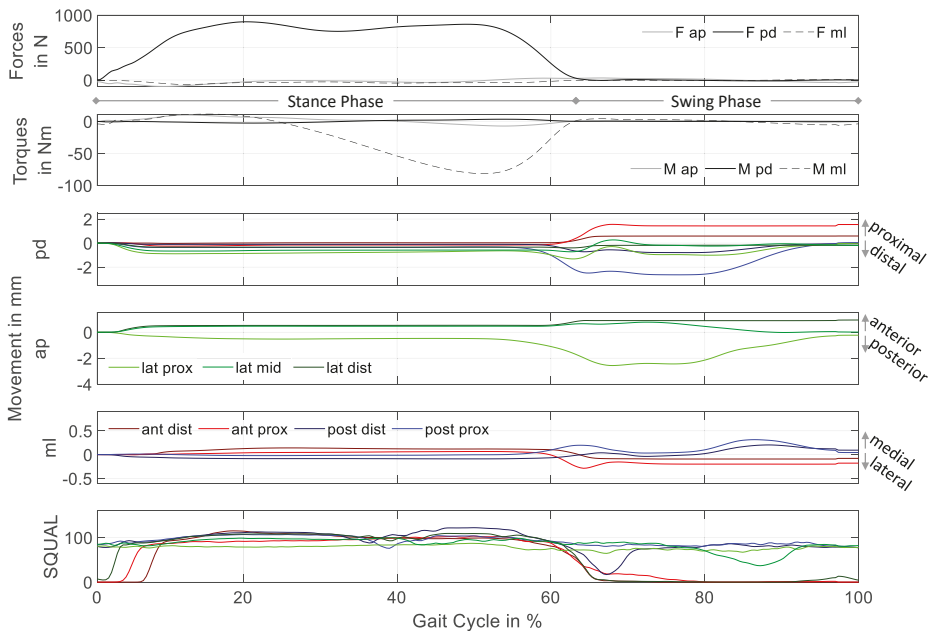
Figure 5. Mean of measurement data of 54 gait cycles; from top to bottom: forces, torques, movement in proximodistal, anteroposterior, mediolateral direction, and SQUAL-values sensor units. The colors used for the sensor units correspond to Figure 4.

As stated in Section 2, the SQUAL-value is equal to 1/4 of the features the sensor uses to calculate its movement over the reference surface. Consequently, a SQUAL-value of zero means that the sensor cannot detect any features for movement calculation. This most likely stems from an out of bounds distance between sensor and reference surface (liner), which corresponds to movement of the residual limb normal to the sensor lens. During these time instants, the sensor does not continue to measure movement in its main axes. Thus, the shown movement data of the affected sensor units were underestimated for lower SQUAL-values (break-in during swing phase) and inconsequential at time instants with SQUAL-values of zero.

The small movement during stance phase might indicate the tight socket fit. With load bearing in the longitudinal direction, the residual limb is pressed radially against the socket wall. The resulting normal force between residual limb surface and socket wall leads to a high friction force, which prevents most of the relative movement during stance phase. This relatively constant position of the limb within the socket during stance phase has been observed previously [9]. Nevertheless, observed relative movements during stance phase show a uniform behavior in pd-direction while movement observations in ap- and ml-direction depend on the measuring site. Overall, the recorded movement of the residual limb inside the socket during stance phase might be described as sinking deeper into the socket, extending in the Sagittal plane and rotating internally in the Transverse plane.

The detected relative motion at the different socket locations during swing phase is mostly non-uniform: While negative motion in pd-direction is noted at measuring site post prox, the movement detected at ant prox is in the opposite direction. This might indicate a rotation of the socket relative to the residual limb in the Sagittal plane, which would correspond to the shank extending inside the socket. Compared to these two measuring sites, there is little movement detected in pd-direction at the other sites. In ap-direction, most relative motion is detected at lat prox in the first half of swing phase. Corresponding to the residual limb moving in posterior direction, this

matches the shank extending inside the socket. The data are in accordance with [4], which reports residual limb extension inside the socket in the first half of swing phase. In addition to having a smaller magnitude compared to the other directions, the observed movement in ml-direction at the two posterior measurement sites is qualitatively the same, with a maximum at about 90% of the gait cycle. Jointly with the movement indication in the opposite direction at ant prox (60–70% of GC), this might correspond to the shank performing an external rotation in the Transverse plane within the socket during swing phase.

Figure 6 summarizes peak-to-peak (P2P) values measured during the loading (0–55% GC) and unloading (55–100% GC) phase of each gait cycle at the different sites as box plots. The first four rows show the movement of the residual limb inside the socket in the Coronal plane, while the last three rows depict movement in the Sagittal plane. Contrary to data shown in Figure 5, only measurement data of each time instant captured at a corresponding SQUAL-value between 50 and 150 were included in the analysis. The third column shows the subsequent number of included gait cycles at each instant of time. Additionally, mean $\pm$ standard deviation of SQUAL-values of included data are shown in the rightmost column. Vertical lines indicate to which box plot (load vs. unload) the data belong.

The magnitude of relative motion between residual limb and prosthetic socket is quite small; no peak-to-peak values above 3.1 mm are recorded. Except for relative motion at lat dist, more relative motion is recorded during the unloading phase (last 45% of GC) compared to the loading phase (first 55% of GC). A tendency of peak-to-peak values being of greater magnitude during unloading phase compared to loading phase as well as for proximal measurement sites compared to those located more distally was observed.

Taking into account the distal fixation of the liner within the socket (pin lock system) and assuming a constant strain $\epsilon = \frac{\Delta l}{l_0}$ along the pd axis of the liner led to higher peak-to-peak relative motion at proximal compared to distal measuring sites ($l_{0,\text{prox}} > l_{0,\text{dist}}$).

At the two anterior measurement sites as well as at lat dist, the collected movement data are of improvable quality during swing phase; the number of included gait cycles at the three sites was reduced from over 50 at 60% GC to less than 15 at 65% of GC. Time instants of the gait cycle without mean SQUAL-values indicate where the affected sensor units did not detect the necessary features for movement data calculation. At post dist and lat mid, a dip was observed in the number of included gait cycles at 65% and 95% of GC, respectively.

Peak-to-Peak movement data recorded during swing phase by the sensor units on the anterior side of the socket and at lat dist need to be interpreted with caution. The shown values only consider data of parts of the loading and unloading phase (compare SQUAL in rightmost column), which might not include maximal movements. Excluded data at 95% of GC would most likely not affect P2P calculation.

As stated previously, variations in SQUAL-value might stem from too much movement normal to the measurement plane of the corresponding sensor unit. Quantification of the dependency of SQUAL-values on measurement distance between sensor unit and reference surface as well as influence on displacement data might lead to the extension of usable sensor data information.

Compared to most studies quantifying residual limb socket peak-to-peak displacements of transtibial amputees using a socket with pin-lock suspension during walking [3,4,9,12], the detected movement within the pilot study was relatively small. A study investigating the variation of sock thickness on peak-to-peak relative motion shows a distinct decrease as sock thickness increased and vice versa [12]. These findings support the hypothesis of a tight socket fit being responsible for the small magnitudes of detected movement (particularly during stance). Another study presents P2P displacements normal to the socket wall ranging between 0.7 and 9.4 mm and varying for participants as well as measuring sites [10]. The data (see [10] ([Figure 7])) show a tendency of movements of greater magnitude at distal measuring sites compared to measuring sites placed most proximally. Additionally, a comparison of P2P displacements at the proximal measuring sites show consistently greater magnitudes on the anterior side. This supports the hypothesis of residual limb movement

normal to the socket wall being responsible for SQUAL-values of 0 at the measuring sites ant prox and ant dist (as well as lat dist).
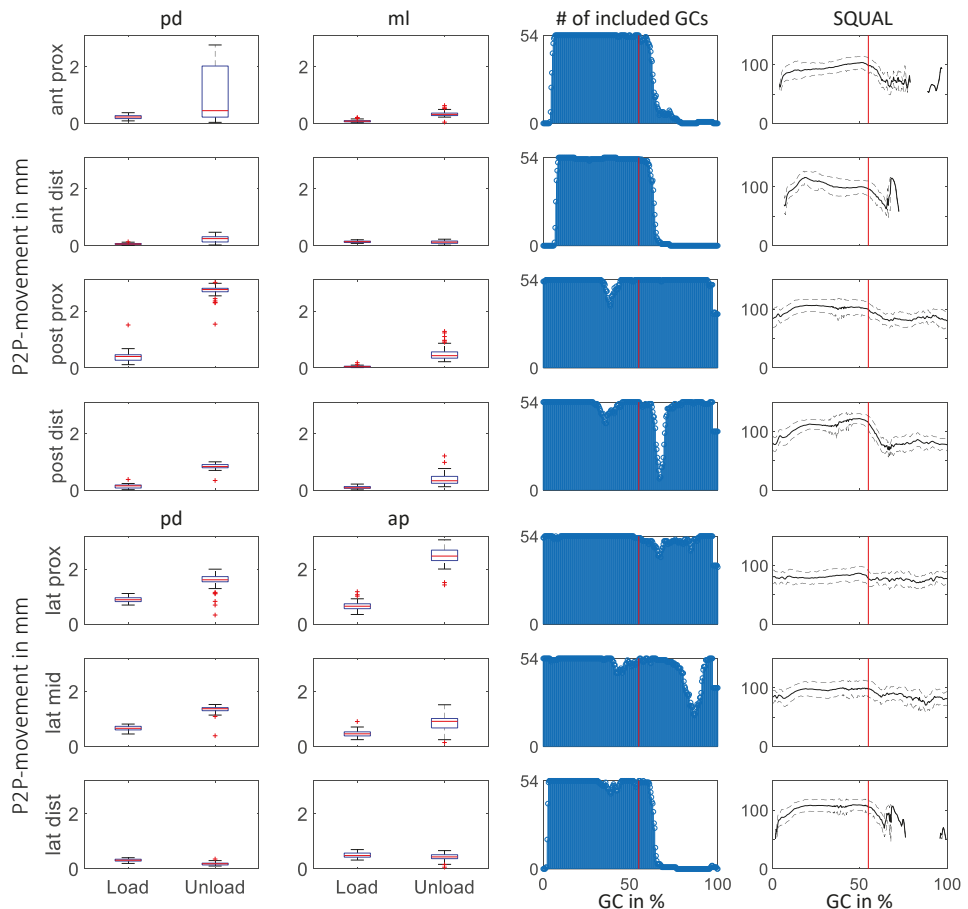

Figure 6. Analysis of peak-to-peak values at the different measuring sites. The data of sites are arranged according to location on anatomical planes: Coronal plane (first four rows) and Sagittal plane (following three rows). The first two columns summarize peak-to-peak values of relative motion in 54 gait cycles in the two anatomical directions corresponding sensor location (pd and ml for Coronal plane, and pd and ap for Sagittal plane). The two columns on the right indicate the data basis underlying peak-to-peak calculations.

5. Conclusions

Overall, the presented data of the pilot study are biomechanically plausible and in accordance with prior findings of studies analyzing residual limb movement within the sockets of individual with transtibial amputation during gait [4,9,10]. Compared to other measuring approaches [4,11,13], the presented measuring system has an additional value as it enables the analysis of relative motion at a number of specific measuring sites, thus facilitating the detection of compliant soft tissue behavior.

Being tethered constrains the area of use of the system; there is potential for removing this impediment by switching to either wireless data transmission or on-board data storage. Nevertheless, even in its current state, the presented measuring system facilitates the analysis of relative movement between residual limb and prosthetic socket at specific measurement sites for different dynamic

gait situations. Future studies will include the exploration and analysis of different gait situations, socket manipulations, as well as prosthetic components and their impact on relative motion. Additionally, the distribution of relative motion with respect to problematic and unproblematic areas will be investigated.

Author Contributions: Conceptualization, V.N. and S.W.; Data curation, V.N. and S.W.; Formal analysis, V.N. and S.W.; Investigation, V.N. and S.W.; Methodology, V.N.; Resources, P.B. and S.R.; Software, V.N. and S.W.; Supervision, P.B. and S.R.; Validation, V.N. and S.W.; Visualization, V.N.; Writing—original draft, V.N. and S.W.; and Writing—review and editing, V.N., S.W., P.B., and S.R.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SPI	Serial Peripheral Interface
PLA	Polylactide
USB	Universal Serial Bus
cpi	counts per inch
pcb	printed circuit board
GC	gait cycle
pd	proximodistal
ap	anteroposterior
ml	mediolateral
P2P	Peak-to-Peak
ant prox	anterior proximal measurement site
ant dist	anterior distal measurement site
lat dist	lateral distal measurement site
lat mid	lateral middle measurement site
lat prox	lateral proximal measurement site
post dist	posterior distal measurement site
post prox	posterior proximal measurement site

References

1. Paterno, L.; Ibrahimi, M.; Gruppioni, E.; Menciassi, A.; Ricotti, L. Sockets for limb prostheses: A review of existing technologies and open challenges. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1996–2010, doi:10.1109/TBME.2017.2775100. [CrossRef] [PubMed]
2. Eshraghi, A.; Osman, N.A.A.; Gholizadeh, H.; Karimi, M.; Ali, S. Pistoning assessment in lower limb prosthetic sockets. *Prosthet. Orthot. Int.* **2012**, *36*, 15–24. [CrossRef] [PubMed]
3. Noll, V.; Wojtusch, J.; Schuy, J.; Grimmer, M.; Beckerle, P. Measurement of biomechanical interactions at the stump-socket interface in lower limb prostheses. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015.
4. Childers, W.L.; Siebert, S. Marker-based method to measure movement between the residual limb and a transtibial prosthetic socket. *Prosthetics Orthot. Int.* **2016**, *40*, 720–728, doi:10.1177/0309364615610660. [CrossRef] [PubMed]

5. Eshraghi, A.; Abu Osman, N.A.; Karimi, M.T.; Gholizadeh, H.; Ali, S.; Abas, W.W.; Azam, M.N. Quantitative and qualitative comparison of a new prosthetic suspension system with two existing suspension systems for lower limb amputees. *Am. J. Phys. Med. Rehabil.* **2012**, *91*, 1028–1038, doi:10.1097/PHM.0b013e318269d82a. [CrossRef] [PubMed]

6. Gholizadeh, H.; Osman, N.A.A.; Kamyab, M.; Eshraghi, A.; Abas, W.W.; Azam, M.N. Transtibial prosthetic socket pistoning: static evaluation of Seal-In® X5 and Dermo® Liner using motion analysis system. *Clin. Biomech. Bristol. Avon.* **2012**, *27*, 34–39, doi:10.1016/j.clinbiomech.2011.07.004. [CrossRef] [PubMed]

7. Appoldt, F.; Bennett, L.; Contini, R. The results of slip measurements in above-knee suction sockets. *Bull. Prosthetics Res.* **1968**, *3*, 106–112.

8. Gholizadeh, H.; Osman, N.A.A.; Lúvíksdóttir, Á.G.; Eshraghi, A.; Kamyab, M.; Abas, W.W.; Azam, M.N. A new approach for the pistoning measurement in transtibial prosthesis. *Prosthetics Orthot. Int.* **2011**, *35*, 360–364, doi:10.1177/0309364611423130. [CrossRef] [PubMed]

9. Sanders, J.E.; Karchin, A.; Fergason, J.R.; Sorenson, E.A. A noncontact sensor for measurement of distal residual-limb position during walking. *J. Rehabil. Res. Dev.* **2006**, *43*, 509–516. [CrossRef] [PubMed]

10. Henrikson, K.M.; Weathersby, E.J.; Larsen, B.G.; Cagle, J.C.; McLean, J.B.; Sanders, J.E. An Inductive Sensing System to Measure In-Socket Residual Limb Displacements for People Using Lower-Limb Prostheses. *Sens. Basel Switz.* **2018**, *18*, 3840. [CrossRef] [PubMed]

11. Gerschutz, M.J.; Hayne, M.L.; Colvin, J.M.; Denune, J.A. Dynamic Effectiveness Evaluation of Elevated Vacuum Suspension. *J. Prosthetics Orthot.* **2015**, *27*, 161–165. [CrossRef]

12. Swanson, E.C.; McLean, J.B.; Allyn, K.J.; Redd, C.B.; Sanders, J.E. Instrumented socket inserts for sensing interaction at the limb-socket interface. *Med. Eng. Phys.* **2018**, *51*, 111–118, doi:10.1016/j.medengphy.2017.11.006. [CrossRef] [PubMed]

13. Wernke, M.M.; Schroeder, R.M.; Haynes, M.L.; Nolt, L.L.; Albury, A.W.; Colvin, J.M. Progress Toward Optimizing Prosthetic Socket Fit and Suspension Using Elevated Vacuum to Promote Residual Limb Health. *Adv. Wound Care* **2017**, *6*, 233–239, doi:10.1089/wound.2016.0719. [CrossRef] [PubMed]

14. Vempala, V.; Liu, M.; Kamper, D.; Huang, H. A Practical Approach for Evaluation of Socket Pistoning for Lower Limb Amputees. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, HI, USA, 17–21 July 2018; pp. 3938–3941.

15. Noll, V.; Weber, P.; Scortecci, S.; Beckerle, P.; Rinderknecht, S. A sensor to acquire the relative movement between residual limb and prosthetic socket. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 6–12 October 2016.

16. Noll, V.; Rinderknecht, S.; Beckerle, P. Systematic Experimental Assessment of a 2D-Motion Sensor to Detect Relative Movement between Residual Limb and Prosthetic Socket. *Sensors* **2018**, *18*, 2170. [CrossRef]

17. Whitmore, S. Realization of a Measuring Device for Recording the Relative Movement between Residual Limb and Prosthetic Socket. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2018.

18. Schuy, J.; Rinderknecht, S. Integrated measurement system for amputee gait analysis: A pilot study. In Proceedings of the 2014 IEEE Healthcare Innovation Conference (HIC), Seattle, WA, USA, 8–10 October 2014; pp. 91–94.

Article

A Dynamic Time Warping Based Algorithm to Evaluate Kinect-Enabled Home-Based Physical Rehabilitation Exercises for Older People

Xiaoqun Yu and Shuping Xiong *

Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea
* Correspondence: shupingx@kaist.ac.kr; Tel.: +82-42-350-3132

Abstract: Older people face difficulty engaging in conventional rehabilitation exercises for improving physical functions over a long time period due to the passive nature of the conventional exercise, inconvenience, and cost. This study aims to develop and validate a dynamic time warping (DTW) based algorithm for assessing Kinect-enabled home-based physical rehabilitation exercises, in order to support auto-coaching in a virtual gaming environment. A DTW-based algorithm was first applied to compute motion similarity between two time series from an individual user and a virtual coach. We chose eight bone vectors of the human skeleton and body orientation as the input features and proposed a simple but innovative method to further convert the DTW distance to a meaningful performance score in terms of the percentage (0–100%), without training data and experience of experts. The effectiveness of the proposed algorithm was validated through a follow-up experiment with 21 subjects when playing a Tai Chi exergame. Results showed that the algorithm scores had a strong positive linear relationship (r = 0.86) with experts' ratings and the calibrated algorithm scores were comparable to the gold standard. These findings suggested that the DTW-based algorithm could be effectively used for automatic performance evaluation of an individual when performing home-based rehabilitation exercises.

Keywords: aging; physical function; rehabilitation exercise; Kinect; dynamic time warping; automatic coaching; exergame

---

1. Introduction

The global healthcare system is under great pressure due to rapid population aging as well as a shortage of healthcare personnel and budget [1]. An increasing proportion of older people is facing serious challenges of impaired physical functions such as muscle strength, balance, and mobility [2]. All these negative changes result in difficulties for older people maintaining independence of daily living, which would further cause anxiety, low self-esteem, and decreased quality of life [3,4]. Epidemiological studies show that a low physical activity level is strongly correlated to functional decline of the elderly. Physical exercise is an effective way to counteract the age-related functional decline [5].

There is strong evidence that appropriate physical rehabilitation exercises can improve physical activity level and activities of daily living of older people [4,6,7]. Conventional exercise therapies for older people are generally conducted in a formal rehabilitation center or clinical setting, which requires direct supervision from a professional therapist. Even though conventional exercise therapies have been shown as effective to increase physical activities as well as improve motor functions and balance, they suffer from low rates of uptake and adherence [8,9] due to a lack of enjoyment, inconvenient transportation, and high cost [10]. For example, Kobayashi et al. [11] examined the effects of physical exercise on fall risks in older people living at home in a rural area. After a 13-week intervention, they

reported that the motor functions of the elderly were improved. However, only 31.7% of the participants were fully adherent to the intervention program. Another study by Liu-Ambrose et al. [12] showed that the Otago Exercise Program (OEP) improved functional mobility and executive functioning in older people based on a 26-week experiment. However, they reported that only 19.4% of the participants finished the whole sessions of OEP and only 25.0% of participants just completed the half proportion of sessions in OEP.

Recently, with the emergence of more affordable motion sensors (such as Kinect and inertial sensors) and gaming technologies, the use of home-based exergame (exercise + gaming) for physical rehabilitation appears promising over passive conventional exercise therapies for the elderly with respect to long-term uptake [9,13–15]. Such home-based solutions can not only eliminate restrictions of distance and cost to the rehabilitation center, but also allow the elderly to flexibly adjust the training schedule and exercise intensity. Another big advantage of exergame is that it provides auto-coaching to compensate for the lack of global healthcare resources [16]. In this kind of auto-coaching system, there is usually a virtual avatar coach in the game environment. The coaching system will guide end users to follow virtual coach's motions as well as track and evaluate their reproduced motions. However, because there is no real coach on site during the exercise, the scientific motion comparison for automatic performance evaluation is critical to guarantee the effectiveness of such a system.

Different algorithms have been proposed to assess exercise performance automatically. Lin et al. [17] developed a Kinect-based rehabilitation system utilizing a "seated Tai Chi" exercise to assist patients with movement disorders. They decomposed each form of Tai Chi into four poses and calculated the difference of joint angles between skeletons of standard pose and user's actual pose, which were tracked by Kinect sensors. Muangmoon et al. [18] adopted a similar method to evaluate Thai dance performance. Even though this kind of motion comparison algorithm is easy to implement due to the simplification of extracting a few discrete poses from continuous motions, the quality of exercise could not be evaluated comprehensively. This is because the simplification has resulted in the loss of some essential exercise information such as motion continuity, pace, and intensity. Another group of researchers proposed to assess motion similarity based on the correlation coefficient between two time series of human motions from a virtual coach (standard motion) and a subject (actual motion) [19,20]. Even though this approach can evaluate the overall performance from the continuous motion data, it fails to take into account the typical time lag and speed variance between the older subject's actual motion and the virtual coach's standard motion during the exercise therapy. It is difficult for the older people to exactly follow the pace of standard motion due to the physical and cognitive impairments as well as the lack of exercise skills. Both previously mentioned disadvantages can be elegantly overcome by dynamic time warping.

Dynamic time warping (DTW) is a well-known approach for measuring time-series similarity, which minimizes the effects of time lag and distortion in the time axis due to speed variation [21,22]. DTW outputs the optimal alignment (least matching cost or cumulative distance) between two time series and it is widely applied in speech and gesture recognition [23]. Due to the advantages of DTW, some researchers applied it into the auto-coaching system for physical rehabilitation exercises at home. Saraee et al. [24] applied DTW into developing a remote monitoring system to evaluate home-based physical exercises. Since DTW itself could not generate a meaningful scaled score for performance evaluation, a physical therapist was required to remotely monitor the patient in real time through Webcam and determine whether a patient's performance was acceptable or not. Semblantes et al. [25] and Saenz-de-Urturi and Garcia-Zapirain Soto [26] used DTW and binary classification to discriminate between correct and incorrect motions. Su et al. [27] utilized DTW and Kinect sensors to evaluate patients' supplementary exercise at home for shoulder rehabilitation, when compared with the pre-recorded standard motion in the hospital. They integrated DTW with fuzzy logic to convert different DTW matching costs (DTW distances) to three performance levels: bad, good, and excellent. Wei et al. [28] applied DTW to measure the similarity between the motion data of the trainee and the standard motion of the trainer in dance teaching. They utilized existing training data

and the experience of experts to determine three boundaries of DTW matching costs for categorizing individual performances into four levels: below average, average, good, and excellent. Most studies require large representative training data of each exercise to convert the matching cost from DTW to the final performance score. This task is resource-intensive and it is difficult to generalize the established conversion criteria for a specific exercise program to different ones. In addition, the final performance evaluation is categorical, and, thus, qualitative and not sensitive to recognize user's gradual progress of exercise interventions. Chatzitofis et al. [29] and Mocanu et al. [30] developed home-based rehabilitation systems for heart health and physical activity. DTW was used to compute the quantitative performance score. However, how to convert the DTW matching cost to a quantitative score was not described and the quantitative scores were not validated with ground-truth ratings. Osgouei et al. [31] recently proposed an objective method for quantitative performance evaluation of rehabilitation exergames. Using shoulder abduction exercise as an example and two angle features (shoulder angle and arm angle), they presented how DTW was applied to compute the motion similarity between the unknown and reference trajectories of the human skeleton joints. A normalization approach with estimated lower and upper bounds for the DTW distance was utilized to further convert any DTW distance to an objective similarity score (0–100). The proposed method is promising since it does not require any training data. However, the proposed objective similarity score was not validated with physicians' evaluation of the performances. In addition, it remains questionable whether the proposed method can be extended from simple repetitive exercises to complex whole-body exercises.

This study aims to develop and validate a DTW-based algorithm for a motion similarity evaluation, in order to support effective Kinect-enabled home-based virtual coaching. We proposed a simple but innovative method to directly convert the DTW matching cost to a meaningful performance score in terms of percentage (0–100%), without training data and experience of experts. We further validated the effectiveness of our algorithm through a follow-up experiment with human subjects performing the complex whole-body exercise (Tai Chi) instead of simple, repetitive exercises, which would show good generalization of our proposed method to different exercise programs. The developed algorithm is expected to provide a similar evaluation on user's performance as domain experts, which could be very promising to apply into home-based physical rehabilitation exercises for better quality of life of the elderly.

2. Methods

2.1. Development of a DTW-Based Algorithm for Performance Evaluation

2.1.1. 3D Pose Comparison

Human motion is the coordinated movement of different body parts, and motion data from the Kinect sensor could be seen as a sequence of frames that comprise 3D coordinates of joint positions of the human skeleton and each bone determined by two connected joints could be seen as a 3D vector in the space. In order to assess the similarity between trainer's motion (coach) and trainee's motion (end user), the first step is to quantify the difference between these two 3D poses at a given frame. In this study, we chose the sum of the angle difference between all corresponding bone vectors of the two 3D poses as the distance measure. Eight major bones of the human skeleton (the upper and lower arms, and the upper and lower legs) were chosen for motion comparison since most human motions during the rehabilitation exercises involve the coordination of upper and lower limb movements. The angle difference (θ) between two corresponding bone vectors of trainer and trainee is illustrated in Figure 1, which can be calculated by the law of cosine (see Equation (1)).

$$\theta = \text{acos}(\frac{\vec{AB} \cdot \vec{A'B'}}{|\vec{AB}| \cdot |\vec{A'B'}|}) \tag{1}$$
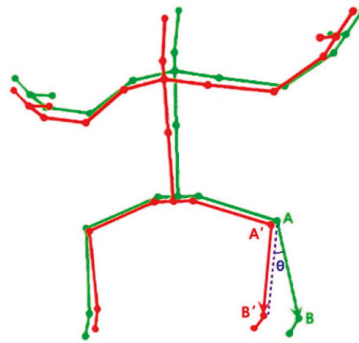
Figure 1. Angle difference (θ) between corresponding bone vectors for trainer (green) and trainee (red), using the left lower leg as an example.

2.1.2. Motion Comparison

Given two sequences of motion data X and Y, each frame of motion data is a 3D pose. We applied DTW to find the optimal matching between the trainer's motion and the trainee's motion while minimizing the effects of shifting and distortion in time [32]. Since we chose eight bone vectors for a motion comparison and the motion of each bone constitutes one dimensional time-series data, both trainer and trainee's data have eight dimensions. The matrix of each dataset has dimension 8-by-*n*, where *n* is the total frames of motion data. To explain how DTW works, let us start with the one-dimensional case. The motion of a body part can be denoted as $S = (s_1, s_2, \ldots, s_n)$ and $T = (t_1, t_2, \ldots, t_m)$, which correspond to the trainer's motion and trainee's motion, respectively. The element in *S* and *T* is a normalized bone vector of that body part in a certain 3D coordinate system. To compare the similarity of sequence *S* and *T* by DTW, an *n*-by-*m* cost matrix is constructed where the (*i*th, *j*th) element denoted as $C(s_i, t_j)$ is the angle difference between s_i and t_j (See Equation (1)). A warping path denoted as *P* defines an alignment between *S* and *T* in the cost matrix, which should satisfy three conditions: boundary condition, monotonicity condition, and step size condition [33]. There could be multiple feasible warping paths in the cost matrix and the total matching cost of one warping path *P* between *S* and *T* is defined by the equation below.

$$c_p(S,\ T)\ =\ \sum_{k\ =\ 1}^{s} C\Big(s_{ik},\ t_{jk}\Big), \tag{2}$$

where *s* is the length of the warping path *P*.

The goal of DTW is to find the optimal warping path, which has the minimal cumulative distance among all the possible warping paths. The DTW distance *DTW(S, T)* is defined as the total matching cost of the optimal warping path. In order to find the optimal warping path, a dynamic programming method is used. The recursive equation is given by the equation below.

$$D(i,j)\ =\ min\{D(i-1,\ j-1),\ D(i-1,\ j), D(i,\ j-1)\} + C\Big(s_i, t_j\Big), \tag{3}$$

where $1 < i < n$ and $1 < j < m$. $D(i, j)$ represents the matching cost between standard data (*S*) and testing data (*T*) from (1, 1) to (*i, j*).

DTW can be generalized from the one-dimensional case to a multi-dimensional case [34]. For a multidimensional case, s_i and t_j are not single bone vector but multiple bone vectors, which represent whole body motion. Multi-dimensional *DTW(S, T)* is calculated in a similar way as the one-dimensional case, except that we need to redefine $C(s_i, t_j)$ as the sum of angle difference among all the dimensions. The output of DTW is the matching cost associated with the cumulative distance along the shortest warping path. Therefore, the lower the matching cost is, the closer the two motion sequences are and the better the motion performance is. In order to quantitatively assess motion performance of a trainee,

we further convert the DTW distance (matching cost) to a meaningful performance score in terms of the percentage (0–100%) using the following equation.

$$Performance\ score\ =\ \frac{\sum_{k=1}^{s}\left[1-\frac{C(s_{ik},\,t_{jk})}{90\times 8}\right]}{s}\ =\ 1-\frac{DTW(S,\,T)}{90\times 8\times s}, \tag{4}$$

where s is the length of optimal warping path, 8 stands for eight bone vectors selected for motion evaluation, and $C(s_{ik}, t_{jk})$ is the element of optimal warping path in the cost matrix, which is the summation of angle differences for eight bone vectors, and DTW distance-$DTW(S, T)$ is a summation of elements $(C(s_{ik}, t_{jk}))$ along the optimal path. We assume the angle difference between two corresponding bone vectors is within 90 degrees based on an earlier study [18], which results in the maximum $DTW(S, T)$ along the optimal path, which would be $90 \times 8 \times s$. Because the output distance ($DTW(S, T)$) is a measure of dissimilarity between the two motion time series (the longer the distance, the greater the deviation), the last part of Equation (4) would be a percentage score (0–100%) to measure the level of similarity between the trainee's motion and trainer's motion.

2.1.3. Body Orientation Offset

Calculation of bone vectors for both trainee's motion and trainer's motion based on the world coordinate system could be problematic if the trainee is not oriented as exactly as the trainer. The difference (error) in body orientation would be directly transferred to all eight bone vectors. In order to solve this problem, instead of using the joint position based on the world coordinate system, we calculate the joint position data based on a local coordinate system of the human model, which would be updated in real time. Establishment of the local coordinate system was adopted from Unity3D Mecanim system [35]. The up vector is defined as middle of left/right upper arm and middle of left/right upper leg. The left vector is an average upper body left (a vector defined by the left upper arm and the right upper arm) and lower body left (a vector defined by the left upper leg and the right upper leg). The forward vector is the cross product of the up vector and the left vector. In order to make these three vectors orthogonal to each other, the final left vector is the cross product of the forward vector and the up vector. Then the up vector is aligned to a normal vector of a ground plane as the y-axis. Hence, the final left vector and forward vector would be finalized as the x-axis and z-axis, respectively, based on the rotation matrix. The main purpose of this step is to guarantee that the x-axis and the z-axis would always be in the ground plane. The origin of this local coordinate system is the projection of center of mass on the ground.

As shown in Figure 2, the only difference between motions of two avatars is the body orientation (see z-axis). Under the local coordinate system, all the bone vectors between two avatars are exactly the same. While under the world coordinate system, there are clear angle differences among all corresponding bone vectors of two avatars, which are caused by the different body orientations.

Calculation of angular differences based on the local coordinate system of the human model can remove the accumulative errors on eight bone vectors caused by different body orientations. However, the difference in body orientation between trainee and trainer should also be considered in this case. Hence, we added one more dimension-body orientation into the evaluation of motion similarity. Lastly, we used nine dimensions to calculate the motion similarity: eight bone vectors under the local coordinate system of the human model and the forward vector of the body, which reflects the body (mainly trunk) orientation (z-axis in Figure 2). Then the final performance score can be updated by Equation (5).

$$Final\ performance\ score\ =\ \frac{\sum_{k=1}^{s}\left[1-\frac{C(s_{ik},\,t_{jk})}{90\times 9}\right]}{s}\ =\ 1-\frac{DTW(S,\,T)}{90\times 9\times s}, \tag{5}$$

where all the notations are the same as in Equation (4) and where 9 stands for nine dimensions.
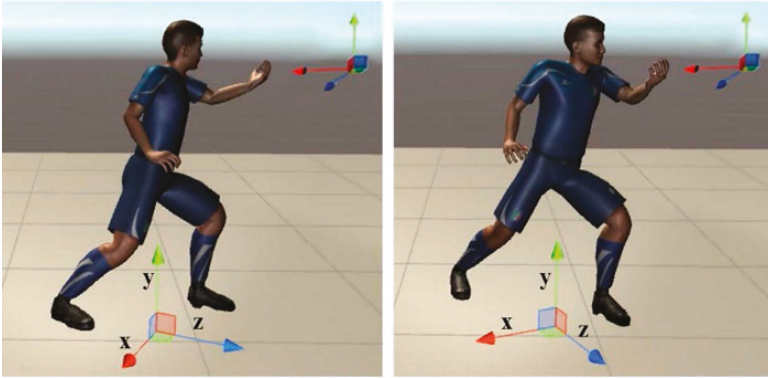

Figure 2. Illustration of motion similarity evaluation under the local coordinate system of the human model (bottom) and the world coordinate system (at the top right corner).

2.2. Validation of the Developed DTW-Based Algorithm

In order to validate the developed algorithm, we performed a follow-up experiment where the final performance scores from the algorithm were compared with ratings given by the domain experts. An 8-form Tai Chi exercise was selected to evaluate the proposed algorithm.

2.2.1. Experimental Participants

Twenty-one middle-aged and older subjects (age: 55.2 ± 4.2 years, height: 166.1 ± 7.9 cm, weight: 65.36 ± 8.3 kg) from a local Tai Chi academy participated in the algorithm validation experiment. For the sake of convenience, the experiment was conducted in the same Tai Chi academy instead of each subject's home. All the subjects were in healthy conditions and without musculoskeletal diseases or injuries that may affect their Tai Chi performance. Prior to the participation of the experiment, each subject signed informed consent on the experiment protocol, which was approved by the KAIST Institutional Review Board (IRB-18-070).

2.2.2. Experimental Setup and Procedure

The Tai Chi exergame was developed with Kinect V2 sensor in Unity3D platform (Unity 5.5.2f1) by C# for the real-world application. The main scene of the exergame is implemented with two avatars: virtual trainer and trainee (Figure 3). The avatar-based rendering of motion preserves the privacy of the user, which is critical for the healthcare systems. The motion of the trainee avatar is updated by the real motion of the user. User's motion is captured by Kinect V2 (Microsoft Corp, Redmond, WA, USA) and mapped to the trainee avatar using the Kinect V2 asset for Unity3D [36]. The motion of the trainer avatar is retargeted by the pre-recorded standard motion from a certified Tai Chi instructor (Master Level).

The whole experimental setting is illustrated in Figure 4. Both Kinect V2 and Xsens motion capture system were used to capture each subject's motion. The Kinect sensor was placed at a height of 0.8 m. Since subjects should stretch their arms frequently because of the characteristics of Tai Chi motion, subjects were instructed to stand around 3.0–3.5 m away from the Kinect sensor. Considering that some Tai Chi motion with body rotation may not be well captured by the Kinect due to the self-occlusion [37], a wearable inertial sensor-based motion capture system-Xsens MVN BIOMECH (Xsens Technologies B.V., Enschede, The Netherlands) was also utilized to obtain high-quality motion data for the algorithm validation [38,39].

Figure 3. Main scene of the Tai Chi exergame (left: trainer avatar, right: trainee avatar).
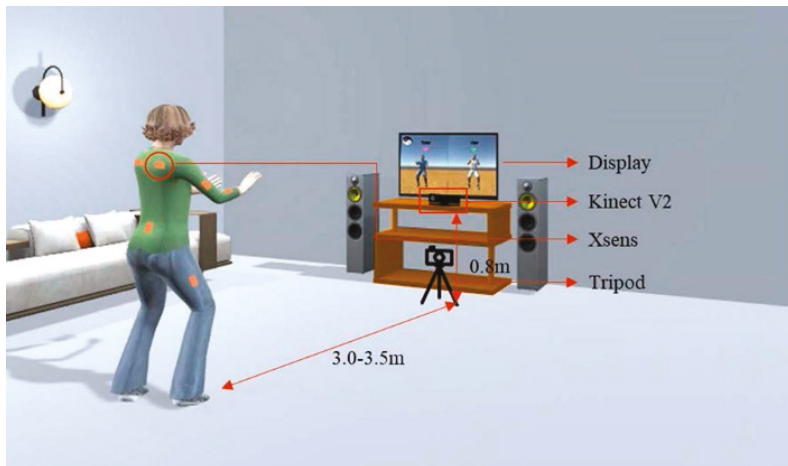
Figure 4. Experimental setup for Tai Chi motion evaluation.

A smart phone, supported by a tripod, was used to record each subject's motion when he/she was playing the Tai Chi exergame. The motion videos were distributed to three Tai Chi experts for independent performance evaluation. Each expert was provided a 10-cm visual analog scale (VAS) [40] and asked to place a vertical mark on the scale to indicate the performance level of motion for each subject. The anchor statements for VAS in this study are "cannot follow Tai Chi at all" (score of 0) on the left and "master level with standard Tai Chi motion" (score of 100) on the right. The raw scale score is then converted to a 0–100 scale.

2.2.3. Statistical Analysis

The intraclass correlation coefficient (ICC) was used to check the inter-rater reliability of experts' subjective ratings [41]. Good consistency and agreement among different experts are the prerequisite to consider experts' rating as the gold standard for validating the developed DTW-based algorithm. ICC is a widely used reliability index and the general guideline of ICC is as follows: ICC < 0.5, poor

reliability, 0.5 < ICC < 0.75, moderate reliability, 0.75 < ICC < 0.9, good reliability, and ICC > 0.9, excellent reliability [42].

More importantly, final performance scores from the developed algorithm were compared with the experts' ratings (as a gold standard). The Pearson correlation coefficient (r) between final performance scores from the developed algorithm and those from experts was calculated to assess the strength of a linear relationship between those two evaluation methods. In addition, linear regression was used to calibrate performance scores from the algorithm so that the scores from two evaluation methods could be consistent. Differences between algorithm scores after calibration and experts' ratings were analyzed. The SPSS statistical package version 20 (IBM Corp., Armonk, NY, USA) was used for statistical analysis.

3. Results

3.1. Inter-Rater Reliability of Experts' Ratings

Figure 5 shows the subjective ratings of 21 subjects from three independent experts. The ICC value for three experts was 0.861 (95% CI: 0.688~0.942), which was within the range of 0.75 to 0.9 [41,42], which indicates the experts' ratings were consistent and with good inter-rater reliability.
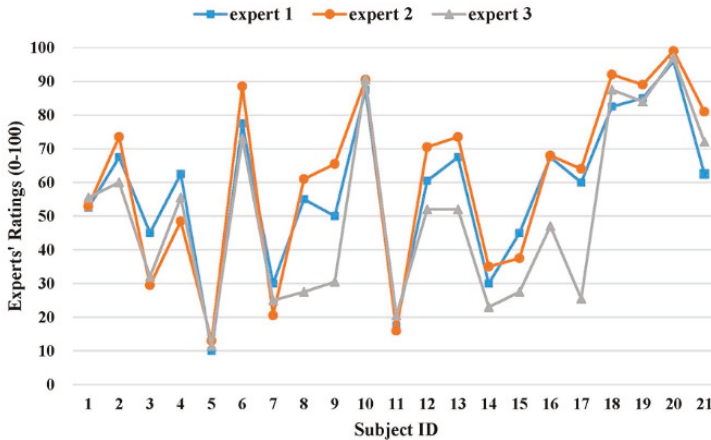
Figure 5. Experts' ratings on motion performance of 21 subjects.

3.2. Evaluation Comparison between Experts and the Developed Algorithm

The subjective ratings from three experts were averaged and regarded as the gold standard to validate the developed DTW-based algorithm. Figure 6 shows the scatter plot of the algorithm's final performance scores and averaged experts' ratings for 21 subjects. The Pearson correlation coefficient (r) was 0.86 ($t = 7.45$, $p < 0.001$), which indicates a strong positive linear relationship between scores from those two evaluation methods. Additional analysis on performance scores between two evaluation methods showed that, under many circumstances, the algorithm would overestimate the performance when compared with the experts' rating. Therefore, we calibrated performance scores of the algorithm using the fitted equation from linear regression (see Figure 6) so that the algorithm would generate similar evaluation scores as the domain experts for practical applications. Figure 7 further demonstrates the calibrated performance scores from the algorithm were comparable to experts' ratings. The score difference between two evaluation methods had a mean of 9.5 and a standard deviation of 7.0 (Maximum: 21.6, Minimum: 0.1).
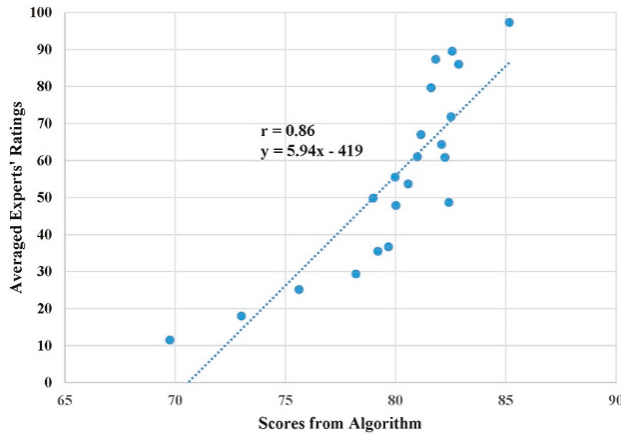
Figure 6. Linear relationship between algorithm scores and experts' ratings.
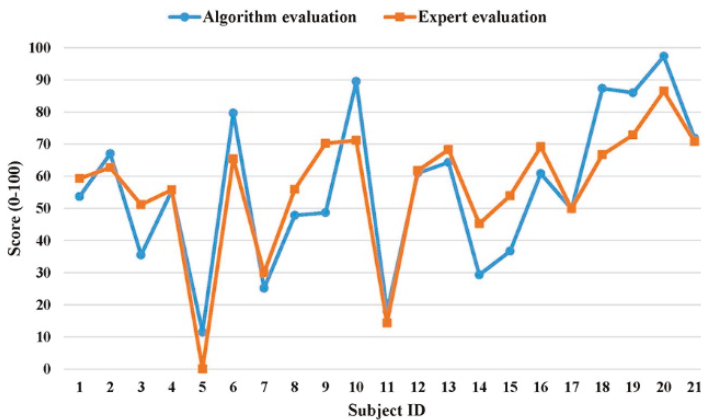
Figure 7. Comparison between algorithm evaluation (after calibration) and experts' evaluation.

4. Discussion

We developed a DTW-based algorithm for assessing motion similarity between an individual user and a virtual coach. DTW was designed to handle local changes in timing (due to speed variations) and, therefore, desirable for evaluating rehabilitation exercises for the elderly self-care at home. The effectiveness of the algorithm was validated through a follow-up experiment. In the validation experiment, the Tai Chi exercise was chosen as the representative physical exercise to verify the proposed algorithm due to two major reasons. First, the effectiveness of Tai Chi exercise for improving physical functions has been proven by many previous studies [43,44]. Second, Tai Chi exercise is a complex and whole-body motion. If the algorithm could perform well in terms of evaluating Tai Chi motion, it should be generalizable to other simpler rehabilitation exercises.

Inter-rater reliability analysis revealed that the reliability level of experts' ratings was "good" $(0.75 < \text{ICC} = 0.861 < 0.90)$. However, the 95% confidence interval of ICC was wide (0.688–0.942), which indicates that, in the worst case, the reliability level was just "acceptable" $(0.5 < \text{ICC} = 0.688 < 0.75)$. The wide confidence interval warned that, even though the overall agreement were high among three experts, there were non-negligible disagreement on their ratings [45]. Paired t-tests also confirmed the significant difference on performance ratings between the third expert and the other two experts

(Figure 5). The inconsistency on subjective ratings from three experts highlighted the potential benefits of applying our developed algorithm to assess the exercise performance automatically and objectively.

Strong linear relationship (r = 0.86) between the algorithm score and experts' evaluation (gold standard) implied the developed algorithm was sensitive in terms of recognizing the performance levels from different subjects as the domain experts. Unexpectedly, a detailed analysis revealed that the algorithm score was significantly higher than the experts' rating. This could be mainly due to different baselines for two evaluation methods. The algorithm evaluation was purely based on the sum of angle differences among nine corresponding body vectors and the subjects with all angle differences at 90 degrees were considered as the worst (performance score = 0). Since even the subjects rated by the experts as the worst in terms of motion performance had most of the angle differences within 45 degrees, the algorithm evaluation would overestimate the subject's performance score due to the ceiling effect [46]. To reduce this overestimation and enable our algorithm to provide similar scores as the domain experts, the linear regression equation (Figure 6) was applied to calibrate the algorithm score. The experimental results showed that the calibrated algorithm scores were comparable to the experts' ratings. Taken together, these findings demonstrated that, even though the developed DTW-based algorithm could be a good evaluation tool to rank the exercise performance among different subjects objectively, the algorithm score should be calibrated by experts' ratings on a small number of representative subjects. In this way, the good consistency between algorithm evaluation and experts' evaluation can be achieved for the practical applications.

Earlier studies used binary classification as well as three-point and four-point Likert scales to obtain experts' ratings for validating their algorithms [24–28]. This kind of validation is rough and likely results in inflated validation accuracy because of the wide performance range between two consecutive points, especially for binary classification and a 3-point Likert scale. To the best of our knowledge, there was only one reported study, which also used 0–100 score as the experts' rating as we did to validate the developed algorithm [21]. However, the highest correlation coefficient between their DTW-based algorithm score and expert's rating was 0.64, which was much lower than ours (r = 0.86). The improved performance from our study could be related to the selection of different motion features. Instead of using simple joint angles as Capecci et al. [21], 3D bone vectors of human skeletons were chosen in our study for better conservation of spatial information of the motion because joint angles could not define spatial information of two bones connected by the same joint. In addition, since there always exist theoretical upper and lower bounds (180 and 0 degrees) for any angle difference between two corresponding bone vectors, converting the DTW matching cost to a final percentage score is straightforward and reasonable in our study. It does not require training data and experience of experts. In this study, we assumed the upper bound was 90 degrees instead of 180 degrees based on an earlier study [18] and our practical exercise scenario.

It is worthwhile to mention that elimination of the confounding effect caused by body orientation offset is a major challenge for the algorithm development. In fact, both bone vector-based and joint position-based algorithms are very sensitive to the body orientation especially when evaluating complex whole-body exercises with rotational motions. Chua et al. [47] also pointed out this issue when they evaluated Tai Chi motion. We calculated the joint positions and bone vectors based on the local coordinate system of the human model instead of the world coordinate system in real-time, which can get rid of the error induced by the body orientation offset during the entire exercise. The compensation of body orientation offset had practical meaning for the elderly because they might not be able to orient themselves precisely as the standard virtual coach during the rehabilitation exercise.

In order to further examine the use and acceptance of exergaming technology for home-based physical rehabilitation by the primary target users (older people), we applied the technology acceptance model [48–51] and designed a questionnaire with 11 constructs (Appendix A) to evaluate user acceptance of our developed Tai Chi exergaming prototype system (Figure 3). Forty-one older adults (age 77.3 ± 5.4 years, height 159.3 ± 8.5 cm, weight 59.0 ± 9.5 kg) from a local senior welfare center participated in this survey. They were asked to try the prototype system and play the Tai Chi exergame

before giving their questionnaire responses in a five-point Likert scale (1 corresponding to "strongly disagree" and 5 corresponding to "strongly agree"). Figure 8 presents a summary of their responses. The results showed that the older people perceived relatively high vulnerability (3.21 out of 5) and severity (3.63) in terms of difficulties in self-care and independent living, and they had high intentions (Behavior intention = 4.08) to use our system in the future. They thought our system was very useful (Perceived usefulness = 4.43), positive (Attitude = 4.29), entertaining (Hedonic motivation = 3.82) and having low privacy risk (Perceived privacy risk = 1.18). Interestingly, even though the older people were somewhat confident in their capabilities to use this system for improving their health conditions (Self-efficacy = 3.75), the expected effort (3.07) and response cost (2.72) were considerably high. Taken all together, these findings implied that our developed Tai Chi exergaming prototype system is useful for the older people performing home-based physical rehabilitation exercises. However, the prototype system needs be improved to make it easy to use and cost-effective. We collected some valuable feedback from the participants to improve our prototype system, which mainly includes the following: (1) Audio effects should be added to make the exergame more entertaining and enjoyable. (2) The standard pace of Tai Chi exergame should slow down and be adjustable by each individual. (3) The size of avatar should be enlarged to be seen clearly and timely feedback for problematic motions should be provided, and (4) social networking functions (such as sharing exergaming performance score with friends) should be further developed.
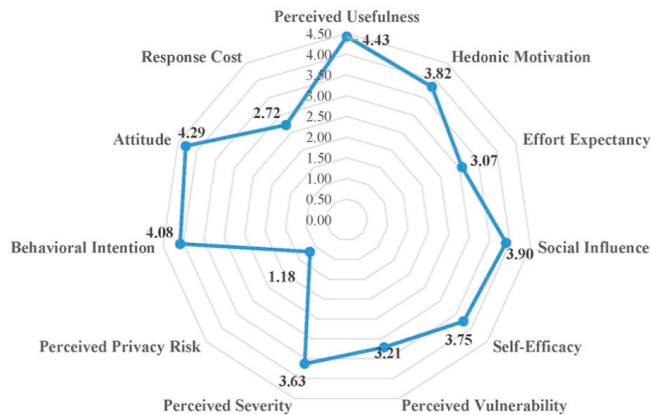
Figure 8. Results of the user acceptance questionnaire for Tai Chi exergaming prototype system. Remarks: (1) Scores from all older participants were averaged for each construct. (2) Except perceived privacy risk and response cost, all constructs are positively associated with user's intention to adopt the system.

There were several limitations in the current study. First, the conversion from DTW distance to a percentage score (0–100%) is based on the assumption that the maximum angle difference between two corresponding bone vectors is 90 degrees. Even though this assumption works fine for most body parts, the exact value of 90 is not always appropriate. Second, we focused on motion correctness for the performance evaluation in this study. The rhythm mismatch was not yet considered in the overall performance evaluation [28]. In addition, detailed feedback for problematic motions from certain body parts should be provided in the future study to timely inform the older individual for further improvements in rehabilitation exercises. Third, even though the primary target users for our developed Tai Chi exergaming system are older adults, 21 participants for validating DTW-based algorithm included both middle-aged and older adults, in order to cover a wide range of Tai Chi proficiency levels under practical constraints. Our next step will be to refine the developed prototype system and test it with a large number of older adults at the home environment for verifying practicality

of the system. Last but not least, a single Kinect sensor often generates poor skeleton tracking performance for some rotational motions during rehabilitation exercises due to self-occlusion and limited sensing range [37]. Further research on combining data from multiple Kinect sensors to achieve more accurate and robust skeleton tracking performance is needed.

5. Conclusions

We developed a DTW-based algorithm to automatically evaluate user's performance during physical rehabilitation exercises. We chose eight bones vectors of the human skeleton and body orientation as the input features and proposed a simple but innovative method to further convert the DTW matching cost to a meaningful performance score in terms of percentage (0–100%), without training data and experience of experts. The effectiveness of the proposed algorithm was tested through a follow-up experiment with 21 subjects when playing a complex whole-body exercise (Tai Chi) instead of simple repetitive exercises. Results showed that the algorithm scores had a strong positive linear relationship ($r = 0.86$) with experts' ratings and the calibrated algorithm scores were comparable to the gold standard. These findings suggested that our algorithm could be effectively used for automatic performance evaluation of an older individual when performing home-based physical rehabilitation exercises.

Author Contributions: Conceptualization, S.X. Data curation, X.Y. Formal analysis, X.Y. Funding acquisition, S.X. Methodology, X.Y. and S.X. Project administration, S.X. Software, X.Y. Supervision, S.X. Writing—original draft, X.Y. Writing—review & editing, S.X.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Questionnaire for user acceptance evaluation.

Construct ID	Questions	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
	Perceived Usefulness (PU)					
1	PU1: Using this system will help improve my health quality.	1	2	3	4	5
	PU2: Using this system will make me more effective in my life.	1	2	3	4	5
	PU3: Overall, I find this system to be useful in my life.	1	2	3	4	5
	Hedonic Motivation (HM)					
2	HM1: Using this system is fun.	1	2	3	4	5
	HM2: Using this system is enjoyable.	1	2	3	4	5
	HM3: Using this system is entertaining.	1	2	3	4	5
	Effort Expectancy (EE)					
3	EE1: Learning how to use this system is easy for me.	1	2	3	4	5
	EE2: I find this system easy to use.	1	2	3	4	5
	EE3: It is easy for me to become skillful at using this system.	1	2	3	4	5
	Social Influence (SI)					
4	SI1: People who are important to me would think that I should use this system.	1	2	3	4	5
	SI2: People who influence me would think that I should use this system.	1	2	3	4	5
	SI3: People whose opinion are valued to me would prefer that I use this system.	1	2	3	4	5
	Self-Efficacy (SE)					
5	SE1: It is easy for me to improve my health conditions by using this system.	1	2	3	4	5
	SE2: I have the capability to use this system to improve my health condition.	1	2	3	4	5
	SE3: I am able to use this system to improve my health condition without much effort.	1	2	3	4	5
	Perceived Vulnerability (PV): having little knowledge about self-care and improving personal healthcare; suffering falls.					
6	PV1: I am at the risk of suffering from the stated problems.	1	2	3	4	5
	PV2: It is likely that I will suffer the stated problems.	1	2	3	4	5
	PV3: It is possible for me to suffer the stated problems.	1	2	3	4	5
	Perceived Severity (PS): having little knowledge about self-care and improving personal healthcare; suffering falls.					
7	PS1: If I suffered the stated problems, it would be severe.	1	2	3	4	5
	PS2: If I suffered the stated problems, it would be serious.	1	2	3	4	5
	PS3: If I suffered the stated problems, it would be significant.	1	2	3	4	5

Table A1. *Cont.*

Construct ID	Questions	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
	Perceived Privacy Risk (PPR)					
8	PPR1: It would be at risk to disclose my personal health information to the doctor.	1	2	3	4	5
	PPR2: There would be high potential for loss associated with disclosing my personal health information to the doctor.	1	2	3	4	5
	PPR3: There would be too much uncertainty associated with giving my personal health information to the doctor.	1	2	3	4	5
	Behavioral Intention (BI)					
9	BI1: I intend to use this system in the future.	1	2	3	4	5
	BI2: I intend to use this system at every opportunity in the future.	1	2	3	4	5
	BI3: I plan to increase my use of this system in the future.	1	2	3	4	5
	Attitude (AT)					
10	AT1: Using this system is a good idea.	1	2	3	4	5
	AT2: Using this system is a wise idea.	1	2	3	4	5
	AT3: I like the idea of using this system.	1	2	3	4	5
	Response Cost (RC)					
11	RC1: This system is expensive to purchase (the price of Kinect is $150).	1	2	3	4	5
	RC2: I have to spend effort on learning how to use this system.	1	2	3	4	5
	RC3: Using this system will change my life style.	1	2	3	4	5

References

1. World Health Organization Ageing and Health. Available online: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed on 9 March 2019).

2. Weening-Dijksterhuis, E.; de Greef, M.H.; Scherder, E.J.; Slaets, J.P.; van der Schans, C.P. Frail institutionalized older persons: A comprehensive review on physical exercise, physical fitness, activities of daily living, and quality-of-life. *Am. J. Phys. Med. Rehabil.* **2011**, *90*, 156–168. [CrossRef]

3. Windle, G.; Hughes, D.; Linck, P.; Russell, I.; Woods, B. Is exercise effective in promoting mental well-being in older age? A systematic review. *Aging Ment. Health* **2010**, *14*, 652–669. [CrossRef]

4. Frändin, K.; Grönstedt, H.; Helbostad, J.L.; Bergland, A.; Andresen, M.; Puggaard, L.; Harms-Ringdahl, K.; Granbo, R.; Hellström, K. Long-term effects of individually tailored physical training and activity on physical function, well-being and cognition in Scandinavian nursing home residents: A randomized controlled trial. *Gerontology* **2016**, *62*, 571–580. [CrossRef]

5. Brach, J.S.; FitzGerald, S.; Newman, A.B.; Kelsey, S.; Kuller, L.; VanSwearingen, J.M.; Kriska, A.M. Physical activity and functional status in community-dwelling older women: A 14-year prospective study. *Arch. Intern. Med.* **2003**, *163*, 2565–2571. [CrossRef]

6. Rydwik, E.; Frändin, K.; Akner, G. Effects of a physical training and nutritional intervention program in frail elderly people regarding habitual physical activity level and activities of daily living—A randomized controlled pilot study. *Arch. Gerontol. Geriatr.* **2010**, *51*, 283–289. [CrossRef]

7. Organization, W.H. A guide for population-based approaches to increasing levels of physical activity. In *Implementation of the WHO Global Strategy on Diet, Physical Activity and Health*; World Health Organization: Geneva, Switzerland, 2007.

8. Lin, S.F.; Lee, J.W.; Modeste, N.; Johnson, E.G. Attitudes and beliefs predicting Taiwanese older adults' intentions to attend strength and balance training programs. *J. Appl. Gerontol.* **2012**, *31*, 260–281. [CrossRef]

9. Choi, S.D.; Guo, L.; Kang, D.; Xiong, S. Exergame technology and interactive interventions for elderly fall prevention: A systematic literature review. *Appl. Ergon.* **2017**, *65*, 570–581. [CrossRef]

10. Yardley, L.; Bishop, F.L.; Beyer, N.; Hauer, K.; Kempen, G.I.; Piot-Ziegler, C.; Todd, C.J.; Cuttelod, T.; Horne, M.; Lanta, K. Older people's views of falls-prevention interventions in six European countries. *Gerontologist* **2006**, *46*, 650–660. [CrossRef]

11. Kobayashi, R.; Nakadaira, H.; Ishigami, K.; Muto, K.; Anesaki, S.; Yamamoto, M. Effects of physical exercise on fall risk factors in elderly at home in intervention trial. *Environ. Health Prev. Med.* **2006**, *11*, 250–255. [CrossRef]

12. Liu-Ambrose, T.; Donaldson, M.G.; Ahamed, Y.; Graf, P.; Cook, W.L.; Close, J.; Lord, S.R.; Khan, K.M. Otago home-based strength and balance retraining improves executive functioning in older fallers: A randomized controlled trial. *J. Am. Geriatr. Soc.* **2008**, *56*, 1821–1830. [CrossRef]

13. Uzor, S.; Baillie, L. Investigating the long-term use of exergames in the home with elderly fallers. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 2813–2822.

14. Arlati, S.; Colombo, V.; Spoladore, D.; Greci, L.; Pedroli, E.; Serino, S.; Cipresso, P.; Goulene, K.; Stramba-Badiale, M.; Riva, G.; et al. A Social Virtual Reality-Based Application for the Physical and Cognitive Training of the Elderly at Home. *Sensors* **2019**, *19*, 261. [CrossRef]

15. Fortino, G.; Gravina, R. A cloud-assisted wearable system for physical rehabilitation. In *ICTs for Improving Patients Rehabilitation Research Techniques*; Springer: Berlin, Germany, 2014; pp. 168–182.

16. Ofli, F.; Kurillo, G.; Obdržálek, Š.; Bajcsy, R.; Jimison, H.B.; Pavel, M. Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 201–212. [CrossRef]

17. Lin, T.Y.; Hsieh, C.H.; Lee, J.D. A kinect-based system for physical rehabilitation: Utilizing Tai Chi exercises to improve movement disorders in patients with balance ability. In Proceedings of the Asia Modelling Symposium 2013: 7th Asia International Conference on Mathematical Modelling and Computer Simulation, Hong Kong, China, 23–25 July 2013; pp. 149–153.

18. Muangmoon, O.-O.; Sureephong, P.; Tabia, K. Dance Training Tool Using Kinect-Based Skeleton Tracking and Evaluating Dancer's Performance. In *International Conference on Industrial, Engineering and other Applications of Applied Intelligent Systems*; Springer: Berlin, Germany, 2017; pp. 27–32.

19. Alexiadis, D.S.; Kelly, P.; Daras, P.; O'Connor, N.E.; Boubekeur, T.; Moussa, M.B. Evaluating a dancer's performance using kinect-based skeleton tracking. In Proceedings of the 19th ACM International Conference on Multimedia, Scottsdale, AZ, USA, 28 November—1 December 2011; pp. 659–662.

20. Ren, W.; Pu, F.; Fan, X.; Li, S.; Sun, L.; Li, D.; Wang, Y.; Fan, Y. Kinect-Based Skeleton-Matching Feedback for Motor Rehabilitation: Transient Performance Effect of Shoulder training. *J. Mech. Med. Biol.* **2016**, *16*. [CrossRef]

21. Capecci, M.; Ceravolo, M.G.; Ferracuti, F.; Iarlori, S.; Kyrki, V.; Longhi, S.; Romeo, L.; Verdini, F. Physical rehabilitation exercises assessment based on hidden semi-markov model by kinect v2. In Proceedings of the 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA, 24–27 February 2016; pp. 256–259.

22. Rybarczyk, Y.; Deters, J.K.; Gonzalo, A.A.; Esparza, D.; Gonzalez, M.; Villarreal, S.; Nunes, I.L. Recognition of physiotherapeutic exercises through DTW and low-cost vision-based motion capture. In *International Conference on Applied Human Factors and Ergonomics*; Springer: Berlin, Germany, 2017; pp. 348–360.

23. Müller, M.; Röder, T. Motion templates for automatic classification and retrieval of motion capture data. In Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Vienna, Austria, 2–4 September 2006; Eurographics Association: Aire-la-Ville, Switzerland, 2006; pp. 137–146.

24. Saraee, E.; Singh, S.; Hendron, K.; Zheng, M.; Joshi, A.; Ellis, T.; Betke, M. ExerciseCheck: Remote monitoring and evaluation platform for home based physical therapy. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, Rhodes, Greece, 21–23 June 2017; pp. 87–90.

25. Semblantes, P.A.; Andaluz, V.H.; Lagla, J.; Chicaiza, F.A.; Acurio, A. Visual feedback framework for rehabilitation of stroke patients. *Inform. Med. Unlocked* **2018**, *13*, 41–50. [CrossRef]

26. Saenz-de-Urturi, Z.; Garcia-Zapirain Soto, B. Kinect-based virtual game for the elderly that detects incorrect body postures in real time. *Sensors* **2016**, *16*, 704. [CrossRef]

27. Su, C.-J.; Chiang, C.-Y.; Huang, J.-Y. Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Appl. Soft Comput.* **2014**, *22*, 652–666. [CrossRef]

28. Wei, Y.; Yan, H.; Bie, R.; Wang, S.; Sun, L. Performance monitoring and evaluation in dance teaching with mobile sensing technology. *Pers. Ubiquitous Comput.* **2014**, *18*, 1929–1939. [CrossRef]

29. Chatzitofis, A.; Monaghan, D.; Mitchell, E.; Honohan, F.; Zarpalas, D.; O'Connor, N.E.; Daras, P. HeartHealth: A cardiovascular disease home-based rehabilitation system. *Procedia Comput. Sci.* **2015**, *63*, 340–347. [CrossRef]

30. Mocanu, I.; Marian, C.; Rusu, L.; Arba, R. A Kinect based adaptive exergame. In Proceedings of the 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 8–10 September 2016; pp. 117–124.

31. Osgouei, R.H.; Soulsbv, D.; Bello, F. An Objective Evaluation Method for Rehabilitation Exergames. In Proceedings of the 2018 IEEE Games, Entertainment, Media Conference (GEM), Galway, Ireland, 15–17 August 2018; pp. 28–34.

32. Salvador, S.; Chan, P. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **2007**, *11*, 561–580. [CrossRef]

33. Müller, M. *Information Retrieval for Music and Motion*; Springer: Berlin, Germany, 2007; Volume 2.

34. Shokoohi-Yekta, M.; Wang, J.; Keogh, E. On the non-trivial generalization of dynamic time warping to the multi-dimensional case. In Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, 30 April–2 May 2015; pp. 289–297.

35. Lanciault, R. Mecanim Humanoids. Available online: https://blogs.unity3d.com/2014/05/26/mecanim-humanoids/ (accessed on 15 March 2019).

36. Filkov, R. K2-Asset On-line Documentation. Available online: https://ratemt.com/k2docs/ (accessed on 9 May 2018).

37. Moon, S.; Park, Y.; Ko, D.W.; Suh, I.H. Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering. *Int. J. Adv. Robot. Syst.* **2016**, *13*. [CrossRef]

38. Karatsidis, A.; Bellusci, G.; Schepers, H.M.; de Zee, M.; Andersen, M.S.; Veltink, P.H. Estimation of ground reaction forces and moments during gait using only inertial motion capture. *Sensors* **2016**, *17*, 75. [CrossRef]

39. Guo, L.; Xiong, S. Accuracy of base of support using an inertial sensor based motion capture system. *Sensors* **2017**, *17*, 2091. [CrossRef]

40. Reips, U.-D.; Funke, F. Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behav. Res. Methods* **2008**, *40*, 699–704. [CrossRef]

41. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [CrossRef]

42. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]

43. Ćwiękała-Lewis, K.J.; Gallek, M.; Taylor-Piliae, R.E. The effects of Tai Chi on physical function and well-being among persons with Parkinson's Disease: A systematic review. *J. Bodyw. Mov. Ther.* **2017**, *21*, 414–421. [CrossRef]

44. Sun, J.; Kanagawa, K.; Sasaki, J.; Ooki, S.; Xu, H.; Wang, L. Tai chi improves cognitive and physical function in the elderly: A randomized controlled trial. *J. Phys. Ther. Sci.* **2015**, *27*, 1467–1471. [CrossRef]

45. Kottner, J.; Audigé, L.; Brorson, S.; Donner, A.; Gajewski, B.J.; Hróbjartsson, A.; Roberts, C.; Shoukri, M.; Streiner, D.L. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int. J. Nurs. Stud.* **2011**, *48*, 661–671. [CrossRef]

46. Kim, T.; Xiong, S. Comparison of seven fall risk assessment tools in community-dwelling Korean older women. *Ergonomics* **2017**, *60*, 421–429. [CrossRef]

47. Chua, P.T.; Crivella, R.; Daly, B.; Hu, N.; Schaaf, R.; Ventura, D.; Camill, T.; Hodgins, J.; Pausch, R. Training for physical tasks in virtual environments: Tai Chi. In Proceedings of the IEEE Virtual Reality 2003, Los Angeles, CA, USA, 22–26 March 2003; pp. 87–94.

48. Gao, Y.; Li, H.; Luo, Y. An empirical study of wearable technology acceptance in healthcare. *Ind. Manag. Data Syst.* **2015**, *115*, 1704–1723. [CrossRef]

49. Kim, S.S. The integrative framework of technology use: An extension and test. *MIS Q.* **2009**, *33*, 513–537. [CrossRef]

50. Bhattacherjee, A.; Hikmet, N. Physicians' resistance toward healthcare information technologies: A dual-factor model. Proceedings of 2007 40th the Annual Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 3–6 January 2007.

51. Lee, Y.; Larsen, K.R. Threat or coping appraisal: Determinants of SMB executives' decision to adopt anti-malware software. *Eur. J. Inf. Syst.* **2009**, *18*, 177–187. [CrossRef]

Neurophysiological Characterization of a Non-Human Primate Model of Traumatic Spinal Cord Injury Utilizing Fine-Wire EMG Electrodes

Farah Masood [1,2,*], Hussein A. Abdullah [1], Nitin Seth [1], Heather Simmons [3], Kevin Brunner [3], Ervin Sejdic [4], Dane R. Schalk [3], William A. Graham [5], Amber F. Hoggatt [6], Douglas L. Rosene [7], John B. Sledge [8] and Shanker Nesathurai [3,5,9]

1   School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada
2   The Department of Biomedical Engineering, Al-Khwarizmi College of Engineering, Baghdad University, Baghdad 47146, Iraq
3   The Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI 53715, USA
4   The Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA
5   The Division of Physical Medicine and Rehabilitation, Department of Medicine, McMaster University, Hamilton, ON L8S 4K1, Canada
6   The Center of Comparative Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
7   The Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA 02118, USA
8   The Lafayette Bone and Joint Clinic, Lafayette, LA 70508, USA
9   The Department of Physical Medicine and Rehabilitation, Hamilton Health Sciences, St Joseph's Hamilton Healthcare, Hamilton, ON L9C 0E3, Canada
*   Correspondence: fmasood@uoguelph.ca or farahmasood16@gmail.com

Abstract: This study aims to characterize traumatic spinal cord injury (TSCI) neurophysiologically using an intramuscular fine-wire electromyography (EMG) electrode pair. EMG data were collected from an agonist-antagonist pair of tail muscles of Macaca fasicularis, pre- and post-lesion, and for a treatment and control group. The EMG signals were decomposed into multi-resolution subsets using wavelet transforms (WT), then the relative power (RP) was calculated for each individual reconstructed EMG sub-band. Linear mixed models were developed to test three hypotheses: (i) asymmetrical volitional activity of left and right side tail muscles (ii) the effect of the experimental TSCI on the frequency content of the EMG signal, (iii) and the effect of an experimental treatment. The results from the electrode pair data suggested that there is asymmetry in the EMG response of the left and right side muscles (p-value < 0.001). This is consistent with the construct of limb dominance. The results also suggest that the lesion resulted in clear changes in the EMG frequency distribution in the post-lesion period with a significant increment in the low-frequency sub-bands (D4, D6, and A6) of the left and right side, also a significant reduction in the high-frequency sub-bands (D1 and D2) of the right side (p-value < 0.001). The preliminary results suggest that using the *RP* of the EMG data, the fine-wire intramuscular EMG electrode pair are a suitable method of monitoring and measuring treatment effects of experimental treatments for spinal cord injury (SCI).

Keywords: fine-wire intramuscular EMG electrode; non-human primate model; traumatic spinal cord injury; wavelet transform; relative power; linear mixed model

---

1. Introduction

Traumatic spinal cord injury (TSCI) is a serious neurological condition. Worldwide, there are approximately 180,000 new cases each year [1,2]. The most common causes of TSCI are motor vehicle

collisions, falls, sports-related activities, and interpersonal violence [2,3]. TSCI typically damages the motor, sensory and autonomic fibre tracts. Patients experience a spectrum of clinical abnormalities such as limb paralysis, dysesthesia, as well as bowel and bladder dysfunction [3–5]. The burden of this condition is shared by the person affected, family members, the community, as well as the healthcare system. Long-term survival and quality of life have improved due to enhanced rehabilitation and medical interventions. However, it has not been convincingly demonstrated that current pharmacological treatments have substantially improved spinal cord function.

Conceptually, the understanding of TSCI impairments and recovery can be advanced through animal models [4–9]. Non-human primate models (NHP) are particularly valuable due to the anatomic and physiological similarities between NHP and human beings [6,7]. The tail in NHPs is analogous to a human limb. The tail is integral to performing functional tasks such as standing on hindlimbs as well as reciprocal movements that aid in balance during ambulation [10]. As described later in this paper, the tail exhibits other features analogous to human beings such as limb dominance or preference.

SCI impairments are usually measured by observing behaviours as well as histopathological assessment [11]. Electromyography (EMG) signals confer certain advantages over these other methods. It is particularly useful for assessing TSCI as EMG data collected on a serial basis as this approach facilitates the characterization of motor unit (MU) activation and recruitment [8]. EMG signals can be recorded through multiple channels, permitting the simultaneous assessment of multiple muscle groups. This is particularly helpful in evaluating agonist–antagonist muscle pairs, in which the assessment of muscle co-contractions are associated with central nervous system disorders [4,9].

EMG data can be obtained through surface recording or intramuscular electrodes placed into the muscles. There is limited literature regarding the nature of intramuscular EMG signals after TSCI in animals and humans. The majority of human studies have been based on surface EMG electrodes. As a general construct, TSCI results in a perturbation of the EMG signal. Calancie et al. [12] studied the recovery of volitional activity after acute SCI in human beings utilizing pairs of surface EMG electrodes. This study demonstrated perturbations in some EMG characteristics including abnormalities in recruitment. Lewko [13] utilized surface EMG electrodes and noted disturbed behaviour in spinal cord conductivity with quiet standing. Nout and Rosenzweig among others [8], ref. [9] developed a NHP model of SCI, and the results demonstrated a significant difference in EMG amplitude and temporal patterns between the healthy and the SCI subjects. Also, they noted uncoordinated muscle activity during the post-lesion condition. Wiegner et al. and Shahani et al. studied the recruitment pattern of MUs post-SCI [14] and cortical pathology [15] using needle EMG signals. The results indicate that MUs fire irregularly with low discharge rate post-SCI. Also, the inter-discharge intervals (IDIs) have a positive serial correlation which results from the decreased variability in length of the adjacent intervals. Capogrosso et al. [16] developed a brain-spine interface to modulate the consequences of TSCI in NHP. EMG signals during continuous locomotion were recorded and averaged to calculate the spatiotemporal maps of the motoneuron activation in monkeys. Their results suggested a practical translational pathway for conceptual analysis studies and investigational applications in human with SCI.

The bio-signals such as EMG are non-stationary signals. Furthermore, EMG signals during volitional muscle contraction have a random nature which means the active MUs have an irregular firing rate [17]. Thus, determining the method to select relevant features of the EMG signal becomes challenging. In this context, several signal processing techniques have been tested, with the goal of developing a robust method [18]. To advance this goal, mathematical transformations can be utilized. Specifically, the EMG signal could be represented in different domains; time, frequency, or a time-frequency/wavelet domain. The amplitude and frequency content of the EMG signal help in understanding the physiology and the pathology aspects of muscle activity. The frequency content of the EMG signal mainly has been calculated using the fast Fourier transform (FFT). However, FFT may not be an appropriate choice for some cases such as the dynamic and variant level of contractions [19,20]. To overcome this problem, the wavelet transforms (WT) is applied. The WT is an

efficient mathematical analysis method for temporally nonstationary and spatially nonhomogeneous bio-signals [21]. Also, this approach has proved its ability to represent precise measurements and to extract useful information from non-stationary biomedical signals [22–25]. Decomposition, denoising, and pattern classification are the most common applications of the WT in the EMG field [18,22,26–34]. Phinyomark et al. [30] investigated the usefulness of extracting EMG features from the multi-resolution wavelet decomposition process. The results showed that the reconstructed EMG of the first and second level (detail coefficients) have improved the class separability. Yamada et al. [26] introduced a new EMG decomposition algorithm by adopting the principal component analysis of wavelet coefficients. This method showed a higher decomposition accuracy when compared to conventional wavelet methods. Fang et al. [27] decomposed EMG signals into their constituent single motor units using wavelet spectrum matching, and the results were satisfying.

The intramuscular fine-wire electrode data of this experiment is particularly unique as it is collected both prior and subsequent to a TSCI. The longitudinal nature of the data, however, presents the challenge of limiting the sampling rate of the sensors to allow for long-term monitoring. In addition, free and dynamic movements must be permitted to allow for data collection that reflected the subjects true volitional control as the limb would be naturally used. Thus, conventional decomposition methods of EMG data are not suitable for this problem as these methods rely on higher sampling frequency (around 25 kHz) [35] as well as the need for isometric contractions [36–39].

In this work, the intramuscular fine-wire EMG data have been analyzed using the WT as an EMG decomposition method and relative power (RP) as a metric of the active MUs within each WT sub-band. Specifically, this work will address the following questions:

1. Is the intramuscular fine-wire electrode pair data capable of detecting limb dominance in the subjects prior to lesion?
2. In the post-lesion period, is there a change in EMG activity attributed to the experimental spinal cord injury and how it could be characterized in term of frequency content?
3. What is the difference in the EMG activity between the control and the treatment group in the post-lesion period (i.e., is there a treatment effect)?

The collected intramuscular EMG dataset contains information obtained before and after an experimental TSCI, allowing each subject to serve as normal control. Accurate measurement of impairment and recovery in a model of TSCI has significant implications for the identification and development of TSCI therapeutics. With no standardized model for combined pre-lesion, post-lesion, and recovery analysis, WT with multi-resolution data could provide evidence of being able to account for multiple effects in a single model. The extended recording period both before and after TSCI in the NHP model is necessary to replicate TSCI; recovery in humans will be recognized months after injury. Repeated measurements in the NHP model permit the long-term evaluation effects of experimental therapies. It is expected that the developed NHP model and its preliminary results will provide a better understanding of the TSCI and may help with the prediction of recovery in human limbs.

2. Materials and Methods

The focus of the SCI model was to create a lesion that mimics human SCI without clinical impairments. The experimental methods to create the lesion are described in detail in previous publications [4,7], but are presented here briefly. The subjects, six adult Macaca fasicularis monkeys, underwent two surgical procedures. The first to insert a small transmitter (PhysioTel D70, Data Sciences International, St. Paul, MN, USA) with attached fine-wire intramuscular electrode pair into the lower back of the subjects. The electrodes were implanted into the left and right flexor cauda longus and brevis muscles (tail muscles). The EMG signals were measured using the fine-wire electrodes with 10 mm inter-electrode distance, 10 mm exposed wire length, 1000 Hz sampling frequency, and common ground.

Baseline pre-lesion EMG data from the fine-wire electrode pair were collected during voluntary movement of NHPs within their home enclosures for 30 days. The data were collected from Monday to Friday (excluding holidays) for approximately 1 h per day. After 30 days, the subjects underwent a second surgical procedure. A small laminotomy was completed at the L5 vertebral level. An epidural balloon catheter was inserted and advanced approximately 10 cm cranial to the level of the lower thoracic spinal cord. The lesion was created by inflation of the balloon with air. The balloon remained inflated for 60 s and then was deflated. The catheter was then removed, and the surgical incision closed. The lesions initially disrupted the grey and white matter at the site of lesion creation. The histopathological features, that were created by this method are similar to human TSCI [4]. To mimic the time frame between human injury and administration of emergency medical treatment, the subjects remained under anesthesia for one hour. Four subjects that received the experimental combination treatment for 90 days while two subjects did not receive any treatment. Combination treatment consisted of a bolus 0.2 mg/kg bolus of thyrotropin releasing hormone (TRH), followed by a continuous infusion of 0.2 mg/kg per hour for 1 h. These subjects were also treated with 60 mg of selenium and 80 IU of vitamin E daily. TRH is a tripeptide produced by the hypothalamus; selenium and Vitamin E are antioxidants. The combination of these agents may modulate the physiological sequelae of TSCI [40]. This protocol was approved by the Institutional Animal Care and Use Committee (IACUC) at Harvard University and the University of Wisconsin at Madison. As such, EMG data from the tail is analogous to EMG data obtained from the limb of a human being with TSCI EMG data were collected from all six subjects. However, in one subject, the data from one side was not recorded due to a technical error. As such, raw data in this paper represents the experience of three subjects that received the combination therapy, and two subjects who did not receive treatment. Figure 1 illustrates a sample of the recorded raw EMG data.
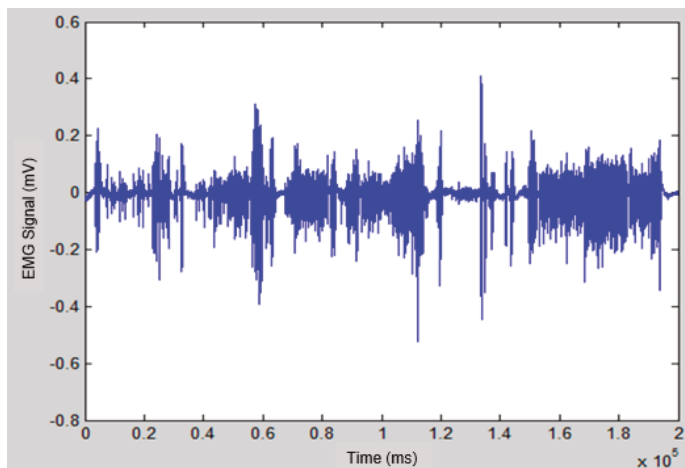
Figure 1. Graphical representation of the raw electromyography (EMG) signal.

In this NHP model, the induced lesion produced a TSCI limited to the upper motor neuron tracts that supply the lower motor neurons of the tail muscles [7]. Lesion of such nature result in a perturbation in the MUs discharge properties; i.e., abnormalities in the inter-discharge interval, firing rate, and floating serial correlation coefficient [14,41]. These abnormalities have been typically characterized in non-physiological experimental conditions such as low force levels and isometric contraction frequency [14,37–39]. In this work, the EMG data were collected when the subjects engage in physiological activities (i.e., locomotion in a cage). The proposed EMG analysis method consists of the following three steps:

A. The raw EMG data obtained from daily recordings were filtered using a bandpass filter (4th order Butterworth filter with a lower and an upper cut off frequency of 10 and 450 Hz respectively). A notch filter with 60 Hz was also applied to eliminate the power line noise, and the input signal was processed forward and backward to solve phase shift problems. The EMG conditioning steps have been implemented using MATLAB software (MathWorks, Natick, MA, USA).

B. A decomposition process was applied using wavelet transforms. Each WT sub-band was assumed to represent the firing rate of a group of MUs. Also, it was assumed that the *RP* of each individual sub-band reflects the level of activity for these MUs. Thus, increases or decreases in *RP* may characterize the recruitment pattern process of the MUs through different conditions of the experiment. The filtered EMG signals were broken down into seven frequency sub-bands using the WT. The discrete wavelet transforms (DWT) was selected for this work because it has non-redundant results, and it required less computational time and costs [42,43]. A Daubechies mother wavelet of fourth order 'db4' was used due to its similarity to the triphasic pattern of the motor unit action potential [44]. Consistent with the analysis of other bio-signals, DWT decomposition was performed using six levels [43,45–48]. The wavelet analysis was performed in two steps, as presented in Figure 2:

1. The EMG signals were decomposed into seven sub-bands, one approximate coefficient (cA6), and six detail coefficients (cD1, … , cD6).
2. The EMG signal was then reconstructed at each level using inverse discrete wavelet transform, and seven EMG reconstructed signals (A6, D1, … , D6) were obtained from their coefficients (cA6, cD1, … , cD1). Table 1 shows the frequency ranges of the seven EMG sub-bands.

C. To evaluate the changes in EMG sub-bands during different phases of the experiment, these changes were characterized using the RP. The probabilistic distribution of the spectral power was quantified by calculating the relative power of each spectral component [49]. To obtain the RP, firstly the power spectral density was determined for each reconstructed EMG sub-band signal. Then, the *RP* for each individual sub-band was calculated using the following formula [50]:

$$RP(\%) = \frac{SBP}{TP}, \tag{1}$$

where:

RP: the relative power of the desired sub-band.
SBP: the power of the desired sub-band (e.g., A6, D1, … or D6).
TP: the total power of all the sub-bands (A6 + D1, … , + D6).

Table 1. The frequency ranges of the seven EMG reconstructed sub-bands.

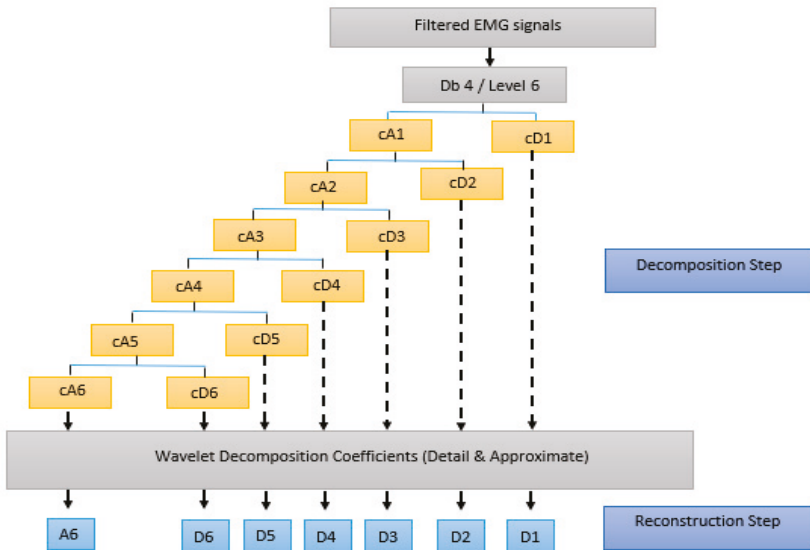| Wavelet Decomposition Level	Frequency Range/Hz	Reconstructed EMG Sub-Bands
1	250–500	D1
2	125–250	D2
3	62.5–125	D3
4	31.25–62.5	D4
5	15.63–31.25	D5
6	7.81–15.63	D6
6	0–7.81	A6

Figure 2. Wavelet analysis; decomposition and reconstruction steps.

The goal was to address the research questions by analyzing specific subsections of the data and testing specific effects. To test the lateralization effect (attributed to limb dominance), only the pre-lesion data were considered for the two sides (left and right). Then, the lesion effect was tested using the data from both pre-and post-lesion periods, and two separate models were fitted one for each side. To test the treatment effect, two separate models (left and right) were fitted using only the post-lesion data. The data in this work is considered a clustered longitudinal dataset with three levels; days are nested in frequency sub-bands, which are in turn nested within subjects (day: is level 1, sub-bands: is level 2, and subject: is level 3). As a result of such hierarchical structure as in Figure 3, the non-independence problem of the observations would arise which requires an appropriate statistical analysis method. A mixed model is a statistical method that was developed particularly to address the non-independence (correlated, or repeated measurements) issue by including the random effect term in its model. The developed mixed models controlled for the within-subject correlation by including a random effect for the subject and the nested sub-bands within each subject, with a variance-covariance structure and restricted maximum likelihood estimation. The models were implemented in the statistical software R using the (lme4) package. The significance level was set to ($\alpha = 0.05$). The *RP* for any given subject at (day) *i* for (sub-band) *j* nested within (subject) *k* denoted as *RPijk*, is represented in the following equation (see a summary of parameters in Table 2):

$$RP_{ijk} = \beta_0 + \beta_1 \, Day_{ijk} + \beta_2 \, EFFECT_{jk} + \beta_3 \, EFFECT_{jk} \, Day_{ijk} + \beta_4 Freq_{jk} +$$
$$\beta_5 \, Freq_{jk} \, Day_{ijk} + \beta_6 \, Freq_{jk} \, EFFECT_{jk} + \beta_7 \, Freq_{jk} \, EFFECT_{jk} \, Day_{ijk} \, (fixed) +$$
$$\alpha_{0k} + \alpha_{1k} \, Day_{ijk} + \alpha_{0jk} + \alpha_{1jk} \, Day_{ijk} + \varepsilon_{ijk} \, (random),$$

(2)

Table 2. Parameters used in Equation (2).

Parameter	Definition
RP_{ijk}	response (relative power)
Day_{ijk}	experiment day
$Freq_{jk}$	frequency sub-band (D1, D2, D3, D4, D4, D6, and A6)
EFFECT	the main effect in the models, so for: Model 1: the side effect (categorical variable with two levels (left and right side)) Model 2: the lesion effect (categorical variable with two levels (pre- and post-lesion)) Model 3: the treatment effect (categorical variable with two levels (control and treatment group))
β_0 **through to** β_7	the fixed effect associated with the (intercept, Day, EFFECT, EFFECT*Day, Freq, Freq*Day, Freq* EFFECT, and Freq* EFFECT*Day)
α_{0k}, α_{1k}	the subject random effect associated with the intercept and Day slope, respectively
α_{0jk}, α_{1jk}	the random effect of a frequency sub-band nested within a subject associated with the intercept and Day slope respectively
ε_{ijk}	random error
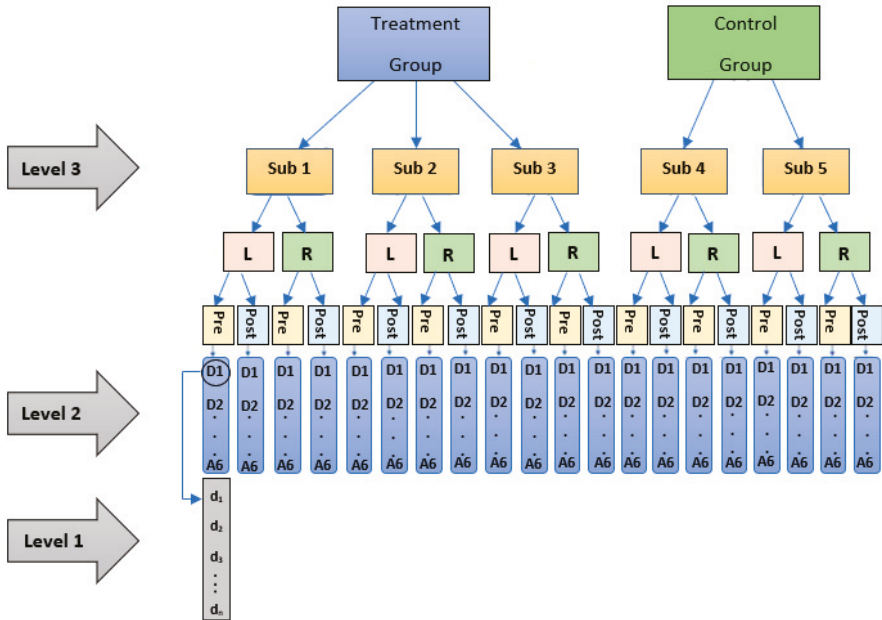
Figure 3. Data Structure: The data in this work was collected from the left (L) and right (R) side of the tail for five subjects during multiple experiment days (d1, d2, ... , dn) for (pre- and post-lesion period). The data of each day was decomposed into seven frequency sub-bands (D1, D2, D3, ... , A6). Three of the subjects received a combination of treatment (Treatment group); the remaining two subjects did not receive any treatment post-lesion (Control group).

3. Results

The results were presented in the form of answers to the three research questions that were stated in the introduction. They are as follows:

(1)   *Is the intramuscular fine-wire electrode pair capable of detecting limb dominance in the subjects prior to lesion?*

To test the symmetry of tail muscle activity on both sides, the *RP* of the EMG signals during the *pre-lesion* period were analyzed using a linear mixed model. The interaction of the frequency sub-band and the side variable in the statistical model demonstrated a significant effect (p-value < 0.001). Given that the (side*frequency) interaction is significant, Tukey's mean comparisons were generated to identify what means differ significantly. These results suggested that side variable had a significant effect on the *RP* value of the D1, D2, D4, and D5 sub-bands. The estimated mean of the *RP* for the D1, D2 and D4 sub-bands of the *left side* were significantly higher than that of the *right side*. On the contrary, the estimated mean of the *RP* for the D5 sub-band on the *left side* was significantly lower than that of the *right side*. The estimated mean for the *RP* across the different frequency sub-bands for both sides is illustrated in Figure 4. Taken together, these results showed that the left and right tail muscles have asymmetry activation.
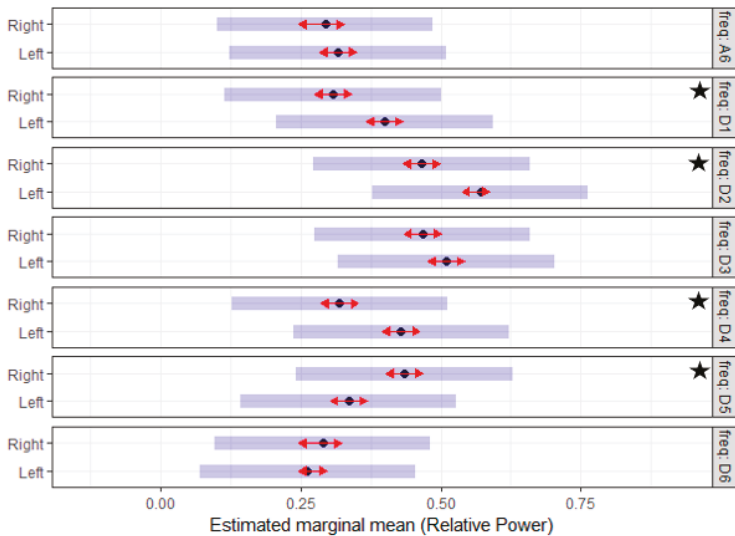
Figure 4. The estimated mean of the relative power (*RP*) of seven reconstructed EMG sub-bands prior to the creation of traumatic spinal cord injury (TSCI) for the left and right side of the tail. Of note for each band, there is a difference in *RP* value when compared to the left and right side of the tail. The D2 sub-band of the left and right side has the maximum RP, and the significant difference between the two sides is at the D1, D2, D4, and D5 sub-bands. The star indicates a significant difference.

(2)   *In the post-lesion period, is there a change in the EMG activity attributed to the experimental spinal cord injury and how it could be characterized in terms of RP?*

To answer this question, the effect of the created lesion was analyzed using a linear mixed model. The lesion effect was studied by testing the difference between the EMG characteristics during the *pre-* and *post-lesion* period. The interaction of the frequency sub-band and the lesion variable in the statistical models of both sides demonstrated a significant effect (p-value < 0.001). Given that the (lesion*frequency) interaction was significant, Tuckey's means comparisons were generated to identify which means differ significantly. Figures 5 and 6 summarized the estimated mean of the *RP* values for different frequency sub-bands of the *pre-* and *post-lesion* group for the two sides. On the *left side*, the estimated mean of the *RP* for the D4, D6 and A6 sub-bands was significantly higher in the *post-lesion* period compared to the *pre-lesion* period. On the *right side*, the estimated mean of the *RP* for the

D4, D6, and A6 sub-band was also higher significantly in the *post-lesion* period. On the other hand, the estimated mean of the *RP* for the D1 and D2 sub-bands was significantly lower in the *post-lesion* period compared to *pre-lesion* period. The results suggested that the created lesion had a clear effect on the discharge properties of MUs, and with this technique, changes in discharge properties can be detected even when there is no clinical evidence.

Figure 5. The estimated mean of the *RP* of seven reconstructed EMG sub-bands prior and post to the creation of TSCI for the *left side* of the tail. Of note, the *RP* values for the frequency sub-bands (D4, D6, and A6) are significantly higher in the post-lesion period. The star indicates a significant difference.

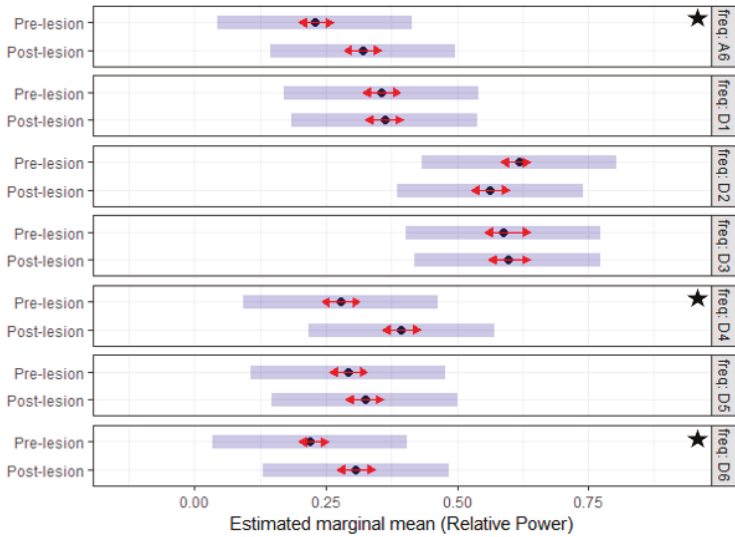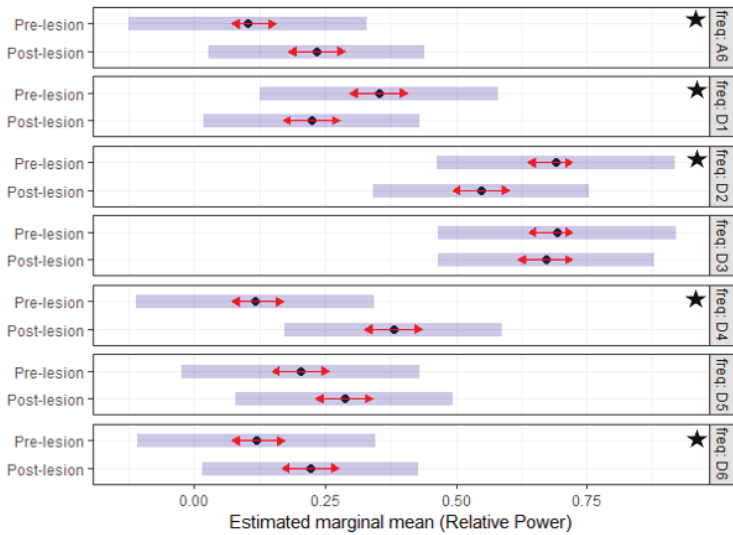**Figure 6.** The estimated mean of the *RP* of seven reconstructed EMG sub-bands prior and post the creation of TSCI for the *right side* of the tail. Of note, the *RP* values for the lower frequency sub-bands (D4, D6, and A6) are significantly higher in the post-lesion period, while the higher frequency sub-bands (D1 and D2) are significantly lower in the post-lesion period. The star indicates a significant difference.

(3)   What is the difference in the EMG activity between the control and the treatment group in the post-lesion period (Treatment effect)?

The *post-lesion data* were utilized to generate two separate mixed models, one for each side. The potential effect of the treatment was analyzed using the *RP* of the left and the right sides incorporating an analysis which compared the treatment and the control groups. The interaction of frequency sub-band and treatment variables in the models of both sides demonstrated a significant effect (*p*-value < 0.001). Given that the (treatment*frequency) interaction was significant, Tukey's mean comparisons were generated to identify what means differ significantly. Figures 7 and 8 summarize the estimated mean of the *RP* values for different sub-bands of the control and treatment group for the two sides. On the *left side*, the D1, D2, D3, and D6 sub-bands have a significant difference, while on the *right side* the effect was significant in all the sub-bands except the D2. The results suggested that there is a significant difference in the discharge properties of MUs of the treatment and the control groups during the post-lesion period.

Figure 7. The estimated mean of the *RP* of seven reconstructed EMG sub-bands post-lesion for the treatment (Tr) and the control (Ctrl) groups of the *left side*; of note, the *RP* values for the frequency sub-bands (D1, D2, D3, and D6) are significantly different. Subjectively, the distribution of the *RP* in the treatment group is similar to the *RP* distribution in the pre-lesion period for all the subjects. The star indicates a significant difference.
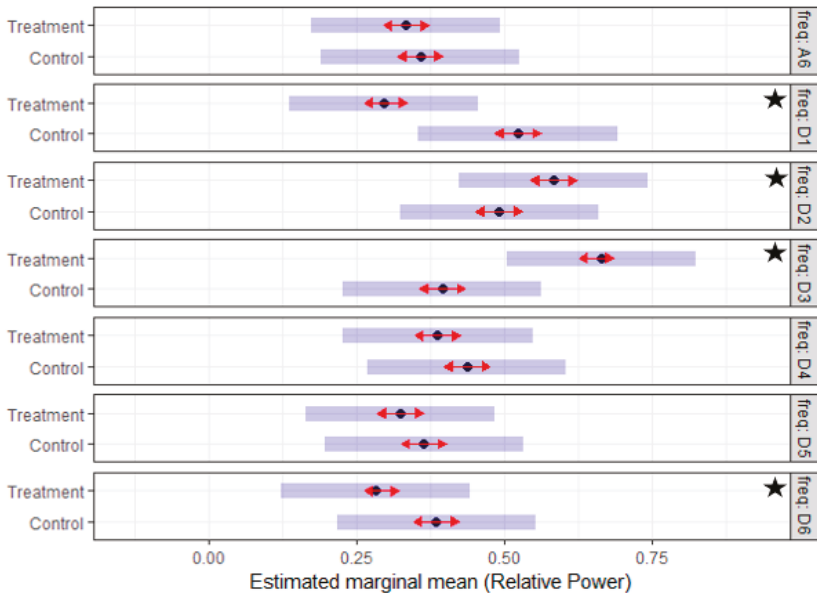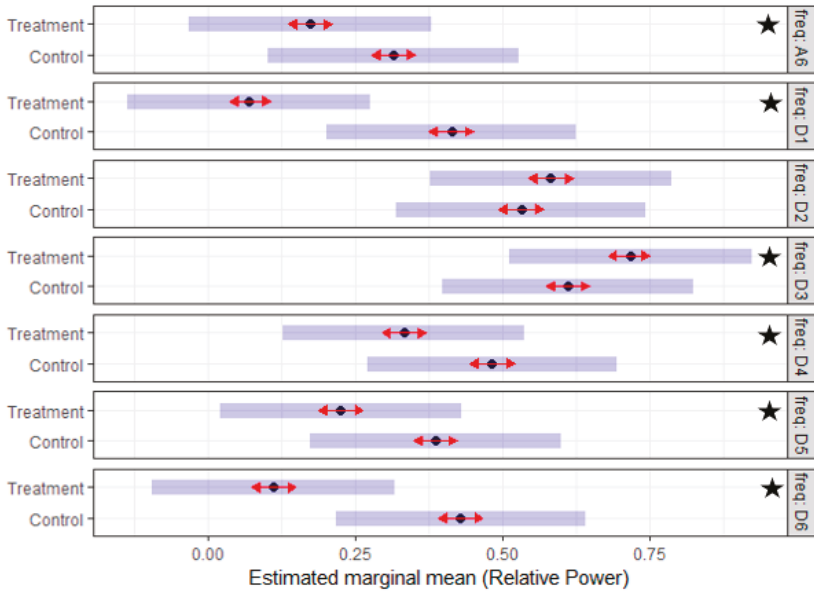
Figure 8. The estimated mean of the *RP* of seven reconstructed EMG sub-bands post-lesion for the treatment (Tr) and the control (Ctrl) group of the *right side*; of note, the *RP* values for frequency sub-bands (D1, D3, D4, D5, D6, and A6) are significantly different. Subjectively, the distribution of the *RP* in the treatment group is similar to the *RP* distribution in the pre-lesion period for all the subjects. The star indicates a significant difference.

4. Discussion

4.1. Non-Human Primates Appear to Exhibit Limb Dominance

In human beings, the construct of limb dominance or preference is well accepted; i.e., most people use their right arm for functional tasks, and a minority use the left arm. A select few people can use both arms equally well (ambidextrous). The question of whether monkeys have limb dominance remains a subject of scientific inquiry. In this analysis, there is asymmetry related to the *RP* of EMG data derived from the left and the right tail muscles. These results suggest that the long-tailed macaque (*Macaca fascicularis*) exhibits limb preference and/or dominance. This is also consistent with the observations of the veterinary staff involved with this project. Generally, limb dominance refers to the preferential use of one limb to perform functional tasks [51]. This asymmetry in the pre-lesion (normal control) period has implications for any analysis subsequent to the experimental TSCI. Conceptually, volitional motor control involves a two-circuit pathway; i.e., the upper motor neurons and the lower motor neurons of the corticospinal tracts. The upper motor neurons originate in the cortex, travel the internal capsule and pyramids, and terminate into the grey matter of the spinal cord. The lower motor neurons originate in the grey matter of the spinal cord, exit via the nerve roots and reach the muscles via the plexus and peripheral nerves. The data suggest that the neural network that controls the left and right neurophysiological circuit is not symmetrical. Therefore, it stands to reason that experimental TSCI will have different effects on the separate circuits. This is consistent with human disease, to the extent that a lesion in one part of the central nervous system may have different manifestations related to the limbs. As such, the effects of the lesion, from a neurological perspective, were analyzed in model 2 separately for both the left and the right side.

4.2. Experimental Traumatic Spinal Cord Injury (TSCI) Causes Perturbation of Electromyographic (EMG) Data

The experimental TSCI resulted in both a physiological and histopathological perturbation [4]. In this analysis, there was a difference in the *RP* values when comparing the pre- and post-lesion data; this was statistically significant. Specifically, the low frequency (LF) sub-bands (D4, D6, and A6) were increased significantly during the post-lesion period on both sides. Also, the high frequency (HF) sub-bands (D1 and D2) were decreased significantly during the post-lesion period on the right side. These results are consistent with the literature [14]. As a general construct, neurological disease is associated with abnormalities in the firing of the motor units and leads to a lower discharge frequency [14,15,52]. This, in turn, can affect the distribution of the high and low-frequency components of EMG signals. This shift in the *RP* value may be related to the disorganized, spontaneous firing of single or multiple motor units, which affect the frequency distribution. Speculatively, a component of these abnormalities was be related to spasticity; in TSCI, there may preferentially affect the recruitment of Type I or Type II motor units, which could also distort the frequency content. The elucidation of these relationships requires further research. With this method, the change in recruitment behaviour post the TSCI might be detectable even before the clinical evidence.

4.3. Combination Treatment Is Associated with Treatment Effect

The combination of TRH, selenium and vitamin E is associated with a treatment effect. Specifically, three subjects received a combination treatment and two subjects did not receive treatment. In the post-lesion period, there was a significant difference in *RP* values for most of the frequency sub-bands of the treatment and the control groups. These results suggested there is a clear difference in the discharge properties of MUs of the treatment group during the post-lesion period. This finding was present for both the left and the right side. However, in the context of a limited number of subjects, this should be considered as a preliminary inference. The rationale for this combination treatment in TSCI and the implications thereof are published elsewhere [40].

4.4. Recording of EMG Signals from Surface, Needle and Wire Electrodes

There are a variety of methods used to obtain EMG data from humans and animals. As a general construct, there are three types of electrodes: surface, needle, and wire electrode. The neurophysiological characteristics and the interpretation of the data derived are not synonymous. Surface electrodes are placed over the skin; they can be manufactured from a variety of materials and formed into various sizes and shapes. Surface recording over the skin is subject to distortion of the EMG signal due to the interposition of the fascia and the subcutaneous tissues. With limb movement, the skin, the subcutaneous tissue and the muscles move asymmetrically. This contributes further to movement artifact. These tissues are effectively a low pass filter. In comparison, needle electrodes are made of metal; subtypes include monopolar, concentric, single fibre and macroelectrodes. Wire electrodes, in contrast, can vary in diameter and length. Wire electrodes allow for long term implantation in subjects and are generally well tolerated. When compared to surface electrodes, there is less cross-talk. As a conceptual argument, data from wire electrodes aggregates the EMG signal from muscle fibres in multiple motor units. The signal is potentially affected by multiple factors including the proximity of the signal generator to the wire, phase cancellation, size of fibres, and type of fibres (i.e., fast twitch, slow twitch). As well, there is some contribution related to the neural network that controls the agonist–antagonist pair.

The fine-wire recordings from this model are particularly unique. Specifically, the EMG data were obtained prior and subsequent to TSCI. Additionally, the data were obtained while the subjects were moving freely (dynamic movements), over an extended period of time. Some potential limitations include the relatively few numbers of subjects; this is, in part related to cost. As well, the collection of longitudinal nature precluded a very high sampling rate.

5. Conclusions

In conclusion, this research characterizes the frequency content of intramuscular EMG signals recorded pre- and post-TSCI in a validated NHP model. The EMG data were collected in a longitudinal nature and were obtained from the left- and right side of the Macaca fasicularis tail muscles engaged in the dynamic/free activity. In the absence of the TSCI, the EMG data demonstrated an asymmetry activation of the two sides; this is consistent with the human phenomenon of limb preference and/or dominance. Perturbation with an experimental lesion resulted in a clear EMG consequence. Of note, the effect of the lesion was different on the left and the right side. Also, there is a preliminary inference that treatment with a combination of TRH, selenium and vitamin E may improve recovery. The results indicate that the *RP* of the decomposed EMG data adds a new dimension to the evaluation of impairment and recovery in this NHP model of experimental TSCI. This analysis might lead to a new assessment index for motor unit activity and progression of recovery from TSCI.

Author Contributions: Conceptualization, F.M., H.A.A., S.N., W.A.G., D.L.R., and J.B.S.; Methodology, F.M.; Data Acquisition, K.B., H.S., D.R.S., W.A.G., A.F.H., and S.N.; Writing—original draft preparation, F.M., S.N., H.A.A.; Writing—review and editing, F.M., S.N., H.A.A., N.S., E.S., H.S., K.B., D.R.S., W.A.G., A.F.H., D.L.R., J.B.S.; Project administration, H.S., K.B., W.A.G., A.F.H., S.N.; Funding acquisition, S.N.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lee, B.B.; Cripps, R.A.; Fitzharris, M.; Wing, P.C. The global map for traumatic spinal cord injury epidemiology: Update 2011, global incidence rate. *Spinal Cord* **2014**, *52*, 110–116. [CrossRef]

2. Nesathurai, S.; Blaustein, D. *The Rehabilitation of People with Spinal Cord Injury*, 3rd ed.; Arbuckle Academic Publishing: Whitinsville, MA, USA, 2013.

3. *National Spinal Cord Injury Statistical Center, Facts and Figures at a Glance*; University of Alabama at Birmingham: Birmingham, AL, USA, 2018.

4. Seth, N.; Simmons, H.A.; Masood, F.; Graham, W.A.; Rosene, D.L.; Westmoreland, S.V.; Cummings, S.M.; Gwardjan, B.; Sejdic, E.; Hoggatt, A.F.; et al. Model of Traumatic Spinal Cord Injury for Evaluating Pharmacologic Treatments in Cynomolgus Macaques (*Macaca fasicularis*). *Comp. Med.* **2018**, *68*, 63–73.

5. Sledge, J.; Graham, W.A.; Westmoreland, S.; Sejdic, E.; Miller, A.; Hoggatt, A.; Nesathurai, S. Spinal cord injury models in non human primates: Are lesions created by sharp instruments relevant to human injuries? *Med. Hypotheses* **2013**, *81*, 747–748. [CrossRef]

6. Graham, W.A.; Rosene, D.L.; Westmoreland, S.; Miller, A.; Sejdic, E.; Nesathurai, S. Humane Non-Human Primate Model of Traumatic Spinal Cord Injury Utilizing Electromyography as a Measure of Impairment and Recovery. *Open J. Vet. Med.* **2013**, *3*, 86. [CrossRef]

7. Nesathurai, S.; Graham, W.A.; Mansfield, K.; Magill, D.; Sehgal, P.; Westmoreland, S.V.; Prusty, S.; Rosene, D.L.; Sledge, J.B. Model of traumatic spinal cord injury in *Macaca fascicularis*: Similarity of experimental lesions created by epidural catheter to human spinal cord injury. *J. Med. Primatol.* **2006**, *35*, 401–404. [CrossRef]

8. Nout, Y.S.; Rosenzweig, E.S.; Brock, J.H.; Strand, S.C.; Moseanko, R.; Hawbecker, S.; Zdunowski, S.; Nielson, J.L.; Roy, R.R.; Courtine, G.; et al. Animal models of neurologic disorders: A nonhuman primate model of spinal cord injury. *Neurotherapeutics* **2012**, *9*, 380–392. [CrossRef]

9. Rosenzweig, E.S.; Courtine, G.; Jindrich, D.L.; Brock, J.H.; Ferguson, A.R.; Strand, S.C.; Nout, Y.S.; Roy, R.R.; Miller, D.M.; Beattie, M.S.; et al. Extensive spontaneous plasticity of corticospinal projections after primate spinal cord injury. *Nat. Neurosci.* **2010**, *13*, 1505. [CrossRef]

10. Ojha, P.R. Tail Carriage and Dominance in The Rhesus Monkey, Macaca Mulatta. *Mammalia* **2009**, *38*, 163–170. [CrossRef]

11. Sharif-Alhoseini, M.; Khormali, M.; Rezaei, M.; Safdarian, M.; Hajighadery, A.; Khalatbari, M.M.; Meknatkhah, S.; Rezvan, M.; Chalangari, M.; Derakhshan, P.; et al. Animal models of spinal cord injury: A systematic review. *Spinal Cord* **2017**, *55*, 714. [CrossRef]

12. Calancie, B.; Molano, M.R.; Broton, J.G. EMG for assessing the recovery of voluntary movement after acute spinal cord injury in man. *Clin. Neurophysiol.* **2004**, *115*, 1748–1759. [CrossRef]

13. Lewko, J.P. Assessment of muscle electrical activity in spinal cord injury subjects during quiet standing. *Paraplegia* **1996**, *34*, 158–163. [CrossRef]

14. Wiegner, A.W.; Wierzbicka, M.M.; Davies, L.; Young, R.R. Discharge properties of single motor units in patients with spinal cord injuries. *Muscle Nerve* **1993**, *16*, 661–671. [CrossRef]

15. Shahani, B.T.; Wierzbicka, M.M.; Parker, S.W. Abnormal single motor unit behavior in the upper motor neuron syndrome. *Muscle Nerve* **1991**, *14*, 64–69. [CrossRef]

16. Capogrosso, M.; Milekovic, T.; Borton, D.; Wagner, F.; Moraud, E.M.; Mignardot, J.B.; Buse, N.; Gandar, J.; Barraud, Q.; Xing, D.; et al. A brainspine interface alleviating gait deficits after spinal cord injury in primates. *Nature* **2016**, *539*, 284–288. [CrossRef]

17. Karlsson, S.; Akay, M. Time-frequency analysis of myoelectric signals during dynamic contractions: A comparative study. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 228–238. [CrossRef]

18. Phinyomark, A.; Nuidod, A.; Phukpattaranont, P.; Limsakul, C. Feature Extraction and Reduction of Wavelet Transform Coefficients for EMG Pattern Classification. *Elektron. Elektrotechnika* **2012**, *122*, 27–32. [CrossRef]

19. Ranniger, C.U.; Akin, D.L. EMG mean power frequency determination using wavelet analysis. In Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. "Magnificent Milestones and Emerging Opportunities in Medical Engineering" (Cat. No.97CH36136), Chicago, IL, USA, 30 October–2 November 1997; Volume 4, pp. 1589–1592.

20. Zhou, P.; Rymer, W.Z. MUAP Number Estimates in Surface EMG: Template-Matching Methods and Their Performance Boundaries. *Ann. Biomed. Eng.* **2004**, *32*, 1007–1015. [CrossRef]

21. Hramov, A.E.; Koronovskii, A.A.; Makarov, V.A.; Pavlov, A.N.; Sitnikova, E. *Wavelets in Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2015.

22. Wang, G.; Zhang, Y.; Wang, J. The analysis of surface EMG signals with the wavelet-based correlation dimension method. *Comput. Math. Methods Med.* **2014**, *2014*, 284308. [CrossRef]

23. Hostens, I.; Seghers, J.; Spaepen, A.; Ramon, H. Validation of the wavelet spectral estimation technique in biceps brachii and brachioradialis fatigue assessment during prolonged low-level static and dynamic contractions. *J. Electromyogr. Kinesiol.* **2004**, *14*, 205–215. [CrossRef]

24. So, R.C.H.; Ng, J.K.-F.; Lam, R.W.K.; Lo, C.K.K.; Ng, G.Y.F. EMG wavelet analysis of quadriceps muscle during repeated knee extension movement. *Med. Sci. Sports Exerc.* **2009**, *41*, 788–796. [CrossRef]

25. González-Izal, M.; Rodríguez-Carreño, I.; Malanda, A.; Mallor-Giménez, F.; Navarro-Amézqueta, I.; Gorostiaga, E.M.; Izquierdo, M. sEMG wavelet-based indices predicts muscle power loss during dynamic contractions. *J. Electromyogr. Kinesiol.* **2010**, *20*, 1097–1106. [CrossRef]

26. Yamada, R.; Ushiba, J.; Tomita, Y.; Masakado, Y. Decomposition of electromyographic signal by principal component analysis of wavelet coefficients. In Proceedings of the IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, Kyoto, Japan, 20–22 October 2003; pp. 118–119.

27. Fang, J.; Agarwal, G.C.; Shahani, B.T. Decomposition of EMG signal by wavelet spectrum matching. In Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 30 October–2 November 1997; Volume 3, pp. 1253–1256.

28. Merlo, A.; Farina, D.; Merletti, R. A fast and reliable technique for muscle activity detection from surface EMG signals. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 316–323. [CrossRef]

29. Sharma, S. Wavelet analysis based feature extraction for pattern classification from Single channel acquired EMG signal. *Elixir Online J.* **2012**, *50*, 0320-1.

30. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. Application of wavelet analysis in EMG feature extraction for pattern classification. *Meas. Sci. Rev.* **2011**, *11*, 45–52. [CrossRef]

31. Salvador, J. Application of the Wavelet Transform for EMG M-Wave Pattern Recognition. Ph.D. Thesis, McMaster University, Hamilton, ON, Canada, 2006.

32. Sharma, T.; Veer, K. EMG classification using wavelet functions to determine muscle contraction. *J. Med. Eng. Technol.* **2016**, *40*, 99–105. [CrossRef]

33. Vetter, R.; Schild, J.; Kuhn, A.; Radlinger, L. Discrimination of Healthy and Post-partum Subjects using Wavelet Filterbank and Auto-regressive Modelling. In Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, Lisbon, Portugal, 12–15 January 2015.

34. Wakeling, J.M.; Lee, S.S.M.; Arnold, A.S.; Miara, M.D.; Biewener, A.A. A Muscle's Force Depends on the Recruitment Patterns of Its Fibers. *Ann. Biomed. Eng.* **2012**, *40*, 1708–1720. [CrossRef]

35. Parsaei, H. *EMG Signal Decomposition Using Motor Unit Potential Train Validity*; University of Waterloo: Waterloo, ON, Canada, 2011.

36. Farina, D.; Crosetti, A.; Merletti, R. A model for the generation of synthetic intramuscular EMG signals to test decomposition algorithms. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 66–77. [CrossRef]

37. Stashuk, D.W.; Farina, D.; Søgaard, K. Decomposition of intramuscular EMG signals. In *Electromyography: Physiology, Engineering, and Noninvasive Applications*; Wiley-Interscience: Hoboken, NJ, USA; IEEE Press: Piscataway, NJ, USA, 2005; pp. 47–80.

38. Negro, F.; Muceli, S.; Castronovo, A.M.; Holobar, A.; Farina, D. Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation. *J. Neural Eng.* **2016**, *13*, 026027. [CrossRef]

39. Nawab, S.H.; Wotiz, R.P.; de Luca, C.J. Decomposition of indwelling EMG signals. *J. Appl. Physiol.* **2008**, *105*, 700–710. [CrossRef]

40. Sledge, J.; Mahadevappa, K.; Stacey, P.; Graham, A.; Nesathurai, S. A Speculative Pharmaceutical Cocktail to Treat Spinal Cord Injury. *Am. J. Phys. Med. Rehabil.* **2016**, *95*, e108–e110. [CrossRef]

41. Luca, C.J.D. Control properties of motor units. *J. Exp. Biol.* **1985**, *115*, 125–136.

42. Messer, S.R.; Agzarian, J.; Abbott, D. Optimal wavelet denoising for phonocardiograms. *Microelectron. J.* **2001**, *32*, 931–941. [CrossRef]

43. Al-Qazzaz, N.K.; Ali, S.H.B.M.; Ahmad, S.A.; Islam, M.S.; Escudero, J. Selection of Mother Wavelet Functions for Multi-Channel EEG Signal Analysis during a Working Memory Task. *Sensors* **2015**, *15*, 29015–29035. [CrossRef]

44. Luca, C.J.D. Physiology and Mathematics of Myoelectric Signals. *IEEE Trans. Biomed. Eng.* **1979**, *BME-26*, 313–325. [CrossRef]

45. Al-Qazzaz, N.K.; Ali, S.H.M.; Islam, S.; Ahmad, S.A.; Escudero, J. EEG Wavelet Spectral Analysis During a Working Memory Tasks in Stroke-Related Mild Cognitive Impairment Patients. In *International Conference for Innovation in Biomedical Engineering and Life Sciences*; Springer: Singapore, 2015; pp. 82–85.

46. Mammone, N.; Foresta, F.L.; Morabito, F.C. Automatic Artifact Rejection from Multichannel Scalp EEG by Wavelet ICA. *IEEE Sens. J.* **2012**, *12*, 533–542. [CrossRef]

47. Inuso, G.; Foresta, F.L.; Mammone, N.; Morabito, F.C. Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings. In Proceedings of the 2007 International Joint Conference on Neural Networks, Orlando, FL, USA, 12–17 August 2007; pp. 1524–1529.

48. Burhan, N.; Kasno, M.; Ghazali, R. Feature extraction of surface electromyography (sEMG) and signal processing technique in wavelet transform: A review. In Proceedings of the 2016 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 22 October 2016; pp. 141–146.

49. Hu, X.; Wang, Z.; Ren, X. Classification of surface EMG signal using relative wavelet packet energy. *Comput. Methods Programs Biomed.* **2005**, *79*, 189–195. [CrossRef]

50. Ko, K.-E.; Yang, H.-C.; Sim, K.-B. Emotion recognition using EEG signals with relative power values and Bayesian network. *Int. J. Control Autom. Syst.* **2009**, *7*, 865. [CrossRef]

51. Sadeghi, H.; Allard, P.; Prince, F.; Labelle, H. Symmetry and limb dominance in able-bodied gait: A review. *Gait Posture* **2000**, *12*, 34–45. [CrossRef]

52. De Carvalho, M.; Eisen, A.; Krieger, C.; Swash, M. Motoneuron firing in amyotrophic lateral sclerosis (ALS). *Front. Hum. Neurosci.* **2014**, *8*, 719. [CrossRef]

sensors

Article

Reexamination of Accelerometer Calibration with Energy Expenditure as Criterion: VO$_{2\text{net}}$ Instead of MET for Age-Equivalent Physical Activity Intensity

Daniel Arvidsson [1,*], Jonatan Fridolfsson [1], Christoph Buck [2], Örjan Ekblom [3],
Elin Ekblom-Bak [3], Lauren Lissner [4], Monica Hunsberger [4] and Mats Börjesson [1,5,6]

[1] Center for Health and Performance, Department of Food and Nutrition, and Sport Science, University of Gothenburg, 40530 Gothenburg, Sweden

[2] Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology—BIPS, 28359 Bremen, Germany

[3] Åstrand Laboratory of Work Physiology, Swedish School of Sport and Health Sciences, 11486 Stockholm, Sweden

[4] Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, 40530 Gothenburg, Sweden

[5] Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, 40530 Gothenburg, Sweden

[6] Sahlgrenska University Hospital/Östra, 41345 Gothenburg, Sweden

* Correspondence: daniel.arvidsson@gu.se

Abstract: Accelerometer calibration for physical activity (PA) intensity is commonly performed using Metabolic Equivalent of Task (MET) as criterion. However, MET is not an age-equivalent measure of PA intensity, which limits the use of MET-calibrated accelerometers for age-related PA investigations. We investigated calibration using VO$_{2\text{net}}$ (VO$_{2\text{gross}}$ − VO$_{2\text{stand}}$; mL·min^{-1}·kg^{-1}) as criterion compared to MET (VO$_{2\text{gross}}$/VO$_{2\text{rest}}$) and the effect on assessment of free-living PA in children, adolescents and adults. Oxygen consumption and hip/thigh accelerometer data were collected during rest, stand and treadmill walk and run. Equivalent speed (Speed$_{\text{eq}}$) was used as indicator of the absolute speed (Speed$_{\text{abs}}$) performed with the same effort in individuals of different body size/age. The results showed that VO$_{2\text{net}}$ was higher in younger age-groups for Speed$_{\text{abs}}$, but was similar in the three age-groups for Speed$_{\text{eq}}$. MET was lower in younger age-groups for both Speed$_{\text{abs}}$ and Speed$_{\text{eq}}$. The same VO$_{2\text{net}}$-values respective MET-values were applied to all age-groups to develop accelerometer PA intensity cut-points. Free-living moderate-and-vigorous PA was 216, 115, 74 and 71 min/d in children, adolescents, younger and older adults with VO$_{2\text{net}}$-calibration, but 140, 83, 74 and 41 min/d with MET-calibration, respectively. In conclusion, VO$_{2\text{net}}$ calibration of accelerometers may provide age-equivalent measures of PA intensity/effort for more accurate age-related investigations of PA in epidemiological research.

Keywords: physical activity; accelerometer; VO$_2$; calibration; MET; VO$_{2\text{net}}$; speed; equivalent speed; free-living; children; adolescents; adults

1. Introduction

The use of accelerometers to measure physical activity (PA) in research have expanded tremendously since the millennium shift. An essential component of PA is the intensity performed, as PA intensity is related to the health benefits achieved [1]. Therefore, it is crucial that accelerometers provide accurate measures of PA intensity, which commonly involves determining time spent in different intensity levels such as light (LPA), moderate (MPA) and vigorous (VPA) physical activity.

These PA intensity measures have been used in cross-sectional and longitudinal epidemiological studies to investigate, for example, the age-related decline in PA from childhood into adulthood and associated factors, in an attempt to identify risk groups for future cardiometabolic disease [2–4].

Researchers working with development of PA intensity measures from accelerometers have mainly used energy expenditure determined from measured oxygen consumption as criterion method. The most common criterion measure of absolute PA intensity is Metabolic Equivalent of Task (MET), calculated as the quotient of gross energy expenditure and resting energy expenditure. However, MET is not directly comparable between children and adults, as the contribution of resting energy expenditure to gross energy expenditure changes from childhood into adulthood [5]. The resting energy expenditure decreases from about 2.0 kcal$\cdot$kg$^{-1}\cdot$h^{-1} at the age of 5 years to 1.0 kcal$\cdot$kg$^{-1}\cdot$h^{-1} at the age of 18 years. When children 8–12 and adolescents 15–18 years old walk at a speed of 5.6 km$\cdot$h^{-1} they reach a MET-value of 4.3 and 4.5, respectively, while running at 8.0 km$\cdot$h^{-1} contributes to a MET-value of 6.7 and 8.1, respectively [6]. When adults walk and run at these speeds, they reach the MET-value of about 5.0 and 9.0, respectively [7]. Importantly, despite their lower MET-values, the younger individuals would actually exert higher degree of effort with higher oxygen consumption per kg body weight compared to the older individuals [8,9] as they are moving with higher step frequency [10,11].

It is common in calibration of accelerometers to use 3.0 and 6.0 METs as cut-points for MPA and VPA, respectively, in both children and adults [12], but even higher values have been implemented in children, i.e., 4.0 and 7.0 METs [2,3,13], with the argument that it adjusts for their higher resting energy expenditure. In fact, we put higher requirements on children than on adults to achieve MPA and VPA. Therefore, the current calibration procedure for absolute PA intensity using MET cannot be used to compare the PA in different age-groups. Theoretically, this may contribute to underestimation of the PA in children relative to adults and incorrect conclusions concerning the decline in PA from childhood into adulthood and identification of individuals being physically inactive.

Attempts have been made to identify alternative criterion measures of activity intensity from oxygen uptake [14]. One such measure is the mass-specific net oxygen consumption (VO$_{2net}$, mL$\cdot$kg$^{-1}\cdot$min^{-1}), subtracting resting oxygen uptake (VO$_{2rest}$) from gross oxygen uptake (VO$_{2gross}$). However, this measure is not equivalent by age, as it decreases from childhood into adulthood [14]. In biomechanical research VO$_{2net}$ is instead calculated by subtracting standing oxygen uptake (VO$_{2stand}$) from VO$_{2gross}$ [10,11]. This definition considers the energy requirement of dynamic movement only and therefore matches the dynamic acceleration captured by accelerometers. The definition of subtracting VO$_{2rest}$ from VO$_{2gross}$ would also include oxygen consumption for muscular support during movement which is not captured by the accelerometer. The literature has not clearly demonstrated whether VO$_{2net}$ (VO$_{2gross}$ − VO$_{2stand}$) as a measure of activity intensity is equivalent by age [8–11,15]. It would make sense that children have higher VO$_{2net}$ than adults at the same absolute speed (Speed$_{abs}$) [8,9], as the they are moving with a higher degree of effort due to the higher step frequency [10,11]. Instead, the VO$_{2net}$ is more similar at the kinematically equivalent speed (Speed$_{eq}$) [8,9,15], which reflects the Speed$_{abs}$ two individuals of different body size are moving at with the same degree of effort, i.e., shorter individuals would move at a slower Speed$_{abs}$ than taller individuals. Therefore, further investigations of the VO$_{2net}$ as a measure of activity intensity equivalent by age is warranted, relating it to the Speed$_{eq}$ to determine whether it reflects a similar effort performed by individuals of different body size and age. If it does, it can be used to calibrate accelerometer PA intensity measures equivalent by age to be implemented into epidemiological research for age-related investigations of PA.

The ActiGraph counts has been the most common accelerometer-based measure of PA activity intensity in research, mostly from hip recordings [12]. However, the processing of raw acceleration to counts demonstrates a measurement error of increasing attenuation of the counts with increasing PA intensity, which is greater in children than in adults [16,17]. This error was explained by the use of a too narrow band-pass filter. When the filter was expanded to a more optimal level, children and adults produced similar accelerometer outcomes from hip recordings for the same absolute movement

speed. This indicates that an accelerometer placed at the hip would capture total mechanical work in both children and adults, which would be identical at the same Speed$_{abs}$ [11,17]. The measurement error with the ActiGraph counts may have contributed to the confusion of the METs in the calibration of accelerometers, where children demonstrated both lower METs and lower counts than adults at the same Speed$_{abs}$. In fact, children should have demonstrated lower METs but similar counts at the same Speed$_{abs}$. These different outcomes would have contributed to different calibration equations and cut-points for PA intensity. While the comparability of accelerometer measures of PA intensity by age has been investigated for the hip placement, less is known about the thigh placement [17–19]. The thigh placement offers additional possibilities for assessment of sedentary time (sitting, standing), biking (i.e., cadence) as well as non-wear time.

Therefore, the aim of the present study was to investigate the calibration of hip and thigh accelerometers in children and adults using VO$_{2net}$ (VO$_{2gross}$ − VO$_{2stand}$) as criterion for absolute PA intensity and the effect on assessment of free-living PA. With the application of a novel calibration procedure, this study may contribute to improve the assessment of PA intensity in children and to establish a measure of PA intensity equivalent by age.

2. Materials and Methods

2.1. Study Design

Two sub-studies were performed:

1. A calibration study including sit, stand, walk and run on a treadmill wearing accelerometers at the hip and thigh and with measured oxygen consumption to develop accelerometer cut-points for PA intensity.
2. Investigation of free-living PA in children, adolescents and adults comparing application of VO$_{2net}$ (VO$_{2gross}$ − VO$_{2stand}$) cut-points to standard MET cut-points.

2.2. Samples

Participants in the first sub-study were 10 children 9–11 years old, 10 adolescents 14–16 years old and 10 adults 23–44 years old. They were recruited among staff and students and their children by written and verbal information at the Department of Food and Nutrition, and Sport Science in Sweden and among their friends and children [17]. All participants provided informed consent and the protocol was approved by the ethical committee at the University of Gothenburg.

The second sub-study used Swedish data collected in the European I.Family study (EC: FP7, No. 266044) [20], which is an extension of the IDEFICS study (EC: FP6, No. 016181) investigating children's health, as well as data collected in the LIV2013 study of adult's health in Sweden [21]. The I.Family sample consisted of 417 Swedish children and adolescents (50% female) with a mean (SD) age of 10.9 (2.4) years, the range being 4–16 years. The LIV2013 study consisted of 735 Swedish adults (24% female) with a mean (SD) age of 48.4 years (11.6), the range being 21–67 years. The participants in these studies have provided informed consents and the studies have been approved by the ethical committees.

2.3. Protocols, Measures and Data Analysis

2.3.1. Calibration Study

The calibration study was performed in the biomechanical–physiological lab at the Center for Health and Performance, Gothenburg, Sweden. The participants were asked to refrain from food intake and strenuous PA 3–4 h prior to measurements. Body weight and height were measured and one Axivity AX3 triaxial accelerometer (Axivity Ltd., Newcastle upon Tyne, UK) was attached over the right hip in an elastic belt around the waist and another at the middle of the front side of the right thigh using medical tape. The accelerometers were set to record data at a sample rate of 100 Hz and an

acceleration range of ±8 g and the raw triaxial data was extracted with the OmGUI software (Axivity Ltd., Newcastle upon Tyne, UK). Oxygen consumption was measured using the stationary metabolic system Oxycon Pro (Jaeger, BD Corporation, Franklin Lakes, NJ, USA) where the participants breath through a face-mask with a turbine flow meter and expired air-sampler. Oxycon Pro has shown high accuracy when evaluated with the Douglas bag method [22]. The activities included consisted of seated rest in an armchair for 20 min to determine resting energy expenditure (REE) followed by 4 min standing, walking at 3, 4, 5 and 6 km·h^{-1} and running at 8 and 10 km·h^{-1} on a treadmill during 4 min at each speed without interruption, in order to reach steady-state [23]. VO$_2$ data was collected during all 20 min and respective 4 min activities.

The triaxial acceleration data were processed using our new algorithm to produce a measure of PA intensity expressed in mg, including a band-pass filter with the cut-point range of 0.29–10 Hz that was shown to include all relevant information to assess activity intensity and to minimize inclusion of noise [17,18]. However, acceleration data was down-sampled to 30 Hz and truncated to ±6 g before processing to match the corresponding ranges of the ActiGraph GT3X+ accelerometer used in the free-living measurements. VO$_2$ data from the 20 min of resting was filtered with a moving average filter with a window size of 2 min and the minimum value was considered the individual REE. One minute of data (accelerometer, VO$_2$) captured between duration 2:45 and 3:45 at standing and at each treadmill speed was used for calibration. VO$_{2net}$ was calculated by subtracting VO$_{2stand}$ from VO$_{2gross}$ divided by body weight (mL·min^{-1}·kg^{-1}), while MET-values were calculated by the quotient of the VO$_{2gross}$ and VO$_{2rest}$.

In each age-group, the relationships between Speed$_{abs}$ and accelerometer output (mg), Speed$_{abs}$ and VO$_{2net}$ and Speed$_{abs}$ and MET were investigated, as well as the relationship between Speed$_{eq}$ and VO$_{2net}$ and Speed$_{eq}$ and MET. The Speed$_{eq}$ was calculated as Speed$_{eq}$ = V^2·g^{-1}·h^{-1} (V = Speed$_{abs}$ (m·s^{-1}), g = gravity (9.81 m·s^{-2}), h = body height (m)), and represent the Speed$_{abs}$ performed with similar kinematical effort in individuals of different body size [8,9,15]. Smoothing splines were fitted to the data from each age group, with the intercept forced to zero VO$_{2net}$ or 1 MET, and zero acceleration. Based on previous biomechanical and physiological research, we made the following assumptions:

- The three age-groups fall on the same regression line between Speed$_{abs}$ and hip accelerometer output, as the hip accelerometer represents total mechanical work which would be the same for the same Speed$_{abs}$ in the three age-groups [10,11,17].
- The three age-groups would have less similarity in the regression line between Speed$_{abs}$ and thigh accelerometer output, as the mechanical work of moving the limbs is higher in shorter than taller individuals [10,11,17].
- The three age-groups would fall on the same regression line between Speed$_{eq}$ and VO$_{2net}$, if VO$_{2net}$ represent the same effort in the age-groups.
- The three age-groups would fall on different regression lines between Speed$_{eq}$ and MET, if MET does not represent the same effort in the age-groups.
- The three age-groups would fall on different regression lines between VO$_{2net}$ and accelerometer output, if the same VO$_{2net}$ in the age-groups would mean the same Speed$_{eq}$ and therefore the same effort, but at the same time lower Speed$_{abs}$ in shorter individuals and correspondingly lower mechanical work and accelerometer output.

MET cut-points at 1.5, 3.0, 6.0 and 9.0 were implemented to represent LPA, MPA, VPA and very vigorous PA (VVPA), respectively, which is in line with previous literature [12]. To calibrate accelerometer data to VO$_{2net}$, the first step was to determine the VO$_{2net}$ corresponding to 3.0, 6.0 and 9.0 METs in adults by linear regression. The next step was to apply the same VO$_{2net}$ cut-points (and therefore the same effort) to all three age-groups to calibrate the corresponding accelerometer mg cut-points based on fitted smoothing splines. For comparison, accelerometer mg cut-points were also determined using the traditional 1.5, 3.0, 6.0 and 9.0 METs in all three age-groups.

2.3.2. Free-Living Study

In both the I.Family study and the LIV2013 study the ActiGraph GT3X+ (ActiGraph, Pensacola, FL, USA) was used to collect acceleration data and was worn over the right hip in an elastic belt around the waist. The participants were instructed to wear the accelerometer for seven days and to remove it during sleep and during water-based activities. The accelerometer was set to record acceleration data at 30 Hz sampling rate with an acceleration range of ±6 g and idle sleep mode were enabled. The raw triaxial data was extracted according to available file specifications [24].

Raw triaxial data was processed using our new algorithm to produce a measure of PA intensity expressed in mg including a band-pass filter with the cut-point range of 0.29–10 Hz [17,18]. The epoch-length was set to 3 s to capture children's activities as well as intermittent activities in adults [18]. Valid days with at least 12 h were included in the analyses. Samples in which the sensor status was idle according to the idle sleep-mode were considered non-wear time. More subjects in the adult group wore the accelerometer during sleep compared to the children; therefore, the acceleration recorded between 23:00–06:00 was not included in the analysis. Our previous research has confirmed that processing of GT3X+ data and AX3 data generate similar output [16].

Accelerometer mg cut-points for LPA, MPA, VPA and VVPA from the VO_{2net}-based calibration and the standard MET-based calibration were applied. The participants were divided into four age-categories: Children (<13 years, n = 321 (50% female), adolescents (13–16 years, n = 96 (50% females)), younger adults (<50 years, n = 366 (75% female)) and older adults (≥50 years, n = 369 (68% female)). The time distribution in the intensity categories was compared between the age-categories with the VO_{2net}-based calibration versus the MET-based calibration. As the VO_{2net}-based calibration would imply lower $Speed_{abs}$ and mechanical work in children and adolescents compared to the MET-based calibration, and consequently lower accelerometer mg cut-points, we assume more free-living PA in these age groups applying the VO_{2net}-based accelerometer mg cut-points.

All data processing and analyses in the two sub-studies were performed in MATLAB 2018b (MathWorks, Natick, MA, USA).

3. Results

3.1. Calibration Study

Figure 1 presents the relationship between $Speed_{abs}$ and the accelerometer output. It demonstrates that the three age-groups fall approximately on the same regression line for hip data, while children show higher values compared to the other age-groups for the thigh placement. These results confirm our assumptions of similar mechanical work captured with the hip placement and less similarity with the thigh placement.

Figure 2 present the relationship between $Speed_{abs}$ and VO_{2net} and between $Speed_{eq}$ and VO_{2net}, respectively. It shows that while children have higher VO_{2net} than the other age-groups for the same $Speed_{abs}$, there was approximately one common regression line with the $Speed_{eq}$. VO_{2net} may therefore be considered a measure of similar effort in different age-groups according to our assumption.

Figure 3 display the relationship between $Speed_{abs}$ and MET and between $Speed_{eq}$ and MET, respectively. The MET-values differ between the three age-groups for both $Speed_{abs}$ and $Speed_{eq}$, which confirms the assumption that MET is not a measure of effort comparable between age-groups.
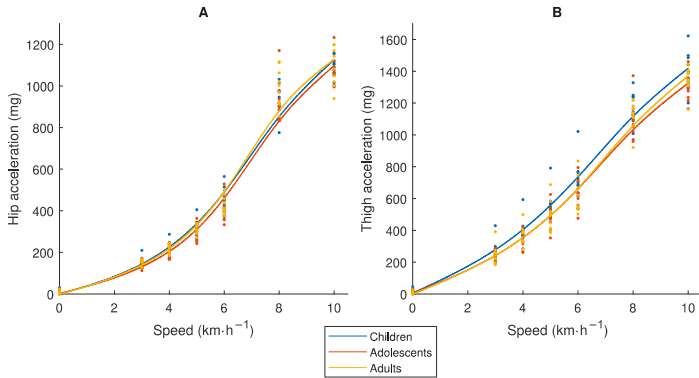
Figure 1. Relationship between absolute speed (Speed$_{abs}$) and the accelerometer measure of physical activity (PA) intensity (mg) for (**A**) hip and (**B**) thigh placement, in children 9–11, adolescents 14–16 and adults 23–44 years old.

Figure 2. Relationship between (**A**) Speed$_{abs}$ or (**B**) equivalent speed (Speed$_{eq}$) and net oxygen consumption (VO$_{2net}$), in children 9–11, adolescents 14–16 and adults 23–44 years old.
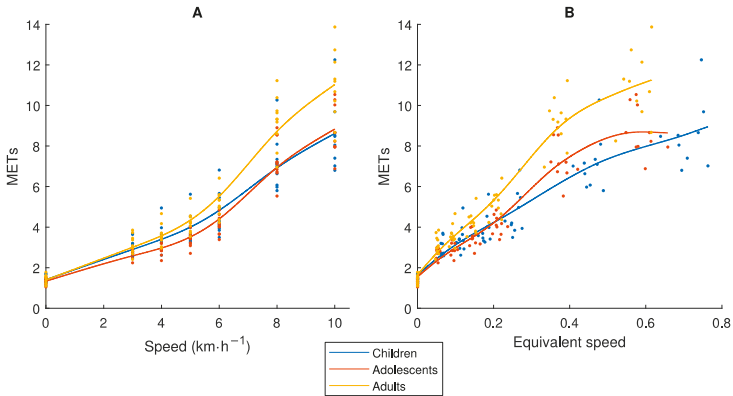
Figure 3. Relationship between (**A**) Speed$_{abs}$ or (**B**) Speed$_{eq}$ and Metabolic Equivalent of Task (MET), in children 9–11, adolescents 14–16 and adults 23–44 years old.

Figure 4 presents the relationship between MET and the corresponding VO_{2net}. It is the starting-point to determine the VO_{2net} cut-points for LPA (1.5 METs), MPA (3.0 METs), VPA (6.0 METs) and VVPA (9.0 METs) based on adult data. If the same MET-values had been applied to the children and adolescent groups, considerable higher VO_{2net}-values (and effort) had been used for accelerometer calibration in these age-groups.
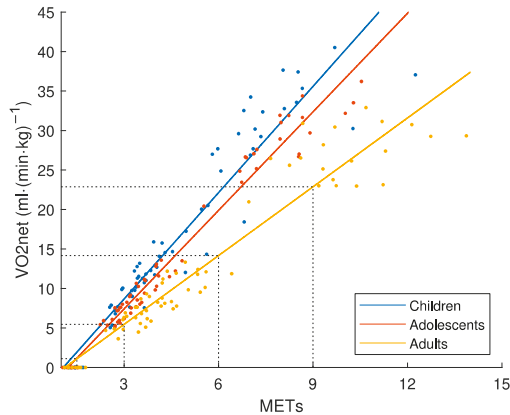
Figure 4. Definition of the common VO_{2net} cut-points representing light PA (LPA), moderate PA (MPA), vigorous PA (VPA) and very vigorous PA (VVPA) based on adult MET values, in children 9–11, adolescents 14–16 and adults 23–44 years old.

The next step was to apply the common VO_{2net} cut-points to all three age-groups to determine the corresponding accelerometer mg values (Figure 5). The three age-groups achieved different regression lines, with the lowest accelerometer cut-points observed in children. Hence, when moving with the same effort, the three age-groups move at different $Speed_{abs}$ and consequently do different mechanical work. Still, there were some overlapping between the age-groups.
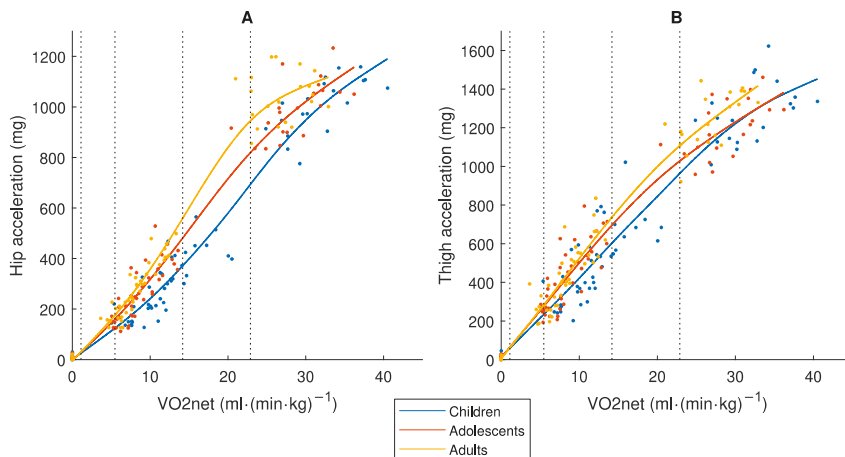
Figure 5. Calibration of accelerometer measures of absolute PA intensity (mg) for (**A**) hip and (**B**) thigh placement in children 9–11, adolescents 14–16 and adults 23–44 years old, using the common VO_{2net} cut-points defined and presented in Figure 4.

In contrast, when MET-based cut-points were applied, the adults achieved the lowest accelerometer values (Figure 6). Table 1 provides the numerical values of the VO_{2net}-based and the MET-based cut-points for LPA, MPA, VPA and VVPA for the two body placements in the three age-groups, expressed in mg.
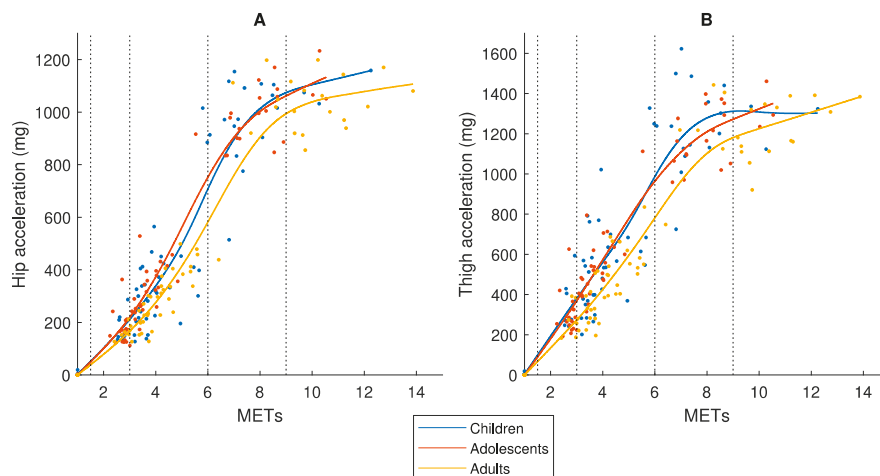
Figure 6. Calibration of the accelerometer measure of absolute PA intensity (mg) for (**A**) hip and (**B**) thigh placement in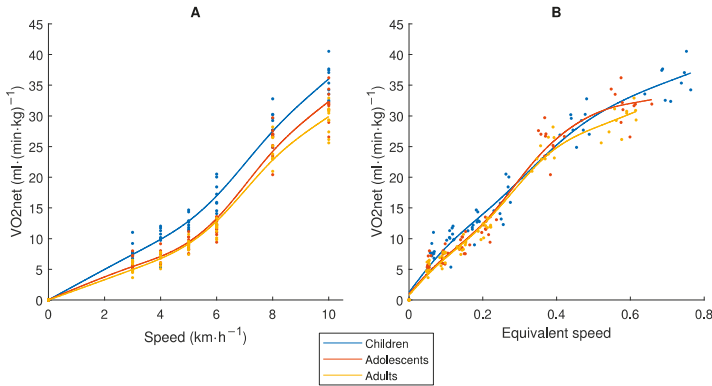 children 9–11, adolescents 14–16 and adults 23–44 years old, using the standard MET cut-points (1.5, 3.0, 6.0, 9.0) in all three groups.

Table 1. Accelerometer cut-points (mg) by body placement, age-group and calibration method.

Unit: mg	Children		Adolescents		Adults	
	VO_{2net}	MET	VO_{2net}	MET	VO_{2net}	MET
Hip placement						
Light	29	52	32	50	39	39
Moderate	124	214	157	219	167	167
Vigorous	368	704	482	753	582	582
Very vigorous	695	1075	830	1062	994	994
	$R^2 = 0.96$	$R^2 = 0.96$	$R^2 = 0.97$	$R^2 = 0.98$	$R^2 = 0.98$	$R^2 = 0.98$
Thigh placement						
Light	60	99	63	90	67	67
Moderate	234	380	274	368	273	273
Vigorous	603	987	700	964	782	782
Very vigorous	964	1312	1034	1272	1181	1181
	$R^2 = 0.93$	$R^2 = 0.95$	$R^2 = 0.95$	$R^2 = 0.98$	$R^2 = 0.98$	$R^2 = 0.98$

Children 9–11, adolescents 14–16 and adults 23–44 years old. VO_{2net} cut-points (mL·min^{-1}·kg^{-1}): LPA 1.1, MPA 5.5, VPA 14.2 and VVPA 22.9. MET cut-points (METs): LPA 1.5, MPA 3.0, VPA 6.0 and VVPA 9.0. Explained variation (R^2) is presented for the fit of the regression line between VO_{2net}/MET and accelerometer mg.

3.2. Free-Living Study

The results from the free-living study shows that when applying the VO_{2net} calibration, there was a considerable increase in the time being physically active in children and adolescents compared with application of the standard MET calibration (Table 2). For example, the time spent in moderate-and-vigorous PA (MVPA = MPA + VPA + VVPA) accounted for 13.8% of the wear-time in children and 8.1% in adolescents with the MET calibration, but increased to 21.2% and 11.2%, respectively, with the VO_{2net} calibration. Consequently, the difference in MVPA between children and adolescents versus adults increases. As an example, with the standard MET calibration the difference

in MVPA between children and younger adults is 6.5% units, while with the VO_{2net} calibration it is 13.9% units. In daily minutes, the difference corresponds to 66 min and 142 min, respectively.

Table 2. Time spent daily (min and % of wear time) in different intensity levels by age-group and calibration method for the hip placement.

Mean (SD)	Children (n = 321)		Adolescents (n = 96)		Adults <50 (n = 366)		Adults ≥50 (n = 369)	
Calibration	Min	%	Min	%	Min	%	Min	%
VO_{2net}								
SED	661 (125)	64.8 (12.2)	786 (114)	77.1 (11.1)	826 (56)	81.0 (5.5)	821 (60)	80.5 (5.9)
LPA	142 (41)	14.0 (4.1)	119 (52)	11.7 (5.1)	119 (37)	11.7 (3.6)	128 (39)	12.5 (3.9)
MPA	154 (57)	15.1 (5.6)	99 (58)	9.7 (5.7)	71 (26)	7.0 (2.5)	69 (29)	6.8 (2.8)
VPA	47 (37)	4.6 (3.6)	12 (12)	1.1 (1.2)	2 (4)	0.2 (0.4)	2 (3)	0.2 (0.3)
VVPA	15 (13)	1.5 (1.3)	4 (4)	0.4 (0.4)	1 (3)	0.1 (0.3)	0 (1)	0.0 (0.1)
MET								
SED	710 (117)	69.6 (11.5)	814 (104)	77.8 (10.2)	826 (56)	81.0 (5.5)	821 (60)	80.5 (5.9)
LPA	170 (51)	16.7 (5.0)	123 (56)	12.0 (5.4)	119 (37)	11.7 (3.6)	128 (39)	12.5 (3.9)
MPA	125 (70)	12.3 (6.9)	78 (54)	7.6 (5.3)	71 (26)	7.0 (2.5)	69 (29)	6.8 (2.8)
VPA	10 (8)	1.0 (0.7)	3 (3)	0.3 (0.3)	2 (4)	0.2 (0.4)	2 (3)	0.2 (0.3)
VVPA	5 (6)	0.5 (0.6)	2 (2)	0.2 (0.2)	1 (3)	0.1 (0.3)	0 (1)	0.0 (0.1)

As the VO_{2net} cut-points are determined from the MET-values for the different intensity levels in adults (i.e., 1.5, 3.0, 6.0 and 9.0), there will be no difference in the accelerometer cut-points in adults applying VO_{2net} calibration vs. MET calibration. Consequently, the time spent in the different intensity levels does not differ between VO_{2net} calibration vs. MET calibration in adults. Children include the age-range 4–12 years, adolescents 13–16 years, and the adult group was divided into younger adults (<50 years, range 21–49 years) and older adults (≥50 years, range 50–67 years).

4. Discussion

The novelty of this study is to propose a criterion method (VO_{2net}, $mL \cdot kg^{-1} \cdot min^{-1}$) to calibrate accelerometer PA intensity measures equivalent by age based on data collected at the hip and at the thigh. It was applied to our improved processing of acceleration data [16–19] to generate new cut-points for LPA, MPA, VPA and VVPA in children, adolescents and adults. These outcomes were supported by the finding that the relationship between $Speed_{eq}$ and VO_{2net} ($VO_{2gross} - VO_{2stand}$) was similar in children, adolescents and adults, indicating that VO_{2net} is a measure of similar effort in the three age-groups, in contrast to the more commonly used MET in accelerometer calibration which was not a measure of PA intensity equivalent by age. When our new cut-points for PA intensity levels were applied to free-living acceleration data from the hip, the difference in PA between children, adolescents and adults increased compared to the application of the standard MET-based cut-points. The importance of the findings in this study is to allow the possibility to directly relate the PA intensity level between age-groups and thereby more accurately investigate the age-related changes in PA from childhood into adulthood and associated factors. A larger decline in PA from childhood into adults may be expected compared to what is anticipated from previous research.

Two individuals of different body size perform an activity with different movement patterns. A short individual takes shorter steps with smaller acceleration amplitude but at higher frequency, while the taller individual takes longer steps with larger acceleration amplitude but at lower frequency [10,11,17]. Biomechanical theory demonstrates that the total mass-specific mechanical work is the same in short and tall individuals moving at the same $Speed_{abs}$, but that the shorter individual generates more internal work related to the higher frequency of moving the limbs while the taller individual generates more external work related to the movement of the center of mass [10,11]. Consequently, one would expect an accelerometer placed at the hip to generate higher values in the taller individual while the opposite would occur with the placement on the thigh. Interestingly, our results (Figure 1A) show that the hip placement generates similar accelerometer output for the same $Speed_{abs}$ in children, adolescents and adults, indicating that all the acceleration signals captured represent total mechanical work. Hence, an accelerometer placed at the hip can be used to compare

the total mechanical work performed between age-groups. Instead, the thigh placement generated higher accelerometer output in the children compared to the other groups (Figure 1B), indicating that this placement captures more internal work which is not comparable between age-groups. In contrast to the results herein, the original ActiGraph counts from hip recordings are lower in children than in adults for the same Speed$_{abs}$ [25,26], while the opposite occur with the Euclidian Norm Minus One (ENMO) accelerometer output [27]. Consequently, none of these methods provide measures of mechanical work equivalent by age. In the case of the ActiGraph counts, this age difference is caused by a well-defined processing error [16]. In the case of the ENMO accelerometer output, it may be a processing error less clearly defined. With the ENMO method, all negative accelerations are set to zero after subtracting 1 g from the vector magnitude [27]. This means that acceleration signals generated with larger amplitude but at lower frequency (as in adults) will be excluded to larger extent compared to acceleration signals with a lower amplitude but at a higher frequency (as in children). In contrast, the ActiGraph counts and the accelerometer method developed in our research group are aggregations of both positive and negative accelerations.

Even if an accelerometer placed at the hip would capture total mechanical work at the same Speed$_{abs}$, the activity is performed with different effort and energy cost in children compared to adults (Figure 2A) [8,9], and therefore has different physiological loads/health effects on the body. Consequently, an accelerometer output needs to be calibrated against a criterion measure of equivalent effort/load by age. The MET was developed to provide a criterion measure of absolute PA intensity. Our study showed that the MET was not equivalent by age as different values were achieved in children, adolescents and adults for the same Speed$_{abs}$ and Speed$_{eq}$ (Figure 3). If we compare the results in Figures 5 and 6, we clearly see the consequence of applying the MET-based accelerometer calibration: Higher accelerometer cut-points are set for younger individuals to reach MPA, VPA and VVPA, when they are actually performing the activity at these PA intensity levels with a higher degree of effort according to the Speed$_{eq}$ (Figure 3B) and with a higher energy cost according to the VO$_{2net}$ (Figure 4) compared to the older individuals. This calibration error will contribute to the underestimation of the PA in children relative to adults.

Alternative criterion measures of PA intensity equivalent by age have been investigated, for example mass-specific VO$_{2gross}$ (mL·kg^{-1}·min^{-1}), VO$_{2net}$ (VO$_{2gross}$ − VO$_{2rest}$, mL·kg^{-1}·min^{-1}) and VO2$_{allom}$ (mL·kg$^{-0.75}$·min^{-1}) [14]. None of the measures was optimal for the PA intensity range. The allometric scaling seems to work well for ambulatory activities and the MVPA intensity range [14] and has previously been proposed as an accurate criterion measures for accelerometer calibration [28]. We based our choice of VO$_{2net}$ (VO$_{2gross}$ − VO$_{2stand}$, mL·kg^{-1}·min^{-1}) on that this measure captures the dynamic movement only and therefore matches the dynamic acceleration captured by an accelerometer. Although, previous research has shown a remaining difference by age when this measure was related to Speed$_{eq}$. For example, at Speed$_{eq}$ = 0.3, the VO$_{2net}$ was about 230 and 200 mL·kg^{-1}·km^{-1} in children and adults, respectively [8]. We also observed some fluctuations in the difference between the age-groups across Speed$_{eq}$ range (Figure 2B), but the relevance of these small differences and those found in the study by McCann et al. [8] is unclear and requires further investigations. We still find our results to provide a strong indication of VO$_{2net}$ as a criterion measure of PA activity intensity equivalent by age and suitable for accelerometer calibrations.

A limitation with the VO$_{2net}$ measure is that there are no established PA intensity cut-points as for the MET measure (i.e., 3.0 and 6.0 METs). We based our calibration procedure on the adult MET-values and cut-points for LPA, MPA, VPA and VVPA (Figure 4), as the MET as a measure of absolute PA intensity was initially developed in adults [29], and translated them to the corresponding VO$_{2net}$. The same VO$_{2net}$ cut-points were thereafter applied to all three age-groups to represent PA intensity levels equivalent by age and used for the accelerometer calibration (Figure 5). As children, adolescents and adults produce the same total mechanical work at the same Speed$_{abs}$ [10,11], which is captured by an accelerometer at the hip (Figure 1A), and as VO$_{2net}$ differs between these age-groups at the same Speed$_{abs}$ (Figure 2A) due to different efforts related to step frequency [10,11] but is similar at the same

Speed$_{eq}$ (Figure 2B), these age-groups will present different regression lines between VO$_{2net}$ and the accelerometer output and consequently also different accelerometer cut-points (children < adolescents < adults). The somewhat smaller difference in the VO$_{2net}$-acceleration regression lines between the age-groups with the thigh placement (Figure 5) is explained by differences in the accelerometer output, which is supported by the biomechanical literature showing that smaller individuals produce more internal work [10,11] and consequently more thigh acceleration. Our calibration procedure requires knowledge about speed of movement. An alternative procedure embraced in calibration and validation studies, is to let the participants perform walking and running at a self-selected pace not involving the measurement of movement speed [13]. It is possible that the self-selected paces correspond to the same Speed$_{eq}$ in different age-groups. The accelerometer calibration research field has postulated that a variety of activity types (including intermittent) should be included in calibration protocols to provide calibration algorithms representative of people's daily activity pattern [30], not only walking and running. Still, that protocol would include activities where the steady-state is not easily attained in order to accurately assess the oxygen consumption of an activity [23], which is the case for protocols including only continuous walking and running as in the present study.

The contribution of our study can be exemplified based on two large epidemiological studies where the PA has been compared between children and adults. Cross-sectional analyses of accelerometer data from the National Health and Nutrition Examination Survey (NHANES) have shown that boys and girls 6–11 years old spend 95 and 75 min daily in MVPA, respectively, while men and women 40–49 years old attained 35 and 19 min, respectively [2]. Longitudinal analyses of accelerometer data in the European Youth Heart Study between ages 9 and 15 years old demonstrated a decline in MVPA from 100 to 52 min daily in Swedish boys and from 73 to 44 min daily in Swedish girls [4]. In another sample in this study, there was a decline between ages 15 and 21 years old from 68 to 58 min daily in Swedish males and from 46 to 40 min daily in Swedish females. These studies used MET-based accelerometer cut-points for MPA and VPA, applying similar (3.0 or 6.0 METs) or higher (4.0, 7.0 vs. 3.0, 6.0 METs) MET-values in children compared to adults. Consequently, higher demands were required from the children to reach MPA or VPA than in the adults. Further, the acceleration data was processed to the original ActiGraph counts with its limitations to assess PA intensity [16–19]. In addition, the aggregation of the acceleration data was performed into 60 s epochs, which would be insensitive to the movement pattern of children [1]. The application of our improved processing of accelerometer data [17,18] and calibration procedure presented herein together with the use of 3 s epochs would have demonstrated a larger decline in MVPA from childhood into adulthood. Consequently, there may have been an underestimation of the change in the risk of future cardiometabolic disease in those studies.

A further complication with the widely adopted 3.0 MET cut-point is that it may have been set too low to accurately represent MPA in the general population. The American College of Sports Medicine Position Stand classify 4.8–7.1 METs and 4.0–5.9 METs as MPA in young respective middle-aged adults [31]. Further, 3.0 METs was already reached at comfortable walking by all age-groups in the present study as well as in other studies of children and adults [6,7,13,14]. Consequently, people would be considered physically active too easily. This issue requires further discussion to be settled, but a higher cut-point for MPA might be considered in investigations of the general population.

The present study has several strengths and limitations. Raw acceleration data were processed with algorithms improving assessment of PA intensity [17,18]. Further, short 3 s epochs were applied to improve assessment of children's PA pattern [1] as well as intermittent PA. Calibration of activity intensity cut-points was performed on a treadmill with four minutes at each speed. Previous studies with other accelerometers have showed inconsistent results concerning the impact of setting (treadmill vs. ground) on accelerometer output [32,33]. Calibration performed with walking and running on a treadmill may not be applicable to other types of activities. As we have commented on earlier in the discussion section, using oxygen consumption as criterion measure of PA intensity requires steady-state of oxygen consumption data [23]. Still, the steady-state requirement may be a challenge as a large part of free-living PA does not fulfill it. The present calibration included less data points in

the VPA range (see Figure 5) compared to the MPA and VVPA ranges which may affect the precision of resultant regression lines and cut-points. The calibration sample age-range of 9–44 years did not fully represent the free-living sample age-range of 4–67 years, which may affect the results of the free-living stub-study.

5. Conclusions

We propose VO_{2net} ($VO_{2gross} - VO_{2stand}$, $mL \cdot kg^{-1} \cdot min^{-1}$), instead of MET ($VO_{2gross}/VO_{2rest}$), to be a criterion measure for accelerometer calibration in children, adolescents and adults. This conclusion is based on the finding that these age-groups showed similar relationship between $Speed_{eq}$ and VO_{2net}, and consequently VO_{2net} may be considered a measure of absolute PA intensity equivalent by age. The same similarity between the age-groups was not found for the relationship between $Speed_{eq}$ and MET. New VO_{2net}-based accelerometer cut-points were developed and when they were applied to free-living data, the difference in PA between children, adolescents and adults was increased compared to applying MET-based accelerometer cut-points. The implications of these findings are that PA can now be directly compared between age-groups in epidemiological research to investigate changes in PA behaviors from childhood into adulthood and associated factors related to risk for cardiometabolic disease. We contribute with new accelerometer cut-points for absolute PA intensity for the hip and thigh placement in children, adolescents and adults (Table 1), to be applicable to accelerometer output generated from raw triaxial acceleration data [17,18]. The algorithms for processing the raw acceleration data are available by request to the authors. Still, this is a small and explorative study and further research is needed to evaluate the robustness of the VO_{2net} calibration method proposed.

Author Contributions: Conceptualization, D.A. and J.F.; methodology, D.A. and J.F.; software, J.F.; formal analysis, D.A. and J.F.; visualization, J.F.; investigation, D.A., J.F., C.B., Ö.E., E.E.-B., L.L., M.H. and M.B.; resources, Ö.E., L.L. and M.B.; data curation, D.A. and J.F.; writing—original draft preparation, D.A. and J.F.; writing—review and editing, C.B., Ö.E., E.E.-B., L.L., M.H. and M.B.; funding acquisition, Ö.E., L.L. and M.B.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aadland, E.; Andersen, L.B.; Anderssen, S.A.; Resaland, G.K.; Kvalheim, O.M. Associations of volumes and patterns of physical activity with metabolic health in children: A multivariate pattern analysis approach. *Prev. Med.* **2018**, *115*, 12–18. [CrossRef] [PubMed]

2. Troiano, R.; Berrigan, D.; Dodd, K.; Mâsse, L.; Tilert, T.; McDowell, M. Physical activity in the United States measured by accelerometer. *Med. Sci. Sports Exerc.* **2008**, *40*, 181–188. [CrossRef] [PubMed]

3. Varma, V.R.; Dey, D.; Leroux, A.; Di, J.; Urbanek, J.; Xiao, L.; Zipunnikov, V. Re-evaluating the effect of age on physical activity over the lifespan. *Prev. Med.* **2017**, *101*, 102–108. [CrossRef]

4. Ortega, F.B.; Konstabel, K.; Pasquali, E.; Ruiz, J.R.; Hurtig-Wennlöf, A.; Mäestu, J.; Löf, M.; Harro, J.; Bellocco, R.; Labayen, I.; et al. Objectively measured physical activity and sedentary time during childhood, adolescence and young adulthood: A cohort study. *PLoS ONE* **2013**, *8*, e60871. [CrossRef]

5. Herrmann, S.D.; McMurray, R.G.; Kim, Y.; Willis, E.A.; Kang, M.; McCurdy, T. The influence of physical characteristics on the resting energy expenditure of youth: A meta-analysis. *Am. J. Hum. Biol.* **2017**, *29*, e22944. [CrossRef] [PubMed]

6. Harrell, J.S.; McMurray, R.G.; Baggett, C.D.; Pennell, M.L.; Pearce, P.F.; Bangdiwala, S.I. Energy costs of physical activities in children and adolescents. *Med. Sci. Sports Exerc.* **2005**, *37*, 329–336. [CrossRef] [PubMed]

7. Kozey, S.; Lyden, K.; Staudenmayer, J.; Freedson, P. Errors in MET estimates of physical activities using 3.5 mL x kg(-1) x min(-1) as the baseline oxygen consumption. *J. Phys. Act. Health* **2010**, *7*, 508–516. [CrossRef] [PubMed]

8. McCann, D.J.; Adams, W.C. The size-independent oxygen cost of running. *Med. Sci Sports Exerc.* **2003**, *35*, 1049–1056. [CrossRef]

9. McCann, D.J.; Adams, W.C. A dimensional paradigm for identifying the size-independent cost of walking. *Med. Sci Sports Exerc.* **2002**, *34*, 1009–1017. [CrossRef] [PubMed]

10. Schepens, B.; Bastien, G.J.; Heglund, N.C.; Willems, P.A. Mechanical work and muscular efficiency in walking children. *J. Exp. Biol.* **2004**, *207*, 587–596. [CrossRef] [PubMed]

11. Schepens, B.; Willems, P.A.; Cavagna, G.A.; Heglund, N.C. Mechanical power and efficiency in running children. *Pflugers Arch.* **2001**, *442*, 107–116. [CrossRef] [PubMed]

12. Migueles, J.H.; Cadenas-Sanchez, C.; Ekelund, U.; Delisle Nyström, C.; Mora-Gonzalez, J.; Löf, M.; Labayen, I.; Ruiz, J.R.; Ortega, F.B. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Med.* **2017**, *47*, 1821–1845. [CrossRef] [PubMed]

13. Trost, S.G.; Loprinzi, P.D.; Moore, R.; Pfeiffer, K.A. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med. Sci. Sports Exerc.* **2011**, *43*, 1360–1368. [CrossRef] [PubMed]

14. McMurray, R.G.; Butte, N.F.; Crouter, S.E.; Trost, S.G.; Pfeiffer, K.A.; Bassett, D.R.; Puyau, M.R.; Berrigan, D.; Watson, K.B.; Fulton, J.E.; et al. Exploring Metrics to Express Energy Expenditure of Physical Activity in Youth. *PLoS ONE* **2015**, *10*, e0130869. [CrossRef] [PubMed]

15. Weyand, P.G.; Smith, B.R.; Puyau, M.R.; Butte, N.F. The mass-specific energy cost of human walking is set by stature. *J. Exp. Biol.* **2010**, *213*, 3972–3979. [CrossRef]

16. Brønd, J.C.; Andersen, L.B.; Arvidsson, D. Generating ActiGraph Counts from Raw Acceleration Recorded by an Alternative Monitor. *Med. Sci. Sports Exerc.* **2017**, *49*, 2351–2360. [CrossRef]

17. Fridolfsson, J.; Börjesson, M.; Arvidsson, D. A Biomechanical Re-Examination of Physical Activity Measurement with Accelerometers. *Sensors* **2018**, *18*, 3399. [CrossRef]

18. Fridolfsson, J.; Börjesson, M.; Buck, C.; Ekblom, Ö.; Ekblom-Bak, E.; Hunsberger, M.; Lissner, L.; Arvidsson, D. Effects of frequency filtering on intensity and noise in accelerometer-based physical activity measurement. *Sensors* **2019**, *19*, 2186. [CrossRef]

19. Brønd, J.C.; Aadland, E.; Andersen, L.B.; Resaland, G.K.; Andersen, S.A.; Arvidsson, D. The ActiGraph counts processing and the assessment of vigorous activity. *Clin. Physiol. Funct. Imaging* **2019**, *39*, 276–283. [CrossRef]

20. Ahrens, W.; Siani, A.; Adan, R.; De Henauw, S.; Eiben, G.; Gwozdz, W.; Hebestreit, A.; Hunsberger, M.; Kaprio, J.; Krogh, V.; et al. Cohort Profile: The transition from childhood to adolescence in European children-how I.Family extends the IDEFICS cohort. *Int. J. Epidemiol.* **2017**, *46*, 1394–1395j. [CrossRef]

21. Olsson, G. Studies of Physical Activity in the Swedish Population. Ph.D. Thesis, Swedish School of Sport and Health Sciences, Stockholm, Sweden, 2016.

22. Rietjens, G.J.; Kuipers, H.; Kester, A.D.; Keizer, H.A. Validation of a computerized metabolic measurement system (Oxycon-Pro) during low and high intensity exercise. *Int. J. Sports Med.* **2001**, *22*, 291–294. [CrossRef] [PubMed]

23. Robergs, R.A. A critical review of the history of low- to moderate-intensity steady-state VO$_2$ kinetics. *Sports Med.* **2014**, *44*, 641–653. [CrossRef] [PubMed]

24. Judge, D.; Maygarden, J. ActiGraph gt3x File Format. April 2016. Available online: https://github.com/actigraph/gt3x-file-format (accessed on 2 February 2019).

25. Freedson, P.S.; Melanson, E.; Sirard, J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med. Sci. Sports Exerc.* **1998**, *30*, 777–781. [CrossRef] [PubMed]

26. Trost, S.G.; Ward, D.S.; Moorehead, S.M.; Watson, P.D.; Riner, W.; Burke, J.R. Validity of the computer science and applications (CSA) activity monitor in children. *Med. Sci. Sports Exerc.* **1998**, *30*, 629–633. [CrossRef] [PubMed]

27. Hildebrand, M.; van Hees, V.T.; Hansen, B.H.; Ekelund, U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med. Sci. Sports Exerc.* **2014**, *46*, 1816–1824. [CrossRef] [PubMed]

28. Eston, R.G.; Rowlands, A.V.; Ingledew, D.K. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *J. Appl. Physiol.* **1998**, *84*, 362–371. [CrossRef] [PubMed]

29. Howley, E.T. Comment on: Errors in MET estimates of physical activities using 3.5 mL·kg^{-1}·min^{-1} as the baseline oxygen consumption. *J. Phys. Act. Health* **2011**, *8*, 141–142. [CrossRef] [PubMed]

30. Bassett, D.R.; Rowlands, A.; Trost, S.G. Calibration and validation of wearable monitors. *Med. Sci. Sports Exerc.* **2012**, *44*, S32–S38. [CrossRef]

31. Garber, C.E.; Blissmer, B.; Deschenes, M.R.; Franklin, G.A.; Lamonte, M.J.; Lee, I.-M.; Nieman, D.C.; Swain, D.P. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Med. Sci. Sports Exerc.* **2011**, *43*, 1334–1359. [CrossRef]

32. Yngve, A.; Nilsson, A.; Sjostrom, M.; Ekelund, U. Effect of monitor placement and of activity setting on the MTI accelerometer output. *Med. Sci. Sports Exerc.* **2003**, *35*, 320–326. [CrossRef]

33. Vanhelst, J.; Zunquin, G.; Theunynck, D.; Mikulovic, J.; Bui-Xuan, G.; Beghin, L. Equivalence of accelerometer data for walking and running: Treadmill versus on land. *J. Sports Sci.* **2009**, *27*, 669–675. [CrossRef] [PubMed]

Article

Appropriate Mother Wavelets for Continuous Gait Event Detection Based on Time-Frequency Analysis for Hemiplegic and Healthy Individuals

Ning Ji [1,2], Hui Zhou [3], Kaifeng Guo [4], Oluwarotimi Williams Samuel [2], Zhen Huang [4], Lisheng Xu [1,*] and Guanglin Li [2,*]

[1]   College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China
[2]   CAS Key Lab of Human-Machine Intelligence-Synergy Systems of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen 518055, China
[3]   School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
[4]   Panyu Central Hospital, Guangzhou 511400, China
*   Correspondence: xuls@bmie.neu.edu.cn (L.X.); gl.li@siat.ac.cn (G.L.)

Abstract: Gait event detection is a crucial step towards the effective assessment and rehabilitation of motor dysfunctions. Recently, the continuous wavelet transform (CWT) based methods have been increasingly proposed for gait event detection due to their robustness. However, few investigations on determining the appropriate mother wavelet with proper selection criteria have been performed, especially for hemiplegic patients. In this study, the performances of commonly used mother wavelets in detecting gait events were systematically investigated. The acceleration signals from the tibialis anterior muscle of both healthy and hemiplegic subjects were recorded during ground walking and the two core gait events of heel strike (HS) and toe off (TO) were detected from the signal recordings by a CWT algorithm with different mother wavelets. Our results showed that the overall performance of the CWT algorithm in detecting the two gait events was significantly different when using various mother wavelets. By using different wavelet selection criteria, we also found that the accuracy criteria based on time-error minimization and F1-score maximization could provide the appropriate mother wavelet for gait event detection. The findings from this study will provide an insight on the selection of an appropriate mother wavelet for gait event detection and facilitate the development of adequate rehabilitation aids.

Keywords: gait event detection; hemiplegic gait; appropriate mother wavelet; acceleration signal; wavelet-selection criteria

1. Introduction

Gait disorders are usually associated with an ageing population as well as stroke survivors with hemiplegia, thus leading to both a reduced quality of life and an increased mortality rate. The detection of gait events or gait characteristics is essential to numerous applications including the development of control mechanisms in drop foot correction devices [1–5], human activity recognition for healthcare [6,7], motor recovery assessments for effective rehabilitation strategies [1,8], especially for patients with lower limb motor dysfunction following a severe stroke. Heel strike (HS) and toe off (TO) are regarded as two core gait events in a normal gait cycle, which provide information on the swing, stance, and stride gait parameters. Hence, it is essential to develop an effective algorithm for the accurate detection of HS and TO gait events.

With the rapid development of wearable electronic devices, inertial sensors such as accelerometers are fast becoming widely used in gait analysis due to their portable, low-power consumption, and

low-cost characteristics. In this regard, various algorithms aiming at gait event detection based on acceleration signals have been proposed in the recent years. In previous studies [4,9–11], a number of proposed algorithms have attempted to detect gait events based on a peak detection approach that integrates a filtering technique to the acceleration signal based on a set of predefined thresholds in the time domain. However, prior knowledge of the optimal thresholds is relatively difficult to adapt to different subjects in a real-world environment. Also, some attempts have been made to utilize machine learning algorithms for gait event detection via classification models especially with labelled datasets [12]. Note that such machine learning based approaches are limited because their operational procedures are driven by black-box models. Thus, it is difficult for developers and clinicians to truly understand the mechanism of gait event detection that often hinders proper interpretation for practical applications [9,13,14]. By exploiting the limitations of the above described threshold-based approaches and machine learning based methods [1,8], continuous wavelet transform (CWT) based methods have been considered as an alternative solution. By simultaneously providing a signal's information in the time and frequency domains through a series of decomposition and reconstruction operation, the wavelet transform method has been utilized as an effective tool in various fields, such as signal de-noising [15], speech/signal processing and evaluation [16–18], and health threatening illness diagnosis [19]. In the aspect of gait research, CWT based methods have been proven to offer a time-frequency representation with the capability to capture and analyze varying frequencies of acceleration signals, be adaptive to irregular acceleration patterns even in the presence of frequency variations, as well as to attenuate inherent noise and baseline drift [20]. Considering these advantages, wavelet transforms are being increasingly used and have been reported to have good results particularly for gait event detection [13,20–28].

Table 1 summarizes the recently proposed CWT-based methods for detecting gait events, in which different mother wavelets were used in those previous studies. Note that the selection of an appropriate mother wavelet has been stated as a central and open question in the application of wavelet transforms for gait signal analysis [29]. Thus, finding an appropriate mother wavelet for gait events detection could be a crucial step, which would benefit coefficient reconstruction and feature extraction from the signal in time and frequency domains. Meanwhile, in other research domains such as signal de-noising, several mother wavelet selection criteria have been investigated based on some qualitative or quantitative approaches [15,17,30–33]. Apart from the fact that the quantitative method has been strongly advocated, mother wavelet selection based on accuracy driven criteria has attracted a lot of research attention in the recent years due to its simplicity and reliability [30]. However, how to determine the appropriate mother wavelet in gait event detection was rarely reported in the previous studies, so the criteria for the mother wavelet selection is still lacking. Additionally, while most of the previous studies concentrated on the gait analysis of healthy subjects and Parkinson's disease (PD) patients, as showed in Table 1, few studies have been conducted to detect the gait events for hemiplegic patients who often suffer from gait disorders.

Towards determining an appropriate mother wavelet on gait event detection for both healthy subjects and hemiplegic patients, this study firstly constructed a general CWT algorithm. Secondly, we investigated the performance of 32 commonly applied mother wavelets in the recognition of two important gait events (HS and TO) using the acceleration signals obtained from 16 participants (including healthy subjects and hemiplegic patients) during level ground walking. By comparing the performance of different mother wavelets based on both accuracy (time-error, F1-score) and quantitative criteria (cross-correlation coefficient, energy-to-Shannon entropy ratio), the appropriate mother wavelet would be obtained. In this paper, the remaining part is organized as follows. Section 2 describes the improved general CWT algorithm and the criteria for selecting the appropriate mother wavelet. Section 3 presents the experimental results for gait event detection amongst the different mother wavelets and the different wavelet selection criteria in both healthy subjects and hemiplegic patients. Section 4 presents the discussion, and finally Section 5 concludes the work.

Table 1. A brief summary of some previous studies for gait event detection based on wavelet transforms.

Ref.	Subject	Sensor Position	Sensor Type	Wavelet	Detected Gait Parameters
[13]	Healthy subjects	Left and right ankles	Tri-axial accelerometer	Morlet (morl)	HS and TO events
[20]	Healthy subjects	Waist	Tri-axial accelerometer	Gaussian (gaus1)	HS and TO events
[21]	Healthy subjects	Ankle, thigh, waist, chest, upper arm and wrist	Tri-axial accelerometer	Morlet (morl)	HS and TO events
[22]	Parkinson's disease (PD) patients	Lower back	Tri-axial accelerometer	Gaussian (gaus1)	HS and TO events
[24]	Healthy subjects	Foot, ankle, shank and waist	Tri-axial accelerometer	Daubechies (db2)	HS and TO events
[26]	Healthy subjects	Tibialis anterior muscle of the lower leg	Tri-axial accelerometer	Morlet (morl)	HS and TO events
[27]	Parkinson's disease (PD) patients	Shank, thigh and lower back	Tri-axial accelerometer	Daubechies (db4)	Freezing of gait

2. Materials and Methods

2.1. Subjects Information and Experimental Procedure

A total of 16 subjects (13 healthy subjects and 3 hemiplegic patients) participated in this study for the analysis of their gait events. The 13 healthy subjects consisted of seven males and six females with an age range of 20 to 35 years, a height of 1.55 to 1.78 m, and a weight of 47 to 78 kg. The three post-stroke patients were composed of two males and one female with an age range of 44 to 59 years, a height of 1.63 to 1.71 m, and a weight of 54 to 78 kg. For the recruitment of healthy subjects, a preliminary assessment was conducted to ensure that they had no physical problems that might affect their gait patterns. For the recruitment of hemiplegic patients, the following four inclusion criteria were ensured: They had a unilateral hemiplegia after stroke, were able to walk at least 10 m independently without any help or assistive devices after motor rehabilitation treatment, were not diagnosed with other diseases that might affect their walking patterns, and were able to understand and cooperate with the experimental protocols. With these criteria, three patients who have been diagnosed with cerebral ischemic stroke were recruited from Guangzhou Panyu Central Hospital. Each patient's lower extremity motor functional level was evaluated using the Brunnstrom stage assessment method, which was used to reflect post-stroke motor recovery stage in clinical settings [34]. The clinical and demographic details of the hemiplegic subjects are summarized in Table 2. All the subjects gave the written informed consent to participate in the study and also provided the permission for the publication and educational purposes of their data/photographs. The experimental protocol of the study was approved by the Ethics Committee for Human Research, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

In the experiments, a wireless tri-axial accelerometer sensor (Delsys Inc., Natick, MA, USA) was attached to the tibialis anterior muscle of the right leg of each healthy subject and the paretic leg of each hemiplegic patient. As a reference for the ground true values of HS and TO, two force sensing resistor (FSR) sensors (Delsys Inc., Natick, MA, USA) were used to record the foot-switch signals of the right foot for comparison. To be precise, the two FSR sensors were placed under the toe and heel of each subject, respectively. The FSR and acceleration signals were concurrently recorded using a commercial data acquisition system (Delsys Inc., Natick, MA, USA) and transmitted wirelessly to a computer system for storage. To ensure high quality signal recordings, a bandage (Kindmax Inc., Irvine, CA, USA) was used to firmly fix the sensors to the leg so as to minimize their displacement and vibration during walking. A diagram showing the experimental setup is presented in Figure 1. During the experiments, each subject was instructed to walk continuously along a 10 m pathway on level ground at their preferred speed (regarded as the norm). The subjects repeatedly walked from

one side of the pathway to another side, and then made a U-turn to walk back for around one minute. Both the FSR and acceleration signals were simultaneously sampled at a rate of 148.15 Hz and then the acquired signals were analyzed using an offline MATLAB R2018b programming tool.

Table 2. Clinical and demographic details of the hemiplegic subjects.

No.	Age	Height (cm)	Weight (kg)	State of Illness	Brunnstrom Stage (Lower Limb)	Diagnosis	Symptom
1	59	163	78	12 months	V	Cerebral Ischemic Stroke	Left limb hemiplegia
2	44	171	54	6 months	IV	Cerebral Ischemic Stroke	Right limb hemiplegia
3	53	167	61	7 months	IV	Cerebral Ischemic Stroke	Right limb hemiplegia
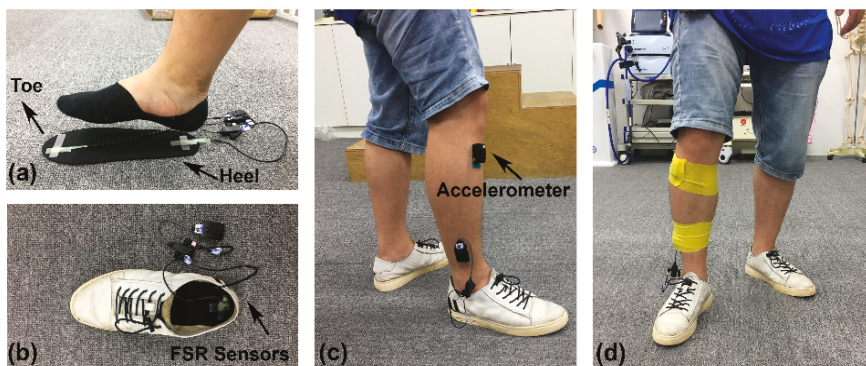
Figure 1. Experimental setup for force sensitive resistor (FSR) and accelerometer sensors placement in preparation for gait data acquisition. (**a**) A pair of FSR sensors were placed under the big toe and heel of the insole; (**b**) the placement of the insole into the shoe; (**c**) the placement of the accelerometer on the tibialis anterior muscle under the knee joint of right leg; (**d**) a bandage was used to firmly fix the sensors.

2.2. CWT Based Gait Event Detection Algorithm

A cycle of human gait can be divided into a sequence of repeated events and phases. More specifically, one gait cycle consists of stance and swing phases. The HS and TO events mark the beginning of stance and swing phases, respectively. In this case, it is obvious that the occurrence frequency of the two gait events (HS and TO) is twice of one gait cycle. This time-frequency relationship between the gait event and the gait cycle through the CWT can be used to detect gait events [13]. Moreover, it should be noted that the human gait is chaotic in nature in the real-world environment [35]. Thus, continuous wavelet transform methods have been proven to be robust and stable in several previous studies in effectively detecting gait events in practical applications that may be subject to various disturbances [13,20–22,26,28]. Here we propose a general CWT algorithm for the detection of two gait events (HS and TO), which is a combination of the methods in the two previous studies by Minh H. Pham et al. [22] and Siddhartha Khandelwal et al. [13].

The operational procedure of the general CWT based gait event detection algorithm is presented in Figure 2. Here the components of the tri-axially recorded acceleration signals from the anterior-posterior (AP) axis during level ground walking were used for the analysis of gait events [22,36]. In order to improve the quality of the AP signal recordings, the algorithm began with a pre-processing of the AP acceleration signals that were composed of three sequential procedures. Firstly, a linear de-trending algorithm was adopted to reduce the effect of noise and interference on the AP signal baseline. Secondly, a low-pass filtering at 10 Hz with a second-order Butterworth filter was used to

remove the high-frequency interferences from the AP signals. Finally, the AP signals were smoothed by integration using the inbuilt function in MATLAB called *cumtrapz*, where the approximation of the cumulative integral of each sampling points along the time intervals was calculated.
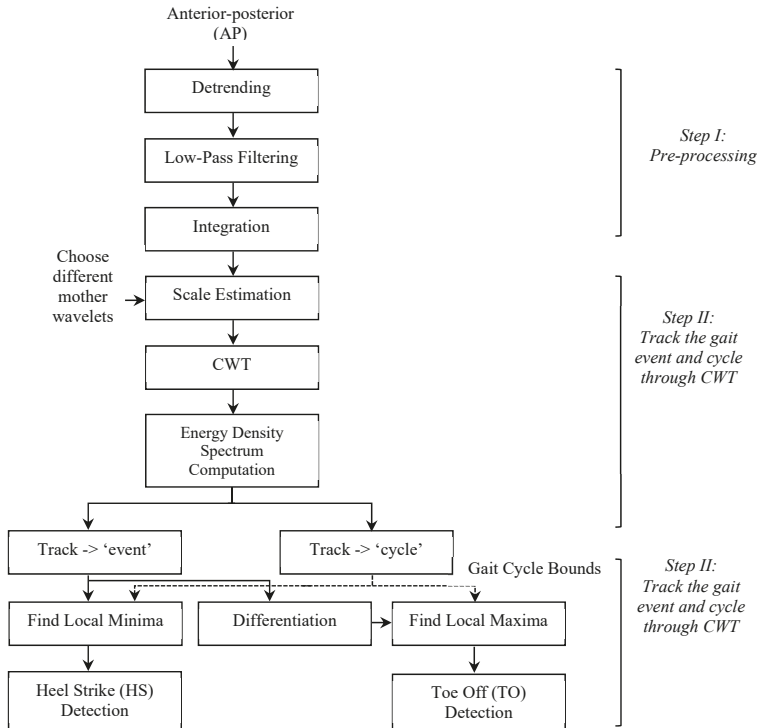
Figure 2. Conceptualization of the proposed general CWT algorithm with three main phases: (1) Pre-processing of the acceleration signals; (2) Tracking the gait events and gait cycles through time-frequency analysis; (3) Distinguishing HS and TO gait events.

After the AP acceleration signal preprocessing, the time-frequency analysis was performed to track the inherent gait events as well as gait cycles. The CWT of a discrete time signal x_n with equal time spacing δ_t is defined as the inner product of x_n with a scaled and translated mother wavelet ψ as expressed in Equation (1).

$$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi * \left[\frac{(n'-n)\delta_t}{s} \right] \tag{1}$$

where $W_n(s)$ denotes the wavelet transform, s is the wavelet scaling factor, n is the localized time index, and the (*) indicates the complex conjugate. Detailed definitions of different mother wavelets (ψ) and the prescription of wavelet selection criteria are presented in Section 2.4. The range of scales for CWT analysis $[1, s_{max}]$ was chosen using the frequency scale relationship of the chosen wavelet [26], which is presented in Equation (2).

$$s_{max} = \frac{f_c \times F_s}{f} \tag{2}$$

where f_c is the central frequency of the wavelet, F_s is the data sampling frequency, and f is the gait frequency. Since the healthy subjects usually walk at an average speed of 1.5 m/s, a minimum gait frequency of 0.5 Hz was utilized in this study as the norm [26]. In the experiments, the hemiplegic

subjects walked with an average speed of around 0.7–0.8 m/s. Thus, a minimum gait frequency of 0.25 Hz was assumed for the hemiplegic subjects.

Given a mother wavelet, the corresponding frequency scale can be calculated by Equation (2) for CWT analysis, and then the CWT coefficients can be obtained by Equation (1). As a typical example, Figure 3 graphically demonstrates the CWT plots of the acceleration signals during walking from one healthy subject (Figure 3a) and one hemiplegic subject (Figure 3b), where two mother wavelets ("db6" and "morl") were adopted for comparison. It was obvious from Figure 3 that the frequency of gait events underlying the corresponding scales is approximately two times of gait cycles. To distinguish the gait event and the gait cycle, the scale-dependent energy density spectrum E_s was computed based on the obtained CWT coefficients by using Equation (3).

$$E_s \ = \ \sum_{n=0}^{N-1} |W_n(s)|^2, \quad s \in [1, \ s_{max}] \tag{3}$$

where N is the total number of wavelet coefficients and $|W_n(s)|^2$ is the 2-D wavelet energy density function that measures the total energy distribution of the signal associated with the s scale. The peaks in E_s represent the dominant energy scales that contribute most to the signal's energy in the spectral domain [13]. Typically, two distinct spectral energy peaks can be obtained from the above computed E_s parameter and the corresponding peak scales have a ratio of two due to the frequency relationship between the gait event and the gait cycle. Since the candidate scales of gait events and gait cycles were found from the peaks of E_s, the peaks were stored as the set of local maxima points {(s_m, E_{sm}), $m \in [1, M]$}; where s_m is the scale corresponding to a peak value of E_{sm} and M is the total number of peaks. Afterwards, the following three possible cases were checked to determine the scale of the associated gait cycle and gait event in the energy density spectrum.
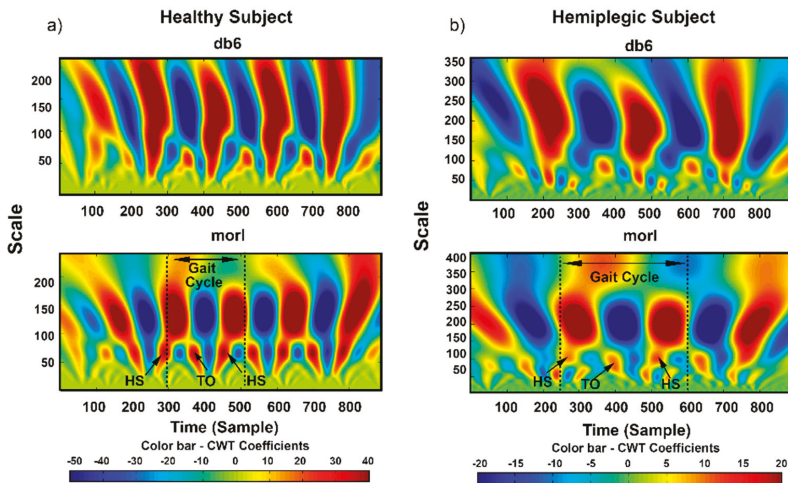
Figure 3. Continuous wavelet transform (CWT) plots of the acceleration signals during walking from one healthy subject (**a**) and one hemiplegic subject (**b**), where two mother wavelets ("db6" and "morl") were adopted for comparison. The underlying scale relationship between the gait events (HS and TO) and gait cycles are illustrated.

Case I: If there was only a single peak ($M = 1$) in the energy density spectrum, then the dominant spectral scale was tracked as the cycle scale using $s_{cycle} = s_1$ and the event scale was tracked using the $s_{event} = round(s_{cycle}/2)$ relationship.

Case II: If there were two dominant peaks ($M = 2$) in the energy density spectrum, then the cycle scale was tracked using $s_{cycle} = s_2$ and the event scale was tracked using the $s_{event} = s_1(s_1 < s_2)$ relationship.

Case III: If there were more than two peaks ($M \geq 3$) in the energy density spectrum, then the ratio of any two consecutive (neighboring) peaks denoted as $\delta = s_{m+1}/s_m$ ($s_m < s_{m+1}, m \in [1, M)$) was checked with the relaxed spectral relationship that $\delta \in [1.6, 2.4]$ to attenuate interferences in the signal and manage the occurrence of low frequency [14]. Thus, the two latest peaks that satisfied the above condition were tracked as the dominant peaks using the $s_{cycle} = s_{m+1}$ and $s_{event} = s_m$ ($s_m < s_{m+1}$) relationships.

Based on the conditions specified in Cases I–III, the scales of gait event and gait cycle were determined with the associated CWT coefficients (x_{event} and x_{cycle}) as shown in Equations (4) and (5).

$$x_{event} \triangleq W_n\left(s_{event}\right) \tag{4}$$

$$x_{cycle} \triangleq W_n\left(s_{cycle}\right) \tag{5}$$

After de-trending both the above temporal signals (x_{event} and x_{cycle}), the gait cycle bounds could be defined as the maximum points of x_{cycle}, which marked the beginning of each gait cycle. Within every gait cycle, the HS event was defined as the first local minima of x_{event} and TO event was defined as the second local maxima of the further differentiated x_{event}. An illustration of the temporal representation of HS and TO gait event detection based on the proposed general CWT algorithm is illustrated in Figure 4.
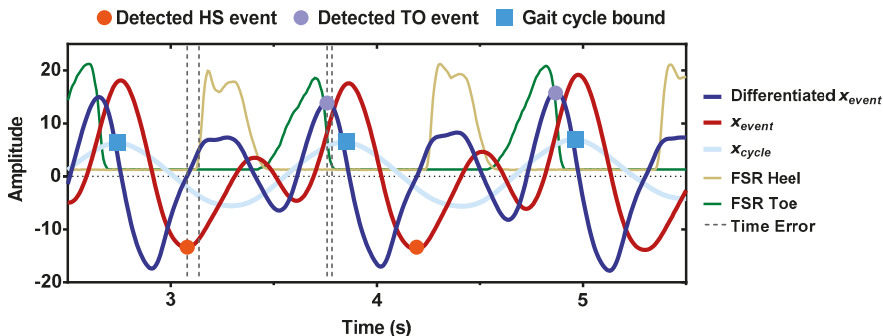
Figure 4. Temporal representation of HS and TO gait event detection based on the proposed general CWT algorithm. The first local minima of x_{event} corresponded to the estimated HS event (orange circle). The second local maxima of the further differentiated x_{event} corresponded to the estimated TO event (purple circle). Each gait cycle bound was defined as the maximum points of x_{cycle}. Vertical dashed lines indicated the time-errors between the estimated event and the reference event from the FSR method.

2.3. Verification of the CWT Based Gait Event Detection Algorithm

As a direct measure of pressure distribution under the foot during walking, the FSR method is commonly used as the gold standard for gait event detection in previous gait literatures [8]. Thus, the FSR method was utilized as the reference method to evaluate the performance of the proposed CWT based gait event detection method in our study. In the FSR method, the HS and TO events were detected by a set of threshold values. The threshold value for the detected HS event was calculated as 5% of the maximum heel FSR amplitude at the foot force increasing stage. Meanwhile, the threshold value for the detected TO event was calculated as 5% of the maximum toe FSR amplitude at the foot force decreasing stage. Detailed description is provided in previous literatures [26,37]. It should be noted that the heel and toe FSR signals of each subject were collected simultaneously with the walking

acceleration signals through a pair of FSR sensors, which was described in the experimental procedure (see Section 2.1).

In this study, the time-error and F1-score were utilized as the metrics to assess the gait event detection performances, which were commonly adopted in the previous gait literatures [13,20,38]. It is worthy to note that the gait events detected with the FSR recordings were used as the reference events for each subject. And the gait events detected by using the proposed general CWT method based on the AP acceleration signals were designated as the estimated events. Time-error, a measure of the time agreement between the reference events and the estimated events, was utilized to assess the time accuracy among the correctly detected gait events when the proposed method was applied. The time-errors of the estimated HS and TO gait events, designated as TE_{HS} and TE_{TO}, respectively, were defined as the following Equations (6) and (7):

$$TE_{HS} = \left| T_{HS\_CWT} - T_{HS\_FSR} \right| \tag{6}$$

$$TE_{TO} = \left| T_{TO\_CWT} - T_{TO\_FSR} \right| \tag{7}$$

where T_{HS_CWT} and T_{TO_CWT} represent the time indexes of the estimated HS and TO gait events via the CWT-based algorithm, and T_{HS_FSR} and T_{TO_FSR} represent the time indexes of the reference HS and TO gait events via the FSR method, respectively.

In cases of possibly missed and wrongly detected gait events, the F1-score was adopted as a measure of the precision and recall of the gait event detection, which is calculated as follows:

$$\text{F1} = \frac{2PR}{P + R} \tag{8}$$

where P denotes Precision and R denotes Recall, which are computed as Equations (9) and (10), respectively:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

where TP denotes the true positives, FN denotes false negatives, and FP denotes false positives. Note that in the context of this study, TP actually represented the number of correctly estimated gait events, FN denoted the number of unidentified/missed gait events, and FP represented the number of wrongly estimated gait events.

2.4. Appropriate Mother Wavelet Selection

Considering the fact that mother wavelet selection plays an important role in the overall performance of CWT based algorithms, there is a need for an effective means of determining the appropriate mother wavelet especially in the context of gait event detection for hemiplegic patients. In order to determine an appropriate mother wavelet, we systematically investigated the performance of 32 commonly applied mother wavelets in detecting gait events using the measures of both accuracy and quantitative criteria. The targeted mother wavelet functions included ten Daubechies (db1–db10), five Coiflets (coif1–coif5), seven Symlets (sym2–sym8), eight Gaussian (gaus1–gaus8), one Morlet (morl), and one Meyer (meyr). Table 3 briefly summarizes the definitions and main properties of the investigated wavelet families in this study.

Table 3. Brief summary of the definitions and main properties of the studied wavelet families.

Wavelet Family	Order N	Orthogonality	Symmetry	Explicit Expression
Haar	db 1	Orthogonal	Symmetric	$\psi_{Haar}(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & otherwise \end{cases}$
Daubechies	db 2–10	Orthogonal	Asymmetric	No
Coiflets	coif 1–5	Orthogonal	Near symmetric	No
Symlets	sym 2–8	Orthogonal	Near symmetric	No
Gaussian	gaus 1–8	No	Symmetric	$\psi_{Gaussian}(t) = -\frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}}$
Morlet	morl	No	Symmetric	$\psi_{Morlet}(t) = \frac{1}{\sqrt[4]{\pi}} e^{i 2\pi f_c t} e^{-\frac{t^2}{2}}$
Meyer	meyr	No	Symmetric	$\psi_{Meyer}(f) =$ $\begin{cases} \sqrt{2\pi} e^{i\pi f} \sin\left[\frac{\pi}{2} v\left(3|f| - 1\right)\right] & \frac{1}{3} \leq |f| \leq \frac{2}{3} \\ \sqrt{2\pi} e^{i\pi f} \cos\left[\frac{\pi}{2} v\left(\frac{3}{2}|f| - 1\right)\right] & \frac{2}{3} \leq |f| \leq \frac{4}{3} \\ 0 & |f| \notin \langle \frac{1}{3}, \frac{4}{3} \rangle \end{cases}$

N is the order of the mother wavelet. ψ is the mother wavelet function if the wavelet has the explicit expression. f_c is the wavelet central frequency, t denotes the time, and f denotes the frequency.

Generally, the basic properties of each mother wavelet, such as the orthogonality and symmetry, are considered in order to help select the appropriate mother wavelet in the first step. The orthogonality characteristics can ensure that the signal will not be decomposed into overlapping sub-frequency bands. Symmetry property can help avoid phase distortion, which is mainly concerned in the wavelet-based filtering operation because the mother wavelet can be served as a linear phase filter. As pointed out in [30], there is possibly more than one mother wavelet sharing similar fundamental properties. Thus, the selection of an appropriate mother wavelet cannot solely rely on the basic properties of the wavelets. Besides, the visual similarity generally applied in the optimal mother wavelet selection is also reported to be not always proper for all wavelet-based processing [30,31]. Additionally, the wavelet selection that is based on accuracy related to specific application is more recommended [30]. In this case, both accuracy (time-error, F1-score of the gait event detection) and quantitative (cross-correlation coefficient, energy-to-Shannon entropy ratio) criteria were investigated to search for an effective mother wavelet associated with HS and TO event detection for healthy and hemiplegic subjects. For the accuracy criteria, kindly note that the definition of time-error and F1-score can be referred in Section 2.3.

For the quantitative criteria, cross-correlation coefficient (*Xcorr*) and energy-to-Shannon entropy ratio (*ESER*) metrics were applied to evaluate the performances of the 32 mother wavelets in gait event detection in this study. Note that they were two commonly used quantitative wavelet selection criterion in various applications such as signal de-noising, signal processing/decoding, and vibration signal analysis [15,17,39]. In the cross-correlation measure, the cross-correlation coefficient (*XCorr*) between the recorded acceleration signal (*X*) and a specific mother wavelet (Y_i) was obtained and utilized to quantify the performance of the mother wavelet (Y_i) in detecting the associated gait events [31,32]. It is noteworthy that the higher the absolute $XCorr(X, Y_i)$, the stronger the correlation will be and a value of 0 indicates that the two variables are linearly independent. The cross-correlation coefficient $XCorr(X, Y_i)$ is computed using the formula in Equation (11).

$$XCorr(X, Y) = \frac{\sum\left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{\sqrt{\sum\left(X - \overline{X}\right)^2\left(Y - \overline{Y}\right)^2}} \tag{11}$$

In the energy-to-Shannon entropy ratio (*ESER*) measure, the goal is to obtain a mother wavelet that provides the maximum energy, and at the same time minimum Shannon entropy associated with the wavelet coefficients [39]. Thus, the energy-to-Shannon entropy ratio is computed using the formula in Equation (12).

$$ESER(s) = \frac{E(s)}{S_{entropy}\ (s)} \tag{12}$$

where the energy $E(s)$ associated with s scale was computed using Equation (3) and the Shannon entropy $S_{entropy}(s)$ was defined as the Equation (13).

$$S_{entropy}(s) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i \qquad (13)$$

where p_i is the energy probability distribution of each wavelet coefficient and is computed using the formula in Equation (14). $W_n(s)$ represents the wavelet transform of the signal associated with the s scale, which was defined in Equation (1). E_s represents the energy density spectrum which is computed based on the above mentioned CWT coefficients $W_n(s)$ and was defined in Equation (3). N is the total number of wavelet coefficients. Specifically, if $p_i = 0$, then $\sum_{i=1}^{N} p_i = 1$.

$$p_i = \frac{|W_n(s)|^2}{E(s)} \qquad (14)$$

In summary, the mother wavelet that provides the maximum F1-score, *XCorr*, and *ESER* as well as the minimum time-error is generally considered to be the most appropriate mother wavelet for gait event detection.

2.5. Statistical Analysis

Statistical tests were conducted using a SPSS 21 software package to further validate the findings of the study. Specifically, the most appropriate mother wavelets among the healthy and hemiplegic subjects were determined by evaluating the total time-error, F1-score, *XCorr*, and *ESER* of the HS and TO events between the proposed general CWT algorithm and the FSR method through one-way ANOVA. Afterwards, a post-hoc comparison was performed based on Duncan's test, and the significance level was set to $p < 0.05$ for all the analysis.

3. Results

3.1. Appropriate Mother Wavelet Selection Based on Accuracy Criteria

The gait event detection results for the healthy and hemiplegic subjects using the two accuracy criteria, the time-error and F1-score measures for mother wavelet selection, are presented as follows. Figure 5 presents the averaged time-errors associated with HS and TO estimated gait events for the healthy and hemiplegic subjects across the 32 mother wavelets. Notably, the lower the time-error value is, the higher the time-agreement with the FSR reference is. It can be seen from Figure 5a that the "sym2", "db7", and "db6" mother wavelets achieved a relatively lower averaged time-error value of 0.06 ± 0.03 s, 0.09 ± 0.07 s, and 0.10 ± 0.03 s, respectively, for the total of the estimated HS and TO gait events over all the healthy subjects in comparison to the other mother wavelets. Meanwhile, "db6", "sym5", and "db5" achieved a lower averaged time-error value of 0.18 ± 0.05 s, 0.21 ± 0.07 s, and 0.26 ± 0.07 s for gait event detection over all the hemiplegic participants, respectively, as shown in Figure 5b. Note that using the "db6" mother wavelet, the averaged time-error values were lowest over the hemiplegic participants (Figure 5b) and were relatively lower over the healthy subjects (Figure 5a). These findings suggest that the "db6" is an appropriate mother wavelet in gait event detection for both healthy and hemiplegic subjects.
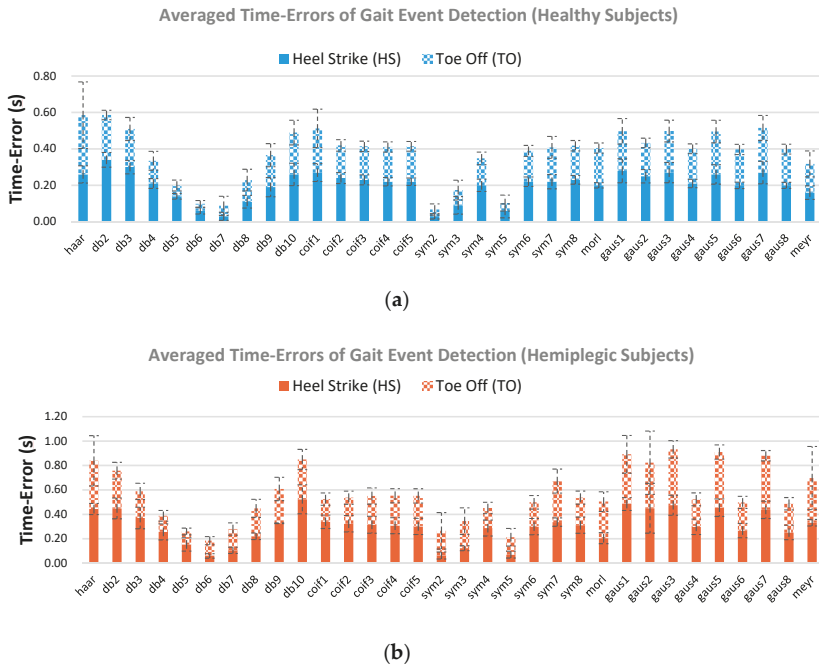
Figure 5. Averaged time-error values of the estimated HS and TO gait events over all the healthy subjects (**a**) and all the hemiplegic subjects (**b**) when using 32 commonly applied mother wavelets. The vertical dashed lines indicate the standard deviation.

As a comparison, Figure 6 plots the Bland–Altman time agreements between the reference gait events and the estimated gait events, where the sample points were obtained from the time-error values of the healthy (Figure 6a) and hemiplegic subjects (Figure 6b) when using the appropriate mother wavelet "db6" and one previously used wavelet "gaus1" in another study [15], respectively. Note that the absolute value of time-errors is presented in Figure 5, whereas the true value of the time agreement is plotted in Figure 6. Here a negative time-error value corresponds to a time-delay in the estimated gait events with respect to the FSR reference gait events, and a positive value means a time-advance. We can see from Figure 6 that for the healthy subjects, the mother wavelet "db6" had a mean time-error of -0.06 s with a 95% confidence interval (CI) of $[-0.12, -0.01]$ for HS gait event detection and -0.02 s $[-0.19, 0.14]$ for TO gait event detection. While using the mother wavelet "gaus1", the mean time-error of 0.16 s $[-0.23, 0.55]$ was achieved for HS gait event detection and 0.12 s $[-0.20, 0.44]$ for TO gait event detection. For the hemiplegic subjects, the "db6" mother wavelet achieved a mean time-error of -0.04 s $[-0.11, 0.02]$ for HS gait event detection and 0.18 s $[-0.01, 0.37]$ for TO gait event detection, whereas the "gaus1" mother wavelet achieved a mean time-error of -0.08 s $[-1.30, 1.20]$ for HS gait event detection and -0.21 s $[-1.40, 0.95]$ for TO gait event detection. Thus, the performances of "db6" were rather good in comparison to "gaus1" for both the healthy and the hemiplegic subjects.
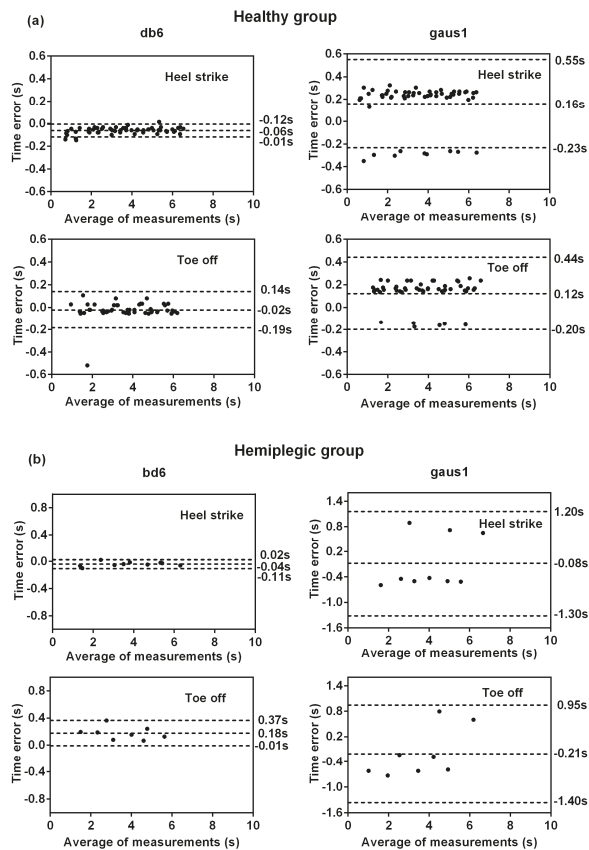
Figure 6. Bland–Altman plots of time agreement between the proposed modified CWT algorithm and the FSR method for HS and TO gait event detection in the healthy group (**a**) and the hemiplegic group (**b**). The time agreement results of selecting the optimal wavelet "db6" are shown on the left side whereas results of the commonly used wavelet "gaus1" with rather poor performance are shown on the right side. Positive times correspond to delays in the gait event detection of the proposed modified CWT algorithm with respect to the reference FSR method. The horizontal axis represents the average time measures for detecting gait events by both methods, and the vertical axis is the time error between the two methods. The dashed line from top to down respectively represent the 95% CI upper limit, the mean, 95% CI lower limit of the time difference (in seconds).

Figure 7 shows the averaged HS and TO gait event detection results obtained by using the F1-score criterion across all the 32 mother wavelets. The vertical dotted line indicates the standard deviation. It can be observed from Figure 7 that four of all the mother wavelets, "sym5", "db5", "db6", and "sym7" achieved relatively higher average F1-scores (1.00 ± 0.01, 0.99 ± 0.02, 0.99 ± 0.02, and 0.99 ± 0.02 across the healthy subjects (blue line), respectively). For the hemiplegic patients (red line), "db5", "db6", and "morl" mother wavelets achieved higher averaged F1-scores (0.98 ± 0.02, 0.98 ± 0.03, and 0.98 ± 0.02, respectively). Note that "db5" and "db6" mother wavelets are more stable with relatively high F1-scores for both the healthy and hemiplegic subjects, indicating that the two mother wavelets would be the appropriate mother wavelets for the accurate detection of gait events in comparison to the other mother wavelets regardless of the status of the subject.

Figure 7. Averaged F1-scores of HS and TO gait event detection over all the healthy subjects (blue line) and over all the hemiplegic subjects (red line) across different mother wavelets. The vertical dashed lines indicate the standard deviation.

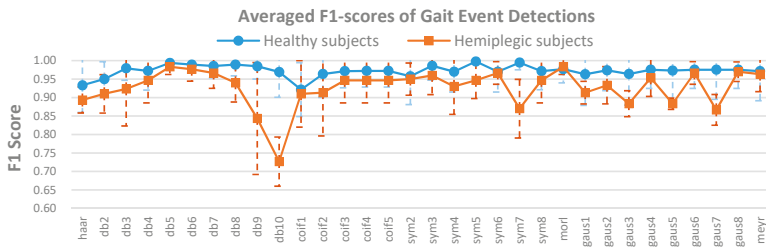Furthermore, when using the different mother wavelets, the ANOVA analysis of the time-errors (Figure 5) between the FSR reference events and the estimated gait events was significantly different ($p < 0.001$) for both healthy and hemiplegic subjects. This statistical result suggests that the time-error should be a proper criterion for selection of an appropriate mother wavelet in gait event detection for all the subjects. Meanwhile, the ANOVA analysis of the F1-scores (Figure 7) was significantly different for the hemiplegic subjects ($p < 0.01$) and was not significantly different for the healthy subjects ($p > 0.05$). This suggested that the F1-score could also be used as a criterion for the selection of an appropriate mother wavelet in gait event detection for hemiplegic subjects, but probably not for healthy subjects.

3.2. Appropriate Mother Wavelet Selection Based on Quantitative Criteria

Additionally, the detection of gait events was also performed using two quantitative criteria, the cross-correlation coefficient and the energy-to-Shannon entropy ratio measures, to obtain an appropriate mother wavelet. The analysis results of the mother wavelets based on the cross-correlation coefficient criterion for gait event detection are presented in Figure 8. It can be seen from Figure 8a that amongst the Daubechies family, "db2" and "sym2" achieved the highest average correlation coefficient of 0.34 between the acceleration and the underlying mother wavelet followed by "gaus1" in the Gaussian family with an average value of 0.33. In other words, the "db2", "sym2", and "gaus1" mother wavelets have high similarity with the anterior-posterior acceleration signals from which the participants' gaits were detected. Meanwhile, we found that "db3", "sym3", and "gaus3" achieved the best correlation coefficients between the acceleration and the underlying mother wavelet function in their respective categories for the hemiplegic patients (Figure 8b).

The analysis results of mother wavelet characteristics based on the energy-to-Shannon entropy ratio criterion for gait event detection are presented in Figure 9. It can be observed that the average energy-to-Shannon entropy ratios for most of the investigated mother wavelets lies in the range of −0.15 to −0.13 with the "morl" mother wavelet achieving the highest ratio for healthy subjects (blue line). On the other hand, except for the "morl" mother wavelet that also provided the highest average energy-to-Shannon entropy ratio across the hemiplegic subjects (red line), the other mother wavelets had the average energy-to-Shannon entropy ratios that were in the range of −0.16 to −0.14. Although the "morl" mother wavelet function achieved a relatively higher mean energy-to-Shannon entropy ratio, however there was no significant difference between the "morl" and the other investigated mother wavelet functions based on ANOVA statistical analysis ($p > 0.05$).

(**a**)

(**b**)
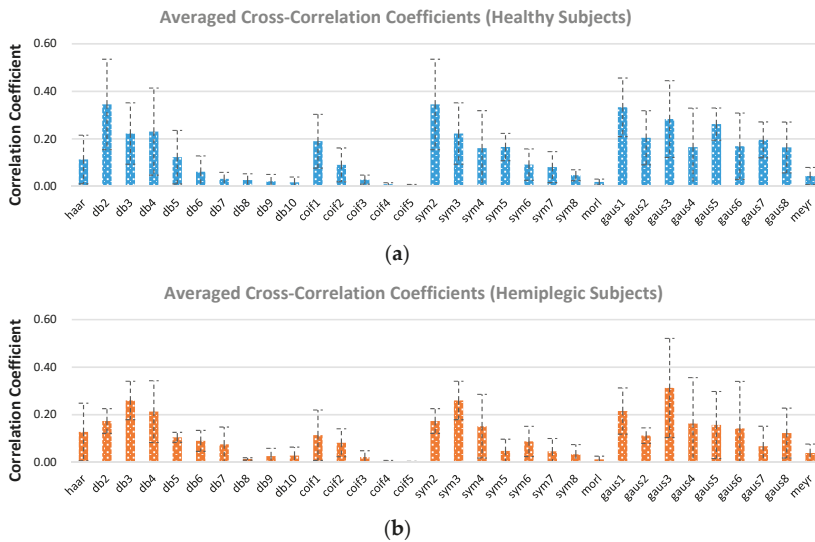
Figure 8. Averaged cross-correlation coefficients between the anterior–posterior knee acceleration signal and specific mother wavelet function for healthy subjects (**a**) and hemiplegic subjects (**b**). Vertical dotted line indicates the standard deviation.
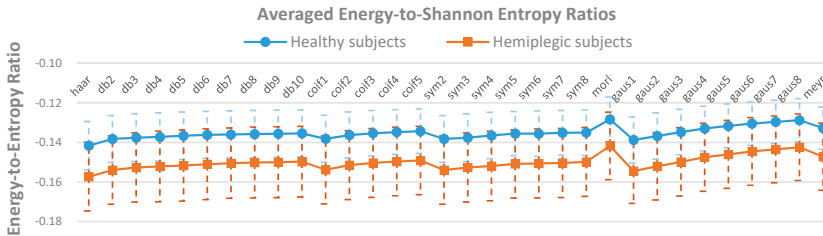
Figure 9. Averaged energy-to-Shannon entropy ratios of CWT coefficients for healthy subjects (blue line) and hemiplegic subjects (red line) across different mother wavelets ($p > 0.05$). Vertical dotted line indicates the standard deviation.

4. Discussion

Although several CWT-based algorithms have been proposed for gait event detection, the validation of such algorithms especially on hemiplegic patients and the investigation into an appropriate mother wavelet selection has rarely been conducted to date. Thus, there is a need to embark on the current study. This study systematically investigated different wavelet selection criteria across the 32 commonly applied mother wavelets towards obtaining the appropriate mother wavelet required for consistently accurate gait event detection especially for hemiplegic patients. The mother wavelets were individually incorporated into a proposed general CWT algorithm for the recognition of HS and TO gait events, and the algorithm was validated with datasets from hemiplegic patients as well as healthy subjects.

It should be noted that the proposed general CWT algorithm showed a good performance for HS and TO gait event detection in terms of consistently achieving high accuracy and low time-error particularly when the appropriate mother wavelet was applied to both healthy and hemiplegic subjects. The gait event detection results for the healthy subjects were observed to be comparatively stable and consistent with that reported in previous studies which utilized only a specific single mother

wavelet [13,22,26]. Interestingly, we also found that the validation results of the estimated gait events for the hemiplegic subjects were observed to be consistent with the results obtained for healthy subjects. One possible reason is due to the use of time-frequency analysis and the domain knowledge about gait events and gait cycles. Since functional gait after stroke is usually affected by spasticity, muscle weakness, and balance disorder [40], hemiplegic gait is often characterized by reduced speed, cadence, stride length and joint angular excursions, and asymmetry in temporal and spatial domains [41]. In addition, limb circumduction is clearly observed on the hemiplegic subjects during walking due to the increased lateral displacement of the foot during the swing in the paretic limb [41], and these factors may increase the complexity of recognizing the gait patterns of hemiplegic patients. However, with the proposed CWT algorithm that incorporates time-frequency information, the gait events of both hemiplegic and healthy subjects can be adequately identified as shown in Figure 3. Therefore, the CWT based algorithm that integrates the optimal wavelet function shows its robustness in HS and TO event detection, which indicates its superiority with no need to readjust the thresholds in comparison with some purely data-driven approaches that are often dependent on threshold and parameter tuning. Hence, this again proves the potential of applying an appropriate mother wavelet to the proposed general CWT algorithm for gait event detection in real-life applications.

When considering the selection of the appropriate mother wavelet, the criteria based on some accuracy and quantitative measures were investigated. For the accuracy-based criteria, time-error and F1-score measures were applied to evaluate the performance of a range of mother wavelets employed by the proposed CWT algorithm to detect the HS and TO gait events. From the analysis results, we found that the "db6" mother wavelet achieved relatively higher F1-scores and yielded the lowest average time-error values for both healthy and hemiplegic subjects. The possible reason is that the "db6" mother wavelet provide the most precise scale for CWT based gait event detection in comparison to the other 31 analyzed mother wavelets. With respect to the quantitative-based criteria, the cross-correlation coefficient (*Xcorr*) and energy-to-Shannon entropy ratio (*ESER*) were investigated to check whether the above two metrics could be considered as effective wavelet selection criteria for the appropriate mother wavelet in detecting gait events. Since the CWT are essentially based on finding the correlation between the analyzed signal (the acceleration signal) and the shifted/scaled mother wavelet, it is obvious that an initial criterion for appropriate wavelet selection could be the cross-correlation (*Xcorr*) that reflects the similarity. However, we found that the results of cross-correlation (*Xcorr*) analysis were not consistent with the actual performances of HS and TO gait event detection in both healthy subjects and hemiplegic patients. This may be due to the fact that cross-correlation (*Xcorr*) is only appropriate for those wavelet-based processing methods based on the resemblance between signals and mother functions [31], but not for our models which extract and distinguish the feature points based on the frequency relationship of the signal itself. After investigating the correlation between the wavelet and the analyzed signal (*Xcorr*), we also wanted to check if there were significant differences that exist in the wavelet coefficients themselves by using different mother wavelets for CWT. Which is to say, the energy concentration and entropy that reflects information lost were integrated to check whether the energy-to-Shannon entropy ratio (*ESER*) could be considered as one of the effective wavelet selection criteria. As pointed out in the previous study [42], maximization of time/frequency energy concentration, minimization of the bias, and the unique relationship that exists between scale and frequency contribute to desirable continuous analytic wavelets. However, we also found that the results of the ESER analysis across different mother wavelets had no significant difference, which might not be appropriate for the wavelet selection in our study. In other words, the results of the *ESER* analysis revealed that the energy concentration and entropy distribution were almost the same across different mother wavelets in the aspect of gait event detection. Therefore, the accuracy-based criteria including time-error and F1-score is suggested as effective wavelet selection criteria in the context of gait event analysis.

Despite the interesting findings observed in this study, there are still some limitations that should be addressed in the future work. For instance, the proposed general CWT algorithm was validated

using the dataset from post-stroke hemiplegic patients who were said to be in stages V and IV according to the Brunnstrom assessment scale, excluding patients with severe stroke (perhaps in stages II and III on the Brunnstrom assessment scale). Therefore, there is need to conduct further studies with this population since they have the most drop foot issues, and their gait patterns maybe somewhat different from those in stages V and IV considered in the current study. Besides that, different sensor placement configuration as well as different walking terrains would normally influence gait patterns in real-life applications. In the future, we would further investigate the performances of different mother wavelets in the presence of both factors, and perhaps make adjustment, if necessary, to the mother wavelet selection criteria examined in this study.

5. Conclusions

While CWT based gait analysis methods have been widely adopted in previous studies, their performances in gait event detection when using different mother wavelets have rarely been studied, especially in hemiplegic patients. In this study, different mother wavelets and wavelet selection criteria were systematically investigated by using acceleration signals recorded from both hemiplegic and healthy subjects. The experimental results demonstrated that an overall significant difference in performance of the CWT algorithm was observed when using different mother wavelet functions for detecting HS and TO gait events, which suggested the need for this study. Additionally, we found that the accuracy criteria based on time-error minimization and F1-score maximization led to the realization of an appropriate mother wavelet named as "db6" which achieved the highest detection accuracy with lowest detection time-error for both hemiplegic and healthy subjects. The outcomes of this study may provide an insight on mother wavelet selection criteria for gait event analysis especially in hemiplegic patients, and may ultimately facilitate the practical development of rehabilitation devices or strategies for them.

References

1. Cui, C.; Bian, G.B.; Hou, Z.G.; Zhao, J.; Su, G.; Zhou, H.; Peng, L.; Wang, W. Simultaneous Recognition and Assessment of Post-Stroke Hemiparetic Gait by Fusing Kinematic, Kinetic, and Electrophysiological Data. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *24*, 856–864. [CrossRef] [PubMed]
2. Lyons, G.M.; Sinkjaer, T.; Burridge, J.H.; Wilcox, D.J. A review of portable FES-based neural orthoses for the correction of drop foot. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2002**, *10*, 260–279. [CrossRef] [PubMed]
3. Kotiadis, D.; Hermens, H.J.; Veltink, P.H. Inertial Gait Phase Detection for control of a drop foot stimulator Inertial sensing for gait phase detection. *Med. Eng. Phys.* **2010**, *32*, 287–297. [CrossRef] [PubMed]
4. Rueterbories, J.; Spaich, E.G.; Andersen, O.K. Gait event detection for use in FES rehabilitation by radial and tangential foot accelerations. *Med. Eng. Phys.* **2014**, *36*, 502–508. [CrossRef] [PubMed]
5. Skelly, M.M.; Chizeck, H.J. Real-time gait event detection for paraplegic FES walking. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2002**, *9*, 59–68. [CrossRef] [PubMed]

6. Damasevicius, R.; Vasiljevas, M.; Salkevicius, J.; Wozniak, M. Human Activity Recognition in AAL Environments Using Random Projections. *Comput. Math. Methods Med.* **2016**, 4073584. [CrossRef]
7. Mannini, A.; Sabatini, A.M. Machine Learning Methods for Classifying Human Physical Activity from on-Body Accelerometers. *Sensors* **2010**, *10*, 1154–1175. [CrossRef]
8. Lopezmeyer, P.; Fulk, G.D.; Sazonov, E.S. Automatic Detection of Temporal Gait Parameters in Poststroke Individuals. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 594–601. [CrossRef]
9. Sant'Anna, A.; Wickström, N. A symbol-based approach to gait analysis from acceleration signals: Identification and detection of gait events and a new measure of gait symmetry. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 1180–1187. [CrossRef]
10. Bejarano, N.C.; Ambrosini, E.; Pedrocchi, A.; Ferrigno, G.; Monticone, M.; Ferrante, S. A Novel Adaptive, Real-Time Algorithm to Detect Gait Events from Wearable Sensors. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2015**, *23*, 413–422. [CrossRef]
11. Encarna, M.A.M.; Idsart, K.; Erik, A.; Stefan, W.; Martijn, N.; Van, L.R.C.; Van, D.J.H. A novel accelerometry-based algorithm for the detection of step durations over short episodes of gait in healthy elderly. *J. Neuroeng. Rehabil.* **2016**, *13*, 38. [CrossRef]
12. Williamson, R.; Andrews, B.J. Gait event detection for FES using accelerometers and supervised machine learning. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 312–319. [CrossRef]
13. Khandelwal, S.; Wickstrom, N. Gait Event Detection in Real-World Environment for Long-Term Applications: Incorporating Domain Knowledge into Time-Frequency Analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 1363–1372. [CrossRef]
14. Lai, D.T.; Begg, R.K.; Palaniswami, M. Computational intelligence in gait research: A perspective on current applications and future challenges. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 687–702. [CrossRef]
15. Singh, B.N.; Tiwari, A.K. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digit. Signal Process.* **2006**, *16*, 275–287. [CrossRef]
16. Shadevsky, A.; Petrovsky, A. Bio-inspired voice activity detector based on the human speech properties in the modulation domain. In *Information Processing and Security Systems*; Saeed, K., Pejas, J., Eds.; Springer Science Business Media, Inc.: New York, NY, USA, 2005; pp. 43–54.
17. Al-Qazzaz, N.K.; Ali, S.H.B.M.; Ahmad, S.A.; Islam, M.S.; Escudero, J. Selection of Mother Wavelet Functions for Multi-Channel EEG Signal Analysis during a Working Memory Task. *Sensors* **2015**, *15*, 29015–29035. [CrossRef]
18. Salyers, J.B.; Dong, Y.; Gai, Y. Continuous Wavelet Transform for Decoding Finger Movements from Single-Channel EEG. *IEEE Trans. Biomed. Eng.* **2018**. [CrossRef]
19. Beritelli, F.; Capizzi, G.; Sciuto, G.L.; Napoli, C.; Wozniak, M. A novel training method to preserve generalization of RBPNN classifiers applied to ECG signals diagnosis. *Neural Netw.* **2018**, *108*, 331–338. [CrossRef]
20. Mccamley, J.; Donati, M.; Grimpampi, E.; Mazzà, C. An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. *Gait Posture* **2012**, *36*, 316–318. [CrossRef]
21. Khandelwal, S.; Wickström, N. Novel methodology for estimating Initial Contact events from accelerometers positioned at different body locations. *Gait Posture* **2018**, *59*, 278–285. [CrossRef]
22. Pham, M.H.; Elshehabi, M.; Haertner, L.; Del, D.S.; Srulijes, K.; Heger, T.; Synofzik, M.; Hobert, M.A.; Faber, G.S.; Hansen, C. Validation of a Step Detection Algorithm during Straight Walking and Turning in Patients with Parkinson's Disease and Older Adults Using an Inertial Measurement Unit at the Lower Back. *Front. Neurol.* **2017**, *8*, 457. [CrossRef]
23. Glowinski, S.; Blazejewski, A.; Krzyzynski, T. Human Gait Feature Detection Using Inertial Sensors Wavelets. *Wearable Robot. Chall. Trends* **2017**, *16*, 397–401. [CrossRef]
24. Aung, M.S.; Thies, S.B.; Kenney, L.P.; Howard, D.; Selles, R.W.; Findlow, A.H.; Goulermas, J.Y. Automated detection of instantaneous gait events using time frequency analysis and manifold embedding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 908–916. [CrossRef]
25. Forsman, P.M.; Toppila, E.M.; Haeggstrom, E.O. Wavelet analysis to detect gait events. In Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine (EMBC), Minneapolis, MN, USA, 2–6 September 2009; pp. 424–427.

26. Zhou, H.; Ji, N.; Samuel, O.W.; Cao, Y.; Zhao, Z.; Chen, S.; Li, G. Towards Real-Time Detection of Gait Events on Different Terrains Using Time-Frequency Analysis and Peak Heuristics Algorithm. *Sensors* **2016**, *16*, 1634. [CrossRef]

27. Saba, R.; Lockhart, T.E. Towards Real-Time Detection of Freezing of Gait Using Wavelet Transform on Wireless Accelerometer Data. *Sensors* **2016**, *16*, 475. [CrossRef]

28. Khandelwal, S.; Wickström, N. Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. *Gait Posture* **2016**, *51*, 84–90. [CrossRef]

29. Chau, T. A review of analytical techniques for gait data. Part 2: Neural network and wavelet methods. *Gait Posture* **2001**, *13*, 102–120. [CrossRef]

30. Ngui, W.K.; Leong, M.S.; Hee, L.M.; Abdelrhman, A.M. Wavelet Analysis: Mother Wavelet Selection Methods. *Appl. Mech. Mater.* **2013**, *393*, 953–958. [CrossRef]

31. Rafiee, J.; Rafiee, M.A.; Prause, N.; Schoen, M.P. Wavelet basis functions in biomedical signal processing. *Expert Syst. Appl.* **2011**, *38*, 6190–6201. [CrossRef]

32. Mojsilović, A.; Popović, M.V.; Rackov, D.M. On the selection of an optimal wavelet basis for texture characterization. *IEEE Trans. Image Process.* **2000**, *9*, 2043–2050.

33. Soubra, R.; Diab, M.O.; Moslem, B. A mother wavelet selection study for vertical ground reaction force signals. In Proceedings of the 2nd International Conference on Bio-Engineering for Smart Technologies, Paris, France, 30 August–1 September 2017; pp. 1–5.

34. Cho, K.H.; Lee, J.Y.; Lee, K.J.; Kang, E.K. Factors Related to Gait Function in Post-stroke Patients. *J. Phys. Ther. Sci.* **2014**, *26*, 1941–1944. [CrossRef]

35. Orellana, J.N.; Sixto, A.S.; Torres, B.; Cachadina, E.S.; Martin, P.F.; Rosa, F.J. Multiscale time irreversibility: Is it useful in the analysis of human gait? *Biomed. Signal Process. Control* **2018**, *39*, 431–434. [CrossRef]

36. González, R.C.; López, A.M.; Rodriguez-Uría, J.; Alvarez, D.; Alvarez, J.C. Real-time gait event detection for normal subjects from lower trunk accelerations. *Gait Posture* **2010**, *31*, 322–325. [CrossRef]

37. Catalfamo, P.; Moser, D.; Ghoussayni, S.; Ewins, D. Detection of gait events using an F-Scan in-shoe pressure measurement system. *Gait Posture* **2008**, *28*, 420–426. [CrossRef]

38. Sprager, S.; Juric, M.B. Robust Stride Segmentation of Inertial Signals Based on Local Cyclicity Estimation. *Sensors* **2018**, *18*, 1091. [CrossRef]

39. Yan, R. Base Wavelet Selection Criteria for Non-Stationary Vibration Analysis in Bearing Health Diagnosis. Ph.D. Thesis, University of Massachusetts Amherst, Amherst, MA, USA, May 2007.

40. Sota, K.; Uchiyama, Y.; Ochi, M.; Matsumoto, S.; Hachisuka, K.; Domen, K. Examination of factors related to the effect of improving gait speed with functional electrical stimulation intervention for patients with stroke. *PM&R* **2018**, *10*, 798–805. [CrossRef]

41. Chen, G.; Patten, C.; Kothari, D.H.; Zajac, F.E. Gait differences between individuals with post-stroke hemiparesis and non-disabled controls at matched speeds. *Gait Posture* **2005**, *22*, 51–56. [CrossRef]

42. Lilly, J.M.; Olhede, S.C. Higher-Order Properties of Analytic Wavelets. *IEEE Trans. Signal Process.* **2009**, *57*, 146–160. [CrossRef]

Article

Preliminary Clinical Application of Textile Insole Sensor for Hemiparetic Gait Pattern Analysis

Changwon Wang [1,2], Young Kim [3], Hangsik Shin [4] and Se Dong Min [1,2,*]

[1] Department of Medical IT Engineering, Soonchunhyang University, Asan 31538, Korea
[2] Department of Computer Science, Soonchunhyang University, Asan 31538, Korea
[3] Wellness Coaching Service Research Center, Soonchunhyang University, Asan 31538, Korea
[4] Department of Biomedical Engineering, Chonnam National University, Yeosu 59626, Korea
* Correspondence: sedongmin@sch.ac.kr; Tel.: +82-41-530-4871

Abstract: Post-stroke gait dysfunction occurs at a very high prevalence. A practical method to quantitatively analyze the characteristics of hemiparetic gait is needed in both clinical and community settings. This study developed a 10-channeled textile capacitive pressure sensing insole (TCPSI) with a real-time monitoring system and tested its performance through hemiparetic gait pattern analysis. Thirty-five subjects (18 hemiparetic, 17 healthy) walked down a 40-m long corridor at a comfortable speed while wearing TCPSI inside the shoe. For gait analysis, the percentage of the plantar pressure difference (PPD), the step count, the stride time, the coefficient of variation, and the phase coordination index (PCI) were used. The results of the stroke patients showed a threefold higher PPD, a higher step count (41.61 ± 10.7), a longer average stride time on the affected side, a lower mean plantar pressure on the affected side, higher plantar pressure in the toe area and the lateral side of the foot, and a threefold higher PCI (hemi: 19.50 ± 13.86%, healthy: 5.62 ± 5.05%) compared to healthy subjects. This study confirmed that TCPSI is a promising tool for distinguishing hemiparetic gait patterns and thus may be used as a wearable gait function evaluation tool, the external feedback gait training device, and a simple gait pattern analyzer for both hemiparetic patients and healthy individuals.

Keywords: conductive textile; gait; stroke; hemiparetic; real-time monitoring

1. Introduction

The post stroke gait disturbance is one of the major complications that requires a long-term rehabilitation and limits the patient's activities of daily living [1]. The improved gait function is an important factor in returning to social life and thus is an essential goal of rehabilitation therapy [2]. While 60% to 80% of stroke patients regain their independent ambulation function, many exhibit hemiparetic gait pattern for the rest of their lives due to unilateral neuromuscular weakness leading to gait asymmetry [2–4]. The critical consequences of the impaired walking ability after a stroke include a reduction in gait speed, shorter step and stride lengths, and an increased fall risk [5,6]. These residual deficits are mainly caused by muscle weakness and imbalance, decreased weight support on the affected side, and asymmetrical intralimb coordination [7,8]. For the patients with these deviations to regain walking abilities, clinical treatments commonly rely on traditional rehabilitation approaches such as neuromuscular re-education, lower limb strength training, and balance training for weight shifting and gait pattern training. While the patients and healthcare providers strive to seek for more effective gait therapy methods, there are not many long-term remedies or set devices for facilitating these treatment options in clinical settings.

With fast growing advances in technology, various forms of wearable measuring-recording devices and sensors have been developed for the health care system. For gait rehabilitation, force sensitive resistor sensors (FSR) are one of the most commonly used sensor types for analyzing gait pattern or

plantar pressure, but is known to have the disadvantage of being deformed in time as a response to repetitive force applied. Another drawback of the FSR sensor is the need to be constantly calibrated (good for approximately 100 uses) and its inability to distinguish between load changes in respect to the weight bearing level difference generated during walking [9,10]. Another popular type considered for gait biofeedback sensors is piezoelectric material, which has a high impedance, low noise, and susceptibility to electrical interference. However, its weakness is having a limited pressure range and non-linearity in repeated signal output [9]. According to a study by Aqueveque et al. (2018), capacitive sensors can be used for developing an in-shoe device that measures plantar pressure [9]. A capacitive sensor basically consists of two conductive plates that are separated by dielectric material. Modifying the distance between the conductive plates generates a variation of the capacitance and this change in capacitive value can be interpreted as plantar pressure change [9]. This study developed a capacitive insole sensor to analyze plantar pressure and with the aim to provide real-time feedback on the patients' gait rehabilitation processes.

The assessment of gait rehabilitation is usually made using clinical scales, but many of the neuromuscular disabilities are still being evaluated manually with analog scales [11–15]. Thus, it is difficult to verify the accuracy in measurement methods, and both the intra- and inter-personal evaluation results have low reliability. A practical device that can quantitatively analyze the characteristics of hemiparetic gait pre and post rehabilitation is substantially needed. A recent study by Ngueleu et al. (2019) stated that, although using pressure-sensing insoles for identification of the step count is a promising approach in gait analysis, the accuracy of the activity monitors in step counting remains limited and that there appears to be no consensus for optimal positioning and the number of sensors for insoles [10].

In stroke rehabilitation, the basic gait analysis parameters include the plantar pressure distribution, the step count, the stride time, velocity (or gait speed), the center of pressure (CoP), the coefficient of variation (CV) and the phase coordination index (PCI) [16–23]. Among these indices, plantar pressure and PCI have shown to be relatively more sensitive measures in analyzing the bilateral coordination or asymmetry of locomotion and balance, which are meaningful variables [16,17,24,25]. According to previous studies, the wearable gait analysis system [26], such as with insole pressure sensors, can analyze gait parameters including gait velocity, cadence, stride length, step length, stride time, single limb support and stance (coordination function) by analyzing pressure between the foot plantar surface and the shoe insole. Therefore, this study developed an insole-type wearable pressure sensor.

In several other previous studies, the insole-type pressure sensor and smart shoes were developed for gait analysis and smart phone applications enabled real-time monitoring of the activities being carried out [27,28]. However, these devices are expensive, and the bio-signals collected and analyzed were reported to be not as accurate for use in clinical research [29]. To solve the cost-related problem of the sensors, researchers tried to apply pressure sensors for a gait analysis system, but they were found to be short-lived, making them still expensive for both users and researchers. Based on these advantages, researchers are conducting studies on an insole type textile pressure sensor considering the properties of conductive textile that is user friendly and inexpensive [30–34]. However, studies on the clinical application of the wearable textile insole sensor are scarce.

Therefore, this study developed a wearable textile capacitive pressure-sensing insole to test its feasibility in analyzing hemiparetic gait patterns and distinguishing its characteristics from normal gait.

2. Materials and Methods

2.1. Textile Capacitive Pressure Sensing Insole (TCPSI)

The textile capacitive pressure sensing insole (TCPSI) was designed based on Coulomb's law as shown in Figure 1. The capacitance is determined by how much electricity is collected between the two conductive plates. The wider the gap between the plates, the higher the capacitance. 1 Farad (F) is

equal to the capacitor charged at 1 C when a voltage of 1 V is applied. Equation (1) is a formula for calculating the parallel capacitance value, and the C value is derived using the above formula.

$$C = Q/V = \varepsilon(A(\text{width} \times \text{length}))/d \tag{1}$$

where, d means the distance between two plates, A means the area of plates (width $\times$ length), ε is the permittivity material between the plates.
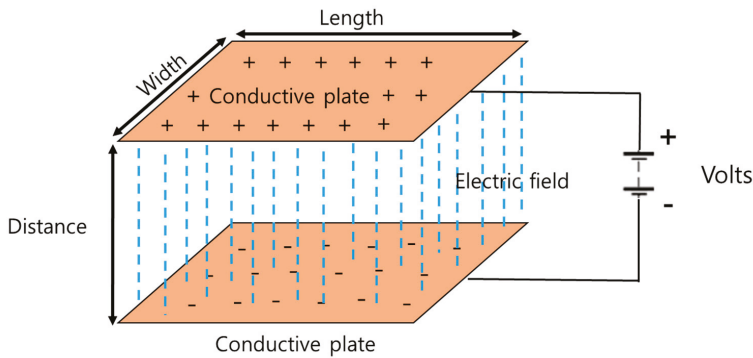
Figure 1. The structure of parallel capacitor used for the textile capacitive pressure sensing insole (TCPSI).

For TCPSI, a parallel capacitance measurement method was applied as illustrated in Figure 2a. The W-290-PCN model (Ajin Electronics, Gangseo, Busan, ROK) conductive textile was used and its characteristics have been described in Table 1. The textile consists of polyester, sequentially plated with nickel, copper, and nickel. Non-conductive rubber with a thickness of 3 mm was placed between a sensor layer and a ground layer, and another two layers of non-conductive woven textiles were placed on the top and bottom of all other layers as a cover and shock absorber. Figure 2b shows the size (2 $\times$ 2 cm^2) and locations of the 10 channels in each insole. Sensor channel 1 was located in the medial upper corner of the insole to detect the pressure mainly under the big toe area (1st, 2nd distal phalanges) and channel 2 was located in the lateral upper corner of the insole to detect the pressure under the 3rd 4th, and 5th distal phalanges. These channels were to detect the plantar pressure in the forefoot during the pre-swing and toe-off phases. Channels 3 and 4 were located below the channels 1 and 2 to detect the pressure under the proximal interphalangeal joint area. Channels 5 and 6 were located in the metatarsal-phalangeal joint area. Channels 7, 8, 9 and 10 were located in the heel area.

Table 1. Characteristics of the W-290-PCN model.

	Value	Structure
Weight (g/m^2)	81 ± 5	
Width (mm)	1100 ± 5	
Density (g/m^3)	188 ± 5	
Thickness (mm)	0.10 ± 0.01	
Shielding effectiveness (dB)	70	
Surface resistance (Ω)	<0.05	

(**a**) Textile capacitive pressure sensing insole

(**b**) Sensor location with anatomical reference
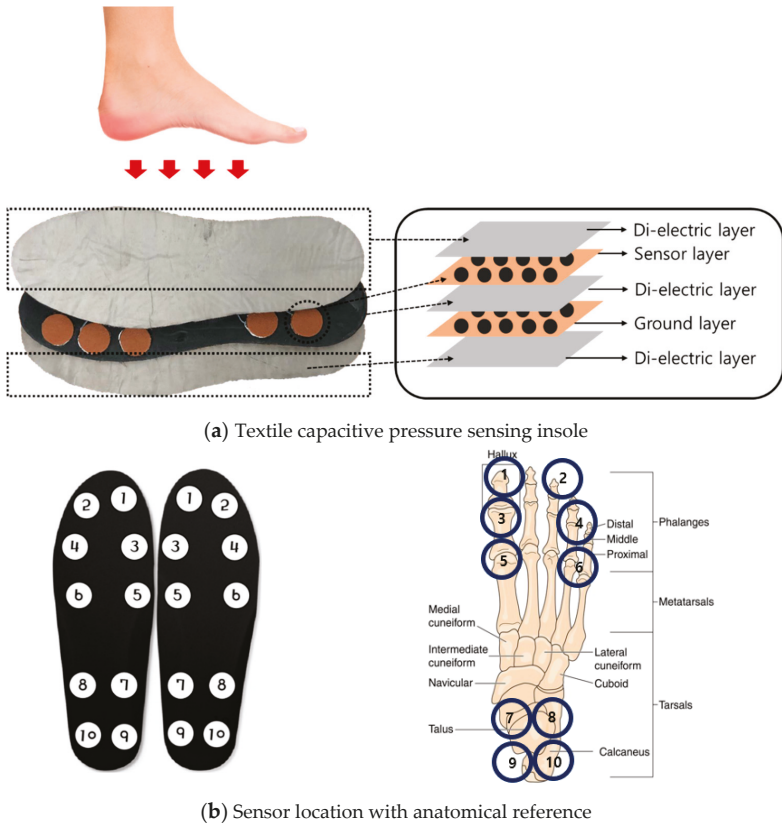
Figure 2. The structure of the textile capacitive pressure sensing insole (TCPSI) and sensor location, (**a**) textile capacitive pressure sensing insole, (**b**) sensor location.

2.2. Gait Monitoring System

Figure 3 shows the system architecture of our proposed gait monitoring system. The hardware collects data from TCPSI and transfers them to the software via Bluetooth communication. The software receives, saves, and analyzes the data using the C#-based gait monitoring system and matlab2018a.
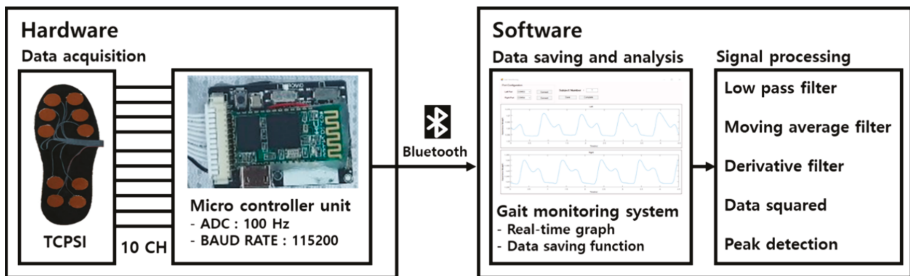
Figure 3. The schematic diagram of the gait monitoring system.

2.2.1. Hardware

For the micro controller unit (MCU), the STM32F103 model developed by STMicroelectronics, USA was used. The response time of MCU was set up at 100 ms and the operation voltage was 3.7 V. The MCU pre-scaler was set up at 3 and the period was 400 based on 72 MHz main clock frequency. Based on this, the sensing period was calculated and it was set at 16 us by using I^2C communication.

For the capacitance measurement function, the MPR121QR2 sensor (Freescale Inc., Austin, TX, USA) was used; it converts capacitance to digital values. This sensor was used in our previous study [35] testing the feasibility of our capacitive insole sensor and showed a high correlation ($R^2 >$ 0.90, $p < 0.05$) with F-scan (Tekscan Inc., Boston, MA, USA). The MPR121QR2 sensor has a resolution of 0.01 pF and the measurement range covers 10 pF to 2000 pF. The data were sampled at 100 Hz [36]. Normalization was performed for all 10 channels of the insole sensor before signal processing. The standard deviation values between the data collected from the 10 channels were lower than 0.03 pF.

The HM-13 chip was used for Bluetooth communication and the baud rate was set to 115,200, the non-parity bit was 0, stop bit was 1.

The printed circuit board (PCB) was developed containing all sensors as presented in Figure 4a,b shows the MCU, MPR121QR2, and Bluetooth communication part of the developed PCB schematics.

2.2.2. Software

A gait monitoring application was designed based on C# language. It visualizes the plantar pressure changes in real time by using capacitance values from the insole sensor. The unit of pressure in our sensing system is capacitance itself, that is, the difference value between pre-pressure (baseline; no pressure applied) and post-pressure at the time of a heel strike (HS) and toe-off (TO). The visualization was to provide feedback on plantar pressure distribution during walking. The software saves the data and exports them in a text file or an excel file. All data are stored as an integer and the software checks in real time whether the data are being received from all 20 channels, 10 channels from each foot.

Figure 5 below is showing the gait data of a left hemiparetic male patient. The graph on top presents the real time gait signals, which are based on the capacitance values received from the left foot (insole), and the graph on the bottom visualizes the data from the right foot. Each of the high peaks at the time of the heel strike and toe-off are marked and the average values of the peaks are shown in a dotted line.

(**a**)

Figure 4. *Cont.*

STM32F103 MCU

[4] BATT_ADC
[4] CHRG_STAT

R11 OPEN
R12 OPEN
CHRG_LED
POWER_LED

[4] VBT_EN
[2] BT_RESET

[2] BT_UART_TX   BT_UART_TX
[2] BT_UART_RX   BT_UART_RX
[4] USB_DN
[4] USB_DP

7 PA0
8 PA1
9 PA2/USART2_TX
10 PA3/USART2_RX
11 PA4
12 PA5
13 PA6
14 PA7
20 PA8
21 PA9/USART1_TX
22 PA10/USART1_RX
23 PA11/USB_DN
24 PA12/USB_DP
25 PA13
28 PA14
29 PA15

PB0 15
PB1 16
PB2 17
PB3 30
PB4 31
PB5 32
PB6/I2C1_SCL 33
PB7/I2C1_SDA 34

SENS_RST
BT_PIO1
ACC_INT1
ACC_INT2
SENS_INT
I2C_SCL
I2C_SDA

CLK_8M_IN
CLK_8M_OUT
2 OSC_IN
3 OSC_OUT

NRST 4
BOOT0 35
VSSA 5

MCU_NRESET
MCU_BOOT0

VDD_3V3

R47 R_1005J_0R   6 VDDA

C29   C30
C_1005_0.1uF_16V   C_1005_0.1uF_16V

19 VDD
27 VDD
1 VDD

VSS 18
VSS 26
VSS 36

STM32F103TBU6

VDD_3V3

C16   C26   C28
C_1005_0.1uF_16V   C_1005_0.1uF_16V   C_1005_0.1uF_16V

(**b**)

MPR121QR2 sensor

C12   C_1005_0.1uF_16V

9 VDD
13 NC

SENS_INT R33   R_1005J_33R   15 /CHANGE
I2C_SCL   R35   R_1005J_33R   14 SCL
I2C_SDA   R36   R_1005J_33R   11 SDA
SENS_RST R38   R_1005J_33R   12 /RESET
                              10 MODE
                              8 VSS
                              21 EXPAD

KEY0 7   R29 R_1005J_0R   1
KEY1 6   R30 R_1005J_0R   2
KEY2 5   R34 R_1005J_0R   3
KEY3 4   R31 R_1005J_0R   4
KEY4 3   R37 R_1005J_0R   5
KEY5 2   R39 R_1005J_0R   6
KEY6 1   R77 R_1005J_0R   7
KEY7 20 R78 R_1005J_0R   8
KEY8 19 R79 R_1005J_0R   9
KEY9 18 R80 R_1005J_0R   10
KEY10 17 R81 R_1005J_0R   11
KEY11 16 R82 R_1005J_0R   12
                            13

J3

(**c**)

HM-13 sensor

BT_UART_RX
BT_UART_TX

R54 R_1005J_33R   1 UART_RTS
R27 R_1005J_33R   2 UART_TX
R60 R_1005J_33R   3 UART_CTS
R28 R_1005J_33R   4 UART_RX
5 NC
6 NC
7 NC
8 NC

PIO0 16 R18 R_1005J_33R
PIO1 15 R20 R_1005J_33R   BT_PIO1 [1]
PIO2 14 R22 R_1005J_33R
PIO3 13 R51 R_1005J_33R
GND   12
RESETB 11 R53 OPEN   BT_RESET [1]
NC   10
VCC   9   VBT_3V3

C10
C_1005_0.1uF_16V

(**d**)

Figure 4. The structure of the printed circuit board (PCB) board, (**a**) PCB configuration, (**b**) STM32F103 MCU, (**c**) MPR121QR2 capacitance to digital converter, (**d**) HM-13 Bluetooth module schematic diagram.
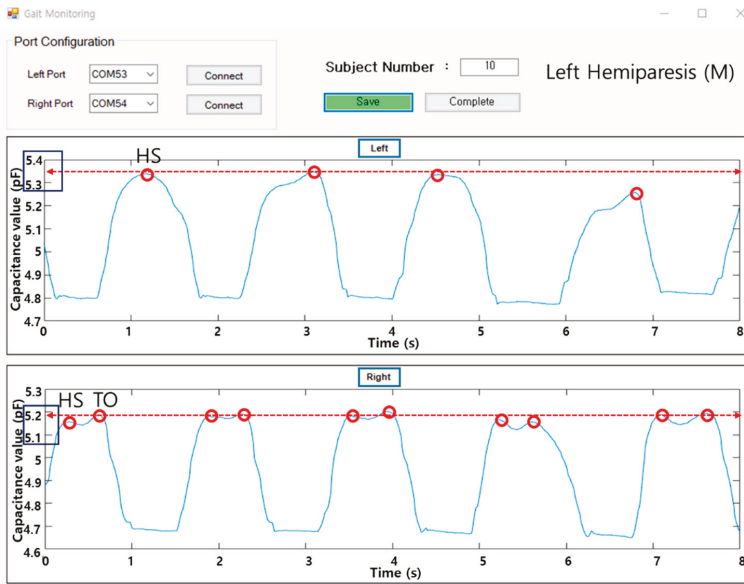
(HS: Heel Strike, TO: Toe-Off)

Figure 5. Gait monitoring application (left hemiparetic male patient data sample).

2.3. Data Analysis

2.3.1. Percentage of Plantar Pressure Difference (PPD)

The PPD was calculated by using Equation (2) as expressed below [17].

$$\text{Percentage of plantar pressure difference } (\%) = \frac{2|Pressure_{Left} - Pressure_{Right}|}{Pressure_{Left} + Pressure_{Right}} \times 100 \tag{2}$$

where, $Pressure_{Left}$ and $Pressure_{Right}$ refers to the summed pressure data on each foot as shown in Figure 6.
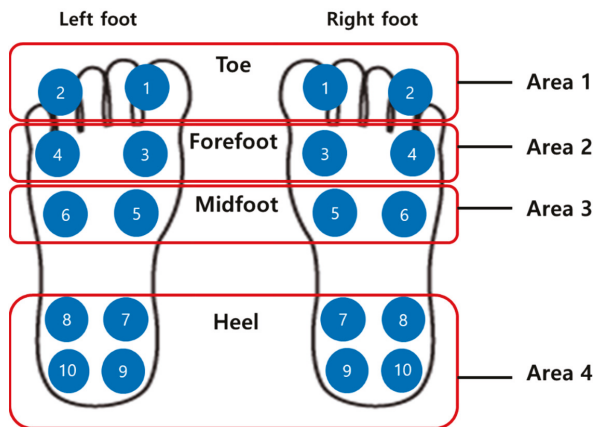
Figure 6. The percentage of plantar pressure difference calculation method diagram.

2.3.2. Plantar Pressure Distribution Comparison Analysis

For analysis of the weight-shifting patterns during gait, the plantar pressure data received from all 10 sensors on each foot were summed and averaged for medial-lateral and right-left weight shifting characteristics comparison between hemiparetic patients and healthy adults as illustrated in Figure 7.
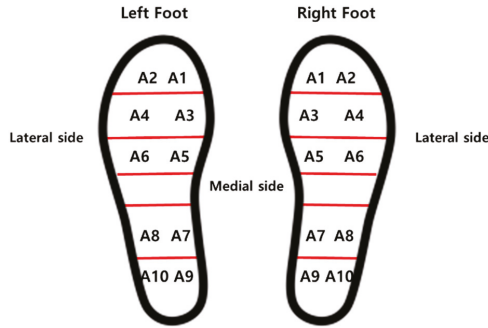
Figure 7. Sensor locations for plantar pressure detection.

2.3.3. Signal Processing

Four stages of signal pre-processing were performed consecutively as shown in Figure 8. The low pass filter with a cut-off frequency of 3 Hz was first applied followed by a moving average filter with 5 point. The gait signal was smoothed and noise was removed. The first-order differential filter was then applied to sharpen the gradients on the y-axis, making the high peaks higher and the low peaks lower. This process made the slope of the original signal steeper and more prominent as the change value on the y-axis increased. This study developed an algorithm to detect the highest peaks at the time of the heel strike and toe-off, at an interval of 300 ms. Considering person-to-person gait pattern difference, the interval range was selected based on our previous gait analysis study results using the prototype insole sensor in healthy young adults [35]. The heel strike and toe off were detected within 300 ms intervals.

This was designed based on a local maxima algorithm.
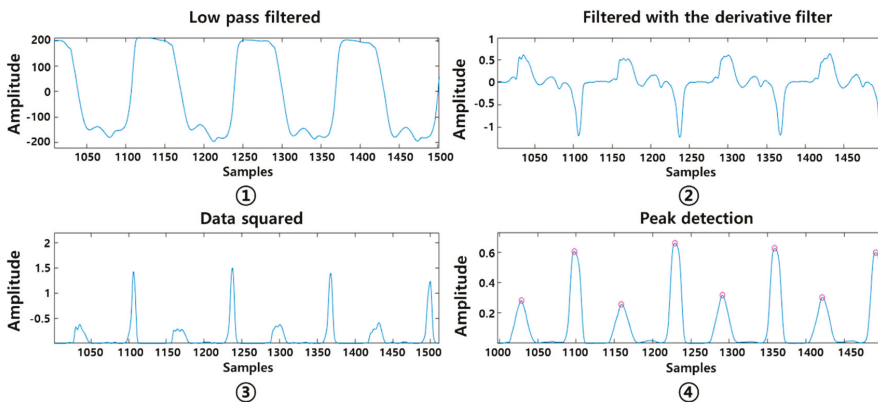
Figure 8. Signal processing order.

Each initial contact (heel strike) of one side of the leg was marked and calculated for the step count and a stride time was defined by the time between two consecutive heel strikes of the same foot [37].

The coefficient of variation (CV) of the stride time was calculated [38].

PCI was calculated by first defining the stride time of one foot as a gait cycle (360°; from one leg's heel strike to the next heel strike) and the phase (φ ideally 180°; time point of contralateral heel strike). The phase of the ith stride (φ_i, in degrees) is defined by normalizing the step time with respect to the stride time Equation (3). The sum of the CV of phase (φ_CV) and the mean absolute difference between phase and 180° $\left(\varphi_{ABS}\right)$ was defined as PCI [21,22]. φ_i is an index that evaluates the symmetry of bilateral stepping phases and is the distance between one heel strike and the next of the opposite leg calculated in degrees, ideally 180° for successful walking. The step time of each leg should normally be equal to the half the gait cycle, but for the asymmetric cases t_{S_i} refers to the shorter swing phase and t_{L_i} for the longer swing phase; the time of heel strike (of step i) during short and long swing phases, respectively. As shown below in the equation, φ_ABS indicates the balance between the two feet Equation (4). φ_CV refers to the coefficient of variation of φ_i which represents a consistency of both feet during walking Equation (5).

$$\varphi_i = 360° \times \frac{t_{S_i} - t_{L_i}}{t_{L(i+1)} - t_{L_i}} \tag{3}$$

$$\varphi\_ABS = \frac{\varphi\_ABS}{180} \times 100 \tag{4}$$

$$\varphi' = \frac{1}{N}\sum_{i=1}^{n}\varphi_i \; \delta = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(\varphi' - \varphi_i)^2}, \; \varphi\_CV = \frac{\delta}{\varphi'} \tag{5}$$

2.4. Experimental Protocol

For this study, a total of 35 subjects were recruited and approved by the institutional review board (IRB No. SCH2016-130). Eighteen hemiparetic patients (12 males, 6 females) with an average age of 63.94 ± 8.75 years and 17 healthy adults (7 males, 10 females) with an average age of 56.47 ± 17.19 years participated. Three of the hemiparetic patients were diagnosed with bilateral hemiparesis, 6 with left hemiparesis, and 9 with right hemiparesis (Tables 2 and 3). The patients who participated in this study were those diagnosed with a stroke between July 2016 and February 2017.

The experiment took place at a local hospital under close supervision by a physical therapist. Each participant was asked to walk down a 20-m long corridor back and forth (1 round trip) at a comfortable speed while wearing the given sneakers with TCPSI inside. The TCPSI sensor was prepared according to the subject's shoe size (all sizes between 235 to 285 mm). Furthermore, TCPSI was developed to fit the size of the shoe and was configured to collect and save the data in real time while the subjects walked at a comfortable speed. For the patient subjects who complained of pain or discomfort during the experiment, a resting time of more than 5 min was given.

The experiment environment is illustrated in Figure 9. All subjects were adequately informed about the experiment procedure and the experiment was conducted after obtaining written consent from all participants.

The collected gait data were analyzed by using the PPD, step count, stride time, CV and PCI of hemiparetic patients and compared their gait coordination functions with those of healthy adults.

Figure 9. Experimental environment.

Table 2. General characteristics of hemiparetic subjects.

Subject	Sex	Age	Height	Weight	BMI	Shoe Size	Hemi Side	Onset
1	M	54	174	80	26.42	260	L	12.2016 2016.12
2	M	65	173	74	24.72	270	R	10.2016
3	M	76	170	90	31.14	270	R	11.2016
4	M	59	167	73	26.17	255	L	11.2016
5	M	72	177	75	23.93	275	R	10.2016
6	M	63	172	70	23.66	275	R	12.2016
7	M	75	175	70	22.85	260	L	09.2016
8	M	63	176	64	20.66	275	B	07.2016
9	F	64	150	45	20.00	235	B	12.2016
10	M	60	178	79	24.93	295	R	01.2017
11	M	78	170	67	23.18	265	R	08.2016
12	F	63	163	67	25.21	245	R	01.2017
13	F	60	150	38	16.88	225	L	07.2016
14	M	68	168	58	20.54	260	L	01.2017
15	F	47	175	75	24.48	230	L	01.2017
16	M	49	162	90	34.29	255	B	02.2017
17	F	73	145	50	23.78	250	L	02.2017
18	F	62	150	45	20.00	240	R	12.2016
AVG		63.94	166.39	67.22	24.05	257.78		
SD		8.75	10.69	14.93	4.05	18.08		

RH: right hemiparesis, LH: left hemiparesis, BH: bilateral hemiparesis, HA: healthy adults.

Table 3. General characteristics of healthy subjects.

Subject	Sex	Age	Height	Weight	BMI	Shoe Size
1	F	34	168	54	19.13	260
2	F	31	161	48	18.51	235
3	M	27	178	75	23.67	265
4	M	48	168	63	22.32	260
5	M	30	178	70	22.09	265
6	M	45	167	65	23.30	260
7	M	75	175	75	25.95	260
8	F	57	159	54	21.35	235
9	F	68	150	51	22.66	225
10	F	58	163	80	30.11	245
11	M	71	165	68	24.97	250
12	F	71	158	62	24.83	245
13	F	71	157	68	27.58	245
14	F	64	156	54	22.18	230
15	M	77	168	80	28.34	260
16	F	69	155	66	27.47	235
17	F	64	162	72	27.43	240
AVG		56.47	164	65	24.22	247.94
SD		17.19	8.04	10.01	3.26	13.11

3. Results

3.1. The Results of Plantar Pressure Analysis

3.1.1. Percentage of Plantar Pressure Difference (PPD)

Table 4 shows the percentage of plantar pressure difference in each subject group. The right hemiparesis had PPD of 13.43% in the toe area, 12.74% in the forefoot area, 12.19% in the midfoot area, and 26.73% in the heel area. The left hemiparesis had 28.27%, 21.38%, 15.25% and 18.86%, respectively. The bilateral hemiparesis had 9.91%, 14.10%, 9.17%, 9.52% and healthy adults had 5.26%, 4.15%, 4.59%

and 6.06%, respectively. Figure 10 shows the percentage of plantar pressure difference results of each subject group.

Table 4. Percentage of the plantar pressure difference (unit: %).

Location	RH	LH	BH	HA
Toe	13.43	28.27	9.91	5.26
Forefoot	12.74	21.38	14.10	4.15
Midfoot	12.19	15.25	9.17	4.59
Heel	25.89	18.86	9.52	6.06
AVG	16.06	20.95	10.68	5.01
SD	6.57	5.49	2.30	0.83

RH: right hemiparesis, LH: left hemiparesis, BH: bilateral hemiparesis, HA: healthy adults.
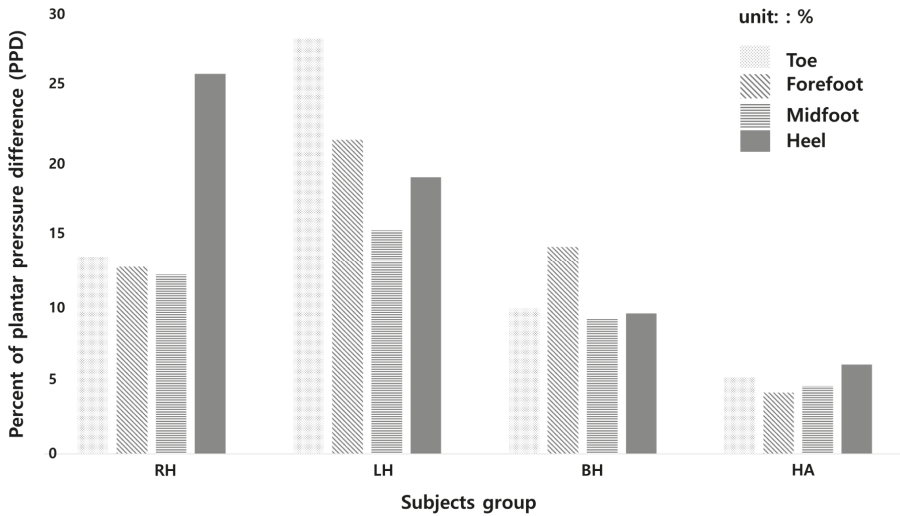
Figure 10. The results of percent of the plantar pressure difference. (RH: right hemiparesis, LH: left hemiparesis, BH: bilateral hemiparesis, HA: healthy adults).

3.1.2. Plantar Pressure Comparison Analysis

The average plantar pressure on the left foot (unaffected) of the right hemiparetic patients was 6.11 ± 0.08 pF (capacitance difference value) in the lateral side of the foot and 6.50 ± 0.27 pF in the medial side. The right foot (affected) showed 5.17 ± 0.46 pF and 5.45 ± 0.29 pF, respectively (Table 5).

Table 5. Average plantar pressure of right hemiparetic patients ($n = 9$) (unit: pF).

Location	Left Foot				Right Foot			
	Medial Side		Lateral Side		Medial Side		Lateral Side	
	Channel	Pressure	Channel	Pressure	Channel	Pressure	Channel	Pressure
Toe	A2	6.59	A1	6.23	A1	5.63	A2	5.81
Forefoot	A4	6.47	A3	6.10	A3	5.51	A4	5.58
Midfoot	A6	6.14	A5	6.03	A5	5.22	A6	5.56
Heel	A8	6.88	A7	6.04	A7	5.03	A8	5.21
Heel	A10	6.40	A9	6.13	A9	4.47	A10	5.10
AVG		6.50		6.11		5.17		5.45
SD		0.27		0.08		0.46		0.29

Table 6 shows the average plantar pressure of left hemiparetic patients. The lateral side of the left foot (affected) showed 5.11 ± 0.18 pF of pressure and the medial side had 5.01 ± 0.16 pF. The unaffected right foot had 6.14 ± 0.18 pF on the lateral side and 6.14 ± 0.11 pF on the medial side.

In Table 7, the average plantar pressure on the lateral side of the left foot was 5.62 ± 0.33 pF and of the medial side was 5.36 ± 0.38 pF in bilateral hemiparetic patients. The medial side of the right foot showed 5.42 ± 0.57 pF and the lateral side 5.66 ± 0.33 pF.

Table 6. Average plantar pressure of left hemiparetic patients (*n* = 6) (unit: pF).

	Left Foot				Right Foot			
Location	**Medial Side**		**Lateral Side**		**Medial Side**		**Lateral Side**	
	Channel	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**
Toe	A2	5.27	A1	5.41	A1	6.41	A2	6.17
Forefoot	A4	5.01	A3	5.12	A3	6.16	A4	6.21
Midfoot	A6	5.03	A5	4.98	A5	5.95	A6	5.95
Heel	A8	4.87	A7	4.98	A7	6.01	A8	6.19
Heel	A10	4.87	A9	4.94	A9	6.16	A10	6.17
AVG		5.01		5.11		6.14		6.14
SD		0.16		0.18		0.18		0.11

Table 7. Average plantar pressure of bilateral hemiparetic patients (*n* = 3) (unit: pF).

	Left Foot				Right Foot			
Location	**Medial Side**		**Lateral Side**		**Medial Side**		**Lateral Side**	
	Channel	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**
Toe	A2	5.51	A1	5.18	A1	4.68	A2	6.07
Forefoot	A4	5.96	A3	5.40	A3	6.20	A4	5.16
Midfoot	A6	5.13	A5	5.64	A5	5.12	A6	5.60
Heel	A8	5.03	A7	5.96	A7	5.56	A8	5.67
Heel	A10	5.17	A9	5.91	A9	5.52	A10	5.82
AVG		5.36		5.62		5.42		5.66
SD		0.38		0.33		0.57		0.33

Table 8 shows the average plantar pressure of healthy adults. The lateral side of the left foot had 6.39 ± 0.43 pF, the medial side had 6.24 ± 0.38 pF, the lateral side of the right foot had 6.41 ± 0.42 pF, and the medial side had 6.39 ± 0.29 pF.

Table 8. Average plantar pressure of healthy adults (*n* = 17) (unit: pF).

	Left Foot				Right Foot			
Location	**Medial Side**		**Lateral Side**		**Medial Side**		**Lateral Side**	
	Channel	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**	**Channel**	**Pressure**
Toe	A2	6.61	A1	6.25	A1	6.66	A2	6.34
Forefoot	A4	6.43	A3	6.60	A3	6.30	A4	6.20
Midfoot	A6	5.65	A5	5.83	A5	5.72	A6	6.08
Heel	A8	6.12	A7	6.30	A7	6.67	A8	6.52
Heel	A10	6.41	A9	6.99	A9	6.68	A10	6.83
AVG		6.24		6.39		6.41		6.39
SD		0.38		0.43		0.42		0.29

3.2. The Results of Gait Parameter Analysis

3.2.1. Step Count

The average step count during 40 m walk in right hemiparetic patients was 46.83 ± 13.89 steps and for the left hemiparetic patients was 38.44 ± 4.53 as shown in Table 9. The bilateral hemiparetic patients was 40.66 ± 16.77 and healthy adults was 31.29 ± 8.97, respectively.

Table 9. The results of step count detection in hemiparetic patients and healthy adults.

Group	Step Count
RH	46.83 ± 13.89
LH	38.44 ± 4.53
BH	40.66 ± 16.77
HA	31.29 ± 8.97

3.2.2. Stride Time

The average stride time of the left foot in the right hemiparetic patients was 1.66 ± 0.20 s and that of the right foot was slowed to 1.77 ± 0.19 s as shown in Table 10. The average time difference between the two feet was 0.11 ± 0.04 s.

Table 10. Average stride time of right hemiparetic patients ($n = 9$) (unit: seconds).

Subject No.	Left Foot	Right Foot	Time Difference (Right–Left)
Subject 2	1.39	1.48	0.09
Subject 3	1.75	1.88	0.13
Subject 5	1.51	1.63	0.12
Subject 6	1.71	1.81	0.10
Subject 10	1.86	1.93	0.07
Subject 11	1.98	2.05	0.07
Subject 12	1.56	1.75	0.19
Subject 16	1.40	1.52	0.12
Subject 18	1.78	1.94	0.16
AVG	1.66	1.77	0.11
SD	0.20	0.19	0.04

The average stride time of left hemiparetic patients showed to be slowed to 1.75 ± 0.42 s on the left (affected) side and 1.64 ± 0.40 s on the right foot as shown in Table 11. The time difference between the two feet was 0.14 ± 0.04 s in average.

Table 11. Average stride time of left hemiparetic patients ($n = 6$) (unit: seconds).

Subject No.	Left Foot	Right Foot	Time Difference (Left–Right)
Subject 1	1.58	1.46	0.12
Subject 4	1.39	1.48	0.09
Subject 7	1.54	1.40	0.14
Subject 13	1.65	1.42	0.23
Subject 14	2.58	2.46	0.12
Subject 17	1.79	1.62	0.17
AVG	1.75	1.64	0.14
SD	0.42	0.40	0.04

The average stride time of the left foot in bilateral hemiparetic patients was 1.48 ± 0.06 s and that of the right foot was 1.52 ± 0.03, and the time difference between the two feet was 0.04 ± 0.04 as illustrated in Table 12.

Table 12. Average stride time of bilateral hemiparetic patients ($n = 3$) (unit: seconds).

Subject No.	Left Foot	Right Foot	Time Difference
Subject 8	1.44	1.54	0.10
Subject 9	1.46	1.48	0.02
Subject 15	1.56	1.54	0.02
AVG	1.48	1.52	0.04
SD	0.06	0.03	0.04

Table 13 shows the average stride time of healthy adults and the time difference between the two feet. The stride time of the left foot was 1.34 ± 0.12 s and of the right was 1.35 ± 0.12 s. The average stride time difference between the two feet was 0.02 ± 0.03 s.

Table 13. Average stride time of healthy adults (*n* = 17) (unit: seconds).

Subject No.	Left Foot	Right Foot	Time Difference (Right–Left)
Subject 19	1.21	1.21	0.09
Subject 20	1.60	1.60	0.13
Subject 21	1.32	1.32	0.00
Subject 22	1.25	1.25	0.00
Subject 23	1.27	1.27	0.00
Subject 24	1.18	1.18	0.00
Subject 25	1.29	1.32	0.03
Subject 26	1.27	1.27	0.00
Subject 27	1.41	1.41	0.00
Subject 28	1.34	1.33	0.01
Subject 29	1.39	1.38	0.01
Subject 30	1.49	1.49	0.00
Subject 31	1.43	1.44	0.01
Subject 32	1.17	1.18	0.01
Subject 33	1.34	1.42	0.08
Subject 34	1.60	1.60	0.00
Subject 35	1.36	1.35	0.01
AVG	1.34	1.35	0.02
SD	0.12	0.12	0.03

The stride time CV of each group is presented in Table 14. The right hemiparetic patients had a CV of 13.33 ± 4.52% (left foot) and 23.88 ± 3.29% (right foot). The left hemiparetic patients had 20.83 ± 6.70%, 12.00 ± 3.34%, the bilateral hemiparetic patients had 19.33 ± 2.51% and 18.00 ± 1.73%, respectively. The healthy adults had 2.82 ± 1.91% and 2.47 ± 1.62%, respectively.

Table 14. Coefficient of Variation (CV) of stride time in each group (unit: %).

Variables	Left Foot	Right Foot
RH	13.33 ± 4.52	23.88 ± 3.29
LH	20.83 ± 6.70	12.00 ± 3.34
BH	19.33 ± 2.51	18.00 ± 1.73
HA	2.82 ± 1.91	2.47 ± 1.62

RH: right hemiparesis, LH: left hemiparesis, BH: bilateral hemiparesis, HA: healthy adults.

Figure 11 below shows the results of average stride time CV of each subject group. The CV of the left foot in left hemiparetic patient group was 22% and that of the right foot was 12%. The CV of the left foot in right hemiparetic patient group was 13% and that of the right foot was 24%. For bilateral hemiparetic patient group, the CV of the left foot was 19% and that of the right was 18%. The healthy group showed to have a CV of 2% on both sides of the foot.

3.2.3. Phase Coordination Index (PCI)

The PCI and bilateral stepping phase symmetry were calculated to evaluate gait coordination of hemiparetic patients and healthy adults as illustrated in Figures 12 and 13, respectively. The average value of PCI in hemiparetic patients was 19.50 ± 13.86% and the φ_CV value was 0.20 ± 0.17%, the φ_ABS was 41.09 ± 37.83° and the φ average was 180.88 ± 56.78°. The PCI value of healthy subjects was 5.62 ± 5.06%, the φ_CV was 0.12 ± 0.11%, the φ_CV was 9.89 ± 8.99°, and the φ average was 176.64 ± 13.14° as summarized in Table 15.
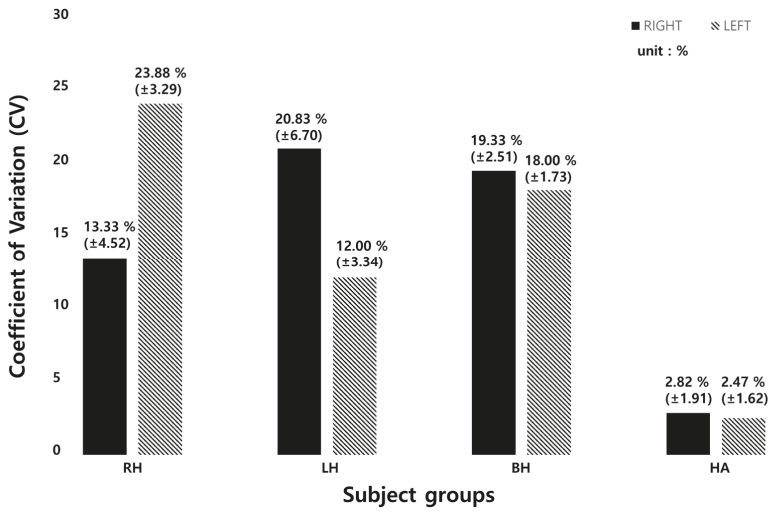
Figure 11. Average stride time CV of each subject group. (RH: right hemiparesis, LH: left hemiparesis, BH: bilateral hemiparesis, HA: healthy adults).
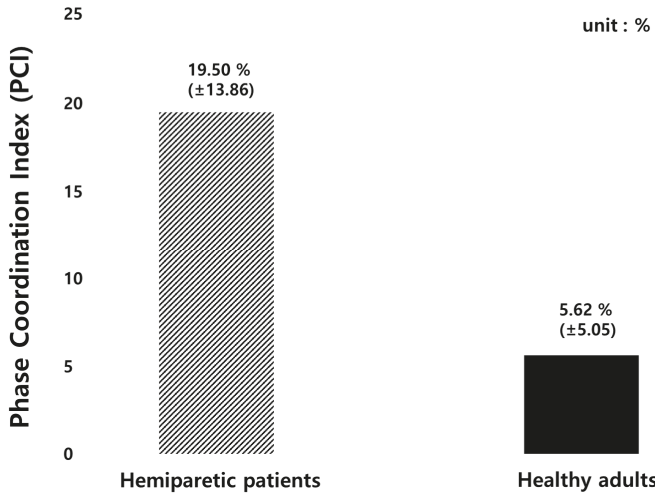

Figure 12. The results of PCI value comparison between hemiparetic patients and healthy adults.

Figure 13 shows a sample comparison result in a scatter plot of bilateral stepping phase symmetry between a hemiparetic (subject #14; left hemi) patient and a healthy adult (subject #34). The hemiparetic patient's bilateral stepping phase was observed to be widely scattered from the standard normal value of 180° compared to that of the healthy subject.

Table 15. The results of the phase coordination index (PCI) detection.

Variables	Hemiparetic ($n = 18$)	Healthy ($n = 17$)
Phase Coordination Index (%)	$19.50 \pm 13.86\%$	$5.62 \pm 5.05\%$
φ_CV (%)	$0.20 \pm 0.17\%$	$0.12 \pm 0.11\%$
φ_ABS (deg°)	$41.09 \pm 37.83°$	$9.89 \pm 8.99°$
φ (deg°)	$180.88 \pm 56.78°$	$176.64 \pm 13.14°$

Figure 13. Sample bilateral stepping phase symmetry values plotted based on the data of, (**a**) left hemiparetic subject number 14; (**b**) healthy subject number 34.

4. Discussion

With the aim of producing a practical and reliable device to quantitatively analyze the human gait, a 10-channeled TCPSI and a real-time monitoring system were developed and applied in this study for hemiparetic gait pattern analysis in comparison with healthy adults. The results of this study showed that our sensor is capable of detecting and distinguishing the differences in plantar pressure (PPD), the step count, the stride time, the coefficient of variation, and the phase coordination index (PCI) between paretic and healthy limbs. These parameters were selected for analysis, because a stroke survivors' gait abnormality is characterized by a pronounced gait asymmetry [25], decreased gait speed and stance phase, shorter stride length, and prolonged swing phase of the paretic limb [39]. These gait abnormalities along with muscle imbalance and weakness leads to a high risk of falls [40,41].

In this study, the subjects with hemiparesis showed the highest PPD in the toe area where the sensor numbers 1 and 2 were located (right hemi: 13.43%, left hemi: 28.27%). The closer to 0% the PPD is, the less the differential pressure of the feet, which means that the feet are well balanced [17]. Sanghan et al. (2015) reported that the PPD of hemiparetic patients was 3 times higher than that of the healthy group [17]. Normally in healthy subjects, the heel contacts first, followed by the midfoot, the forefoot, and then toe pressure. However, since hemiplegic patients do not pressurize the paralyzed feet properly when walking, it may have showed higher pressure in the toe area and in general compared to the healthy adult group. This result was congruent with the results in Perry's research (1992) [39]. Based on this, our proposed sensor was confirmed that it can detect the dynamic pressure difference of the two feet and be analyzed. In gait analysis, plantar pressure distribution is an important parameter for evaluating the balancing capability in terms of weight-shifting, balance strategies, and a risk of falls.

The plantar pressure distribution comparison analysis showed that the plantar pressure moves towards the lateral side of the affected foot in hemiparetic patients. This phenomenon occurs as a form of compensation to avoid falling. The patients shift their bodyweight towards the unaffected side for more secured stability [42]. The plantar dynamics of the involved leg exhibited a transfer of initial contact from the hind to the forefoot, increased lateral plantar support, limited roll-over, and reduced or absent push-off at a terminal stance. There was a tendency for heel support to disappear on the paretic side with reduced functional abilities.

Field et al. (2013) reported that the patients with a stroke reduce daily stepping activities by 27% compared to a healthy adult [43]. This was also found in our study results. The step count of hemiparetic patients was higher than that of healthy adults.

The average stride time difference between the right and left foot of left hemiparesis was 0.14 ± 0.04 s (Table 4), those of the right hemiparesis was 0.11 ± 0.04 s, and of the bilateral hemiparesis was 0.04 ± 0.04 s, respectively (Tables 5 and 6). The temporal aspects of hemiplegic gait are characterized by increased stride times [44]. Hemiplegic patients usually have reduced joint excursion and insufficient forward propulsion, which may lead to an asymmetrical and unstable walking pattern [5]. Since the single support time of the affected limb is significantly shorter than that of the unaffected limb, the

unaffected limb's step length is shorter than the affected limb's step length [4,6]. Likely, the stride time of the affected foot was longer than the unaffected side in our study.

As presented in Table 6, the time difference in the average stride time in healthy adults was 0.02 ± 0.01 s and the average stride time difference was three times lower than the hemiparetic patients. Mackenzie et al. (2006) and Sartini et al. (2010) reported that the stride time was related to falling and an abnormal gait pattern causes a risk of falls [45,46]. The elderly who have experienced falls are associated with a decrease in walking speed, stepping and an increase in step symmetry [47]. Our proposed sensor system can also detect and decrease the risk of falls by real-time gait monitoring.

The CV is used to compare the magnitude of the change regardless of the data dimension. In this study, because the step length and the stride time may be different depending on the subject and the degree of the disease, the CV was calculated and the change amount was examined. The CV of stride time with heathy adults was 5 times higher than hemiparetic patients as shown in Table 8 and Figure 9. This means that hemiparetic patients cannot walk with a constant gait speed and thus walk irregularly. It was confirmed that our sensor can analyze various gait parameters.

This study performed the PCI comparison between hemiparetic patients and healthy adults because hemiparetic gait is characterized by mild to severe asymmetric patterns. Hemiparetic patients have increased stance phase and double support duration. In addition, the stride length and step length decrease, and the paralyzed side has longer step [48]. The PCI is an indicator for evaluating the balance of a pair of feet. A value closer to 0% refers to a higher balance between the two feet [23–25,48]. The PCI was originally developed to evaluate the asymmetry during walking and many studies evaluated the gait asymmetry of patients with Parkinson's disease and a stroke [23–25,48]. In the previous studies evaluating gait asymmetry in patients with stroke patients [48], the PCI value of stroke patients was $19.5\% \pm 2.3\%$ and that of healthy subjects was $6.2\% \pm 1.0\%$. The PCI value was approximately 3 times higher in the patient group than in the healthy group.

In this study, the PCI value of stroke patients was $19.5\% \pm 13.9\%$ and that of healthy subjects was $5.6\% \pm 5.1\%$, showing similar results to the previous studies as summarized in Table 7. Our results showed a higher standard deviation of the PCI value than that of previous studies. This may be because the age range of the subjects participating in this study was larger (63.3 ± 8.6 yrs.) The ages of the subjects (number 2, 6, 8, 10, and 14) were 65, 63, 63, 60, and 68, respectively, and were diagnosed with hemiparesis. The PCI values of the five subjects were 4.07%, 4.00%, 4.64%, 0.98%, 4.03%, respectively, which seemed to be similar to those of healthy adults. According to the previous studies, aging is associated with decreased stride time, velocity, step length during gait. Slowed walking speed leads to larger gait variability [37]. Other previous studies explored age-related changes by analyzing gait speed and PCI values, and found out that the average gait speed decreases and the PCI value increases every decade from the age of 70 years [49–51].

Study limitations: Generalizing the results of this study may be difficult because the number of subjects was small. In our previous study [35], a soft-material-based pressure insole and its performance feasibility test for performing a clinical experiment was developed and analyzed. The study results showed that our pressure insole sensor was confirmed to have a high correlation with F-scan ($R^2 \geq 0.90$, $p < 0.05$). Therefore, the sensor performance analysis was not repeated in this study.

5. Conclusions

The insole-type textile capacitive pressure sensor and a real-time gait monitoring system developed in this study was tested and confirmed for its clinical feasibility in applying the sensor system to hemiparetic patients. Hemiparetic gait patterns were well distinguished from healthy gait. Our proposed sensor may be used as a wearable gait function evaluation tool, external feedback gait training device, and a simple gait pattern analyzer for both hemiparetic patients and healthy individuals.

Author Contributions: Conceptualization, C.W.; Writing-original draft, C.W.; Writing-English version, Y.K.; Review & editing, Y.K.; Data curation, H.S.; Investigation, H.S.; Project administration, S.D.M.; Supervision, S.D.M.

References

1. Moseleyt, A.M.; Stark, A.; Cameron, I.D.; Pollock, A. Treadmill training and body weight support for walking after stroke. *Cochrane Database Syst. Rev.* **2005**, *4*. [CrossRef]

2. Seo, J.S.; Yang, H.S.; Jung, S.; Kang, C.S.; Jang, S.; Kim, D.H. Effect of reducing assistance during robot-assisted gait training on step length asymmetry in patients with hemiplegic stroke: A randomized controlled pilot trial. *Medicine* **2018**, *97*, 11792. [CrossRef] [PubMed]

3. Alexander, L.D.; Black, S.E.; Patterson, K.K.; Gao, F.; Danells, C.J.; Mcllroy, W.E. Association between gait asymmetry and brain lesion location in stroke patients. *Stroke* **2009**, *40*, 537–544. [CrossRef] [PubMed]

4. Patterson, K.K.; Gage, W.H.; Brooks, D.; Black, S.E.; Mcllroy, W.E. Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization. *Gait Posture* **2010**, *31*, 241–246. [CrossRef] [PubMed]

5. Patterson, K.K.; Parafianowicz, I.; Danells, C.J.; Closson, V.; Verrier, M.C.; Staines, W.R.; Black, S.E.; Mcllroy, W.E. Gait asymmetry in community-ambulating stroke survivors. *Arch. Phys. Med. Rehabil.* **2008**, *89*, 304–310. [CrossRef] [PubMed]

6. Titianova, E.B.; Tarkka, I.M. Asymmetry in walking performance and postural sway in patients with chronic unilateral cerebral infarction. *J. Rehabil. Res. Dev.* **1995**, *32*, 236–244. [PubMed]

7. Cheng, P.T.; Liaw, M.Y.; Wong, M.K.; Tang, F.T.; Lee, M.Y.; Lin, P.S. The sit-to-stand movement in stroke patients and its correlation with falling. *Arch. Phys. Med. Rehabil.* **1998**, *79*, 1043–1046. [CrossRef]

8. Balasubramanian, C.K.; Bowden, M.G.; Neptune, R.R.; Kautz, S.A. Relationship between step length asymmetry and walking performance in subjects with chronic hemiparesis. *Arch. Phys. Med. Rehabil.* **2007**, *88*, 43–49. [CrossRef]

9. Aqueveque, P.; Osorio, R.; Pastene, F.; Saavedra, F.; Pino, E. Capacitive Sensors Array for Plantar Pressure Measurement Insole fabricated with Flexible PCB. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 4393–4396.

10. Ngueleu, A.M.; Blanchette, A.K.; Bouyer, L.; Maltais, D.; McFadyen, B.; Moffet, H.; Batcho, C.S. Design and Accuracy of an Instrumented Insole Using Pressure Sensors for Step Count. Design and Accuracy of an Instrumented Insole Using Pressure Sensors for Step Count. *Sensors* **2019**, *19*, 984. [CrossRef]

11. Sungkarat, S.; Fisher, B.E.; Kovindha, A. Efficacy of an insole shoe wedge and augmented pressure sensor for gait training in individuals with stroke: A randomized controlled trial. *Clin. Rehabil.* **2011**, *25*, 360–369. [CrossRef]

12. Alexandros, P.; Nikolaos, G.B. A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2010**, *40*, 1–12. [CrossRef]

13. Shu, L.; Hua, T.; Wang, Y.; Li, Q.; Feng, D.D.; Tao, X. In-Shoe Plantar Pressure Measurement and Analysis System Based on Fabric Pressure Sensing Array. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 767–775. [CrossRef] [PubMed]

14. Hegde, N.; Bries, M.; Sazonov, E. A Comparative Review of Footwear-Based Wearable Systems. *Electronics* **2016**, *5*, 48. [CrossRef]

15. Lemarie, E.D.; Biswas, A.; Kofman, J. Plantar Pressure Parameter for Dynamic Gait Stability Analysis. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006.

16. Nam, S.H.; Son, S.M.; Kim, Y. Changes of gait parameters following constrained-weight shift training in patients with stroke. *J. Phys. Ther. Sci.* **2017**, *29*, 673–676. [CrossRef] [PubMed]

17. Sanghan, S.; Leelasamran, W.; Chatpun, S. Imbalanced Gait Characteristics Based on Plantar Pressure Assessment in Patients with Hemiplegia. *Walailak J. Sci. Technol.* **2015**, *12*, 595. [CrossRef]

18. Storm, F.A.; Heller, B.W.; Mazza, C. Step Detection and Activity Recognition Accuracy of Seven Physical Activity Monitors. *PLoS ONE* **2015**, *10*, e0118723. [CrossRef] [PubMed]

19. Tudor-Locke, C.; Sisson, S.B.; Collova, T.; Lee, S.M.; Swan, P.D. Pedometer-Determined Step Count Guidelines for Classifying Walking Intensity in a Young Ostensibly Healthy Population. *Can. J. Appl. Physiol.* **2005**, *30*, 666–676. [CrossRef]

20. Montero-Odasso, M.; Verghese, J.; Beauchet, O.; Hausdorff, J.M. Gait and Cognition: A Complementary Approach to Understanding Brain Function and the Risk of Falling. *J. Am. Geriatr. Soc.* **2012**, *60*, 2127–2136. [CrossRef]

21. Rampp, A.; Barth, J.; Schulein, S.; Gabmann, K.G.; Klucken, J.; Eskofier, B.M. Inertial Sensor-Based Stride Parameter Calculation From Gait Sequences in Geriatric Patients. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1089–1097. [CrossRef]

22. Studenski, S.; Perera, S.; Patel, K.; Rosano, C.; Faulkner, K.; Inzitari, M.; Brach, J.; Chandler, J.; Cawthon, P.; Connor, E.B.; et al. Gait Speed and Survival in Older Adults. *JAMA* **2011**, *305*, 50–58. [CrossRef]

23. Nanhoe-Mahabier, W.; Snijders, A.H.; Delval, A.; Weerdesteyn, V.; Duysens, J.; Overeem, S.; Bloem, B.R. Walking patterns in Parkinson's disease with and without freezing of gait. *Neuroscience* **2011**, *182*, 217–224. [CrossRef] [PubMed]

24. Plotnik, M.; Bartsch, R.P.; Zeev, A.; Giladi, N.; Hausdorff, J.M. Effects of walking speed on asymmetry and bilateral coordination of gait. *Gait Posture* **2013**, *38*, 864–869. [CrossRef] [PubMed]

25. Olney, S.J.; Richards, C. Hemiparetic gait following stroke. Part I: Characteristics. *Gait Posture* **1996**, *4*, 136–148. [CrossRef]

26. Razak, A.H.; Zayegh, A.; Begg, R.K.; Wahab, Y. Foot Plantar Pressure Measuric: A Review. *Sensors* **2012**, *12*, 9884–9912. [CrossRef] [PubMed]

27. Wu, Y.; Xu, W.; Liu, J.J.; Huang, M.C.; Luan, S.; Lee, Y.J. An Energy-Efficient Adaptive Sensing Framework for Gait Monitoring Using Smart Insole. *IEEE Sens. J.* **2015**, *15*, 2335–2343. [CrossRef]

28. Park, S.Y.; Das, P.S.; Park, J.Y. Development of wearable and flexible insole type capacitive pressure sensor for continuous gait signal analysis. *Org. Electron.* **2018**, *53*, 213–220. [CrossRef]

29. Scilingo, E.P.; Gemignani, A.; Paradiso, R.; Taccini, N.; Ghelarducci, B.; De Rossi, D. Performance evaluation of sensing fabrics for monitoring physiological and biomechanical variables. *IEEE Trans. Inf. Technol. Biomed.* **2005**, *9*, 345–352. [CrossRef]

30. Lou, C.; Wang, S.; Liang, T.; Pang, C.; Huang, L.; Run, M.; Liu, X. A Graphene-Based Flexible Pressure Sensor with Applications to Plantar Pressure Measurement and Gait Analysis. *Materials* **2018**, *10*, 1068. [CrossRef]

31. Yang, C.M.; Chou, C.M.; Hu, J.S.; Hung, S.H.; Yang, C.H.; Wu, C.C.; Hsu, M.Y.; Yang, T.L. A wireless gait analysis system by digital textile sensors. In Proceedings of the 31st Annual International Conference of the IEEE EMBS, Minneapolis, MN, USA, 2–6 September 2009.

32. Guo, X.; Huang, Y.; Cai, X.; Liu, C.; Liu, P. Capacitive wearable tactile sensor based on smart textile substrate with carbon black /silicone rubber composite dielectric. *Meas. Sci. Technol* **2016**, *27*, 1–8. [CrossRef]

33. Cheng, J.; Sundholm, M.; Zhou, B.; Hirsch, M.; Lukowicz, P. Smart-surface: Large scale textile pressure sensors arrays for activity recognition. *Pervasive Mob. Comput.* **2016**, *30*, 97–112. [CrossRef]

34. Holleczek, T.; Ruegg, A.; Harms, H.; Troster, G. Textile pressure sensors for sports applications. In Proceedings of the 2010 IEEE Conference Sensors, Kona, HI, USA, 1–4 November 2010.

35. Wang, C.W.; Kim, Y.; Min, S.D. Soft-Material-Based Smart Insoles for a Gait Monitoring System. *Materials* **2018**, *11*, 2435. [CrossRef] [PubMed]

36. Liu, T.; Inoue, Y.; Shibata, K. A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability. *Sensors* **2010**, *10*, 10240–10255. [CrossRef] [PubMed]

37. Gimmon, Y.; Rashad, H.; Kurz, I.; Plotnik, M.; Riemer, R.; Debi, R.; Shapiro, A.; Melzer, I. Gait Coordination Deteriorates in Independent Old-Old Adults. *J. Aging Phys. Act.* **2018**, *26*, 382–389. [CrossRef] [PubMed]

38. Brach, J.S.; Berthold, R.; Craik, R.; VanSwearingen, J.M.; Newman, A.B. Gait Variability in Community-Dwelling Older Adults. *J. Am. Geriatr. Soc.* **2005**, *49*, 1646–1650. [CrossRef]

39. Yoon, H.W.; Lee, S.Y.; Lee, H.M. The comparison of Plantar Foot Pressure in Normal Side of Normal People, Affected Side and Less Affected Side of Hemiplegic Patients During Stance Phase. *J. Korean Soc. Phys. Med.* **2009**, *4*, 87–92.

40. Perry, J.; Burnfield, J.M. Gait Analysis: Normal and Pathological Function. *J. Sports Sci. Med.* **2010**, *9*, 353. [CrossRef]

41. Dobkin, B.H. Clinical practice. Rehabilitation after stroke. *N. Engl. J. Med.* **2005**, *352*, 1677–1684. [CrossRef] [PubMed]
42. Batchelor, F.A.; Hill, K.D.; Mackintosh, S.F.; Said, C.M.; Whitehead, C.H. Effects of a multifactorial falls prevention program for people with stroke returning home after rehabilitation: A randomized controlled trial. *Arch. Phys. Med. Rehabil.* **2012**, *93*, 1648–1655. [CrossRef] [PubMed]
43. Field, M.J.; Gebruers, N.; Shanmuga Sundaram, T.; Nicholson, S.; Mead, G. Physical activity after stroke: A systematic review and meta-analysis. *ISRN Stroke* **2013**, 464176. [CrossRef]
44. Brandstater, M.E.; de Bruin, H.; Gowland, C.; Clark, B.M. Hemiplegic gait: Analysis of temporal variables. *Arch. Phys. Med. Rehabil.* **1983**, *64*, 583–587. [PubMed]
45. Mackenzie, L.; Byles, J.; D'Este, C. Validation of self-reported fall events in intervention studies. *Clin. Rehabil.* **2016**, *20*, 331–339. [CrossRef] [PubMed]
46. Sartini, M.; Cristina, M.L.; Spagnolo, A.M.; Cremonesi, P.; Costaguta, C.; Monacelli, F.; Garau, J.; Odetti, P. The epidemiology of domestic injurious falls in a community dwelling elderly population: An outgrowing economic burden. *Eur. J. Public Health* **2010**, *20*, 604–606. [CrossRef] [PubMed]
47. Auvinet, B.; Berrut, G.; Touzard, C.; Moutel, L.; Collet, N.; Chaleil, D.; Barrey, E. Reference data for normal subjects obtained with an accelerometric device. *Gait Posture* **2002**, *16*, 124–134. [CrossRef]
48. Meijer, R.; Plotnik, M.; Zwaaftink, E.G.; van Lummel, R.C.; Ainsworth, E.; Martina, J.D.; Hausdorff, J.M. Markedly impaired bilateral coordination of gait in post-stroke patients: Is this deficit distinct from asymmetry? A cohort study. *J. Neuro Eng. Rehabil.* **2011**, *8*, 1–8. [CrossRef] [PubMed]
49. Vaupel, J.W. Biodemography of Human ageing. *Nature* **2010**, *464*, 536–542. [CrossRef] [PubMed]
50. Plotnik, M.; Giladi, N.; Hausdorff, J.M. A new measure for quantifying the bilateral coordination of human gait: Effects of aging and Parkinson's disease. *Exp. Brain Res.* **2007**, *181*, 561–570. [CrossRef]
51. Judge, J.O.; Davis, R.B.; Ounpuu, S. Step length reductions in advanced age: The role of ankle and hip kinetics. *J. Gerontol. A Biol. Sci. Med. Sci.* **1996**, *51*, 303–312. [CrossRef]

Lower Limb Locomotion Activity Recognition of Healthy Individuals Using Semi-Markov Model and Single Wearable Inertial Sensor

Haoyu Li [1,*], Stéphane Derrode [1] and Wojciech Pieczynski [2]

[1] LIRIS, CNRS UMR 5205, École Centrale de Lyon, 69130 Ecully, France; stephane.derrode@ec-lyon.fr
[2] SAMOVAR, CNRS UMR 5157, Telecom SudParis, Institut Polytechnique de Paris, 91011 Evry CEDEX, France; wojciech.pieczynski@telecom-sudparis.eu
* Correspondence: haoyuli1990@gmail.com

Abstract: Lower limb locomotion activity is of great interest in the field of human activity recognition. In this work, a triplet semi-Markov model-based method is proposed to recognize the locomotion activities of healthy individuals when lower limbs move periodically. In the proposed algorithm, the gait phases (or leg phases) are introduced into the hidden states, and Gaussian mixture density is introduced to represent the complex conditioned observation density. The introduced sojourn state forms the semi-Markov structure, which naturally replicates the real transition of activity and gait during motion. Then, batch mode and on-line Expectation-Maximization (EM) algorithms are proposed, respectively, for model training and adaptive on-line recognition. The algorithm is tested on two datasets collected from wearable inertial sensors. The batch mode recognition accuracy reaches up to 95.16%, whereas the adaptive on-line recognition gradually obtains high accuracy after the time required for model updating. Experimental results show an improvement in performance compared to the other competitive algorithms.

Keywords: gait analysis; lower limb locomotion activity; triplet Markov model; semi-Markov model; on-line EM algorithm

1. Introduction

Locomotion activity has recently raised great research interest because of its significant potentials in many fields, e.g., rehabilitation for injured people [1], surveillance systems or health care for the elderly [2], daily activity management. Among these researches [3], many different types of sensors are used, such as camera, wireless beacon, electromyogram (EMG) sensors, electrocardiography (ECG) sensors, and inertial measurement units (IMUs). In a smart home, camera system or wireless beacon can help to understand the activity pattern of the host, and then provide suggestions for a healthy life or make a decision when an emergency is coming [4]. On the other hand, for the wearable sensors, EMGs can measure the electrical signal of muscles, while ECGs placed on specific body parts can monitor the heart rate. These kinds of signals can be used for evaluating the activity intensity. However, camera systems need to be pre-installed and calibrated, and they are also sensitive to the light. EMGs and ECGs have cables with the host, and they are sensitive to the moisture. By contrast, IMU sensors are small enough to be placed on the body and can be taken anywhere, providing information like 3D acceleration, angular rate, and magnetic field readings. In this work, given the advantages of using IMUs, we propose to use these sensors to collect the acceleration and angular rate of motion for the purpose of activity recognition.

Numerous single sensor-based and multiple sensors-based applications were developed under different scenarios. It seems that using multiple sensors is quite interesting and can help to recognize

more complex activities. For example, Hsu et al. [5] utilized two IMU sensors placed on wrist and ankle to detect 10 daily activities and 11 sport activities. Xie et al. [6] used a hybrid system of inertial sensor and barometer to detect locomotion and static activities. In this paper, we are studying a generic model that can be applied to the recognition of lower limb locomotion activity, this kind of model can work for both single sensor-based and multiple sensors-based applications; the difference is that multiple sensors generate a higher observation dimension than a single sensor. For simplification, the proposed model will be validated through only one IMU sensor placed on the lower limb.

The work proposed here is, to some extent, the continuation of our previous work [7], where a non-parametric triplet Markov chain (TMC-HIST) was designed to detect four lower limb locomotion activities: walking, running, stair ascent and stair descent. TMC [8,9] is an extension of the hidden Markov chain model (HMC) that includes: the observation Y and hidden state X processes and a third auxiliary hidden state U process. While it keeps a similar parameter estimation and restoration algorithm as HMC, in the TMC-HIST, the hidden state process represented the considered activities, the auxiliary one modelized the gait cycle, and histograms were used to represent the non-Gaussian observation density conditioned on each hidden state. We also developed an adaptive on-line algorithm that is based on TMC-HIST to recognize the targeted activities. Results showed that the combination of lower limb activity and gait cycle can significantly improve the recognition performance, and the adaptive parameter updating can gradually fit the motion pattern of people. However, the non-parametric histogram represented the marginal density of observation along one sensor axis, it does not involve the correlation among the three axes of sensor. As a consequence, this weakness may cause a failure when recognizing the activity. In addition, the precision of histogram is highly dependent on the volume of data and the width of bins, which require large storage memory and will slow down the processing speed of on-line recognition.

In this work, in order to overcome the weaknesses of TMC-HIST, we focus on developing a new parametric TMC model that can recognize lower limb locomotion activities using one single IMU sensor. Besides, the proposed algorithm should be adaptive and on-line applicable as well, i.e., it can adjust its parameters at run-time to suit for the user. By introducing a sojourn hidden state process to form semi-Markov structure, it allows the hidden states X and U keep the same for a while, which is consistent with the activity and gait transition during the motion. Semi-Markov structure is embedded into the TMC to better mimic the real state transition properties. Multi-dimensional Gaussian mixture model (GMM) is introduced to represent the non-Gaussian conditioned observation densities, in the meantime, it involves the observation correlation among the sensor axes. With the introduction of semi-Markov structure and Gaussian mixture density, the specific TMC model will be referred to as SemiTMC-GMM in the remaining of this paper. Because of the parametric densities, an on-line parameter learning algorithm based on EM is applied. Therefore, our claimed contributions in this paper are:

- Semi-Markov structure is embedded into the TMC model to make the hidden state transition closer to the realistic motion.
- GMM is adopted to overcome the weakness of non-parametric density, while still allowing to model non-Gaussian data.
- An EM-based on-line learning algorithm is adopted to SemiTMC-GMM for making the algorithm work on-line.

The remaining of the paper is organized as follows. Section 2 depicts the state-of-the-art works in the field of activity recognition using wearable sensors. Section 3 gives the definition of conventional TMC model and gradually extends the model to SemiTMC-GMM. Then, how to apply the proposed model to recognize lower limb locomotion activities is presented in detail at the end of this Section. Section 4 depicts both batch mode and on-line mode parameter learning for the proposed model. In Section 5, the proposed recognition algorithm is tested on two datasets: one is the public dataset [10], and the other is our own dataset. In addition, the performance of the proposed algorithm is discussed

compared to the competitive works. Finally, conclusions and future work are presented in the last Section.

2. Related Works

Numerous works have investigated human activity recognition (HAR) in the last decade. The methodologies used recently can generally be classified into two dominant categories: (i) traditional classifiers; (ii) deep learning methods.

For the first category, numerous classiffiers have been investigated. Parri et al. [11] proposed a fuzzy-logical classifier to identify lower limb locomotion mode, with the assistance of gait phases. The authors developed a lower limb wearable robot system that can help impaired people to perform locomotion activity. Chen et al. [12] proposed a robust activity recognition algorithm based on principal component analysis (PCA) and on-line support vector machine (OSVM), the algorithm obtained a robust recognition accuracy over a smartphone dataset collected in six different orientations. In the work [13], the authors compared the performances among the classifiers of SVM, Naive Bayes, k-Nearest Neighbour (kNN) and kStar. Results showed that kNN and kStar obtained the highest accuracy while Naive Bayes obtained the lowest. Zhao et al. [14] proposed a 2-layer model to detect six gait phases of walking, the algorithm used Neural Network (NN) to provide a pre-decision of gait phases to Hidden Markov Model (HMM), the final decision of gait phase from HMM obtained an accuracy of 98.11%. The limitation of this study is that only the activity of walking was considered, and the authors only tested their algorithm on straight forward walking, not free walking. In [15], hidden semi-Markov model (HSMM) and semi-Markov conditional random field (SMCRF) were applied to recognize human activity in smart home. The results showed that HSMM consistently outperformed HMM, while SMCRF obtained a similar result to CRF. However, because daily activities at home do not have stationary property, it is not practical to use a stationary transition matrix to represent the activity switches. Moreover, the authors only used Gaussian density to represent the conditioned observation density, which is quite limited to a complex scenario.

In the second category, deep learning-based methodologies are very prevalent. Generally, this kind of method is more inclined to image processing, so it needs to convert sensor data to image description to support extraction of discriminative features [16]. As reported in [17], convolutional neural network (CNN) is an important category of discriminative deep learning model for HAR. The work [18] proposed convolutional recurrent neural network to recognise daily activity; their algorithm gained an improvement of 6% compared to the state-of-the-art works. Recently, as reported in [19], transfer learning and semantic approach have raised great research interest. Bao [20] and Rokni [21] used transfer learning to automatically construct model for newly added wearable sensors; they obtained an accuracy enhancement between 9.3–10%. However, the recognition accuracy highly depends on the performance of labeling from source devices, thus it still requires a reliable method for recognition on a single sensor.

Some other methods can also be applied to dedicated applications and obtain good results. Schneider et al. [22] proposed an automatic extraction and selection method of highly relevant features, the method was tested on eight datasets and obtained a general accuracy over 90%. Rezaie et al. [23] proposed a feedback controller framework to adapt the sampling rate for better efficiency and higher accuracy. Dao et al. [24] introduced a man-in-loop decision architecture and data sharing among users and gradually obtained a high accuracy.

In fact, people perform lower limb locomotion activities every day, such as moving from one place to another place and doing sports like running and cycling. ... There are a lot of methods that have been proposed for HAR, while to our best knowledge, very few methods can be found that are specially designed for lower limb locomotion activities, including but not limited to activities like walking and jogging [25].

3. Model

In this section, the conventional TMC model is firstly introduced, then it is gradually equipped with more sophisticated structures, i.e., applying Gaussian mixture to TMC to obtain the TMC-GMM model and then applying semi-Markov structure to TMC-GMM to obtain the SemiTMC-GMM model. Afterward, a detailed description of on-line EM algorithm suited for SemiTMC-GMM is given. As a matter of fact, these additional processes can be naturally added because of the high generality of the TMC model through the flexibility of the auxiliary processes.

3.1. Triplet Markov Chain

Consider two discrete stochastic processes $X = (X_1, \cdots, X_N)$ and $U = (U_1, \cdots, U_N)$ as hidden states, where $X_n \in \Lambda = \{1, \cdots, r\}$ and $U_n \in \Gamma = \{1, \cdots, \tau\}$, $n \in \{1, \ldots, N\}$. Let $Y = (Y_1, \cdots, Y_N)$ be a real-valued process representing the observation of the model, each $Y_n \in \mathbb{R}^w$, where w is the observation dimension. In this paper, the hidden state X refers to the activity to be recognized, U refers to the introduced gait or leg cycle, while the observation Y is the features extracted from sensor readings. The details of how to apply the model to recognize lower limb locomotion activity are described in Section 3.4. Then, the triplet $T = (V, Y)$, with $V = (X, U)$ is a TMC if T is Markovian. It should be noted here that, in classic TMC, none of processes $X, U, Y, (X, U), (X, Y), (U, Y)$ are necessarily Markovian.

Let the realizations of X_n, U_n and Y_n be denoted by their lower cases x_n, u_n, and y_n, respectively, so $v_n = (x_n, u_n)$, $t_n = (v_n, y_n)$. In addition, for simplification, we will denote the probabilities $p(X_n = x_n, U_n = u_n | Y_1 = y_1, \cdots, Y_N = y_N)$ by $p(x_n, u_n | y_1^N)$, $p(X_n = x_n, U_n = u_n | Y_1 = y_1, \cdots, Y_n = y_n)$ by $p(x_n, u_n | y_1^n)$, for example. In TMC transitions form $p(t_{n+1} | t_n)$ can be expressed in different forms, let us consider the following one:

$$p(t_{n+1} | t_n) = p(v_{n+1} | v_n, y_n) \, p(y_{n+1} | v_{n+1}, v_n, y_n). \tag{1}$$

In the application of this paper we will assume that $p(v_{n+1} | v_n, y_n) = p(v_{n+1} | v_n)$ and $p(y_{n+1} | v_{n+1}, v_n, y_n) = p(y_{n+1} | v_{n+1})$. So the transition is simplified to

$$p(t_{n+1} | t_n) = p(v_{n+1} | v_n) \, p(y_{n+1} | v_{n+1}), \tag{2}$$

which provides process T with the structure of a classical HMC. For simplification, this simplified TMC is referred to as TMC in the remaining. The first term $p(v_{n+1} | v_n)$ in Equation (2) is the state transition probability, the dimension of the matrix is $(r \times \tau) \times (r \times \tau)$. The second term is the probability of observing y_n conditionally to each state. Most of the time, this kind of density is modeled by Gaussian distributions:

$$p(y_n | v_n = i) \sim \mathcal{N}(\mu_i, \Sigma_i), i \in \Lambda \times \Gamma, \tag{3}$$

where μ_i and Σ_i are the mean and variance. The dependency graph of this particular TMC is shown in Figure 1a. Regardless of the probabilistic links inside the nodes related to V, the dependency of Y and V is just in the form of HMC.

For obtaining the probability of individual x_n and u_n conditioned on y_1^n, y_1^N, we only need to compute the marginal probability of $p(x_n, u_n | y_1^n)$ and $p(x_n, u_n | y_1^N)$. Indeed, we have

$$
\begin{aligned}
p(x_n | y_1^n) &= \sum_{u_n} p(x_n, u_n | y_1^n), \\
p(x_n | y_1^N) &= \sum_{u_n} p(x_n, u_n | y_1^N).
\end{aligned}
\tag{4}
$$

Likewise, $p(u_n | y_1^n)$ and $p(u_n | y_1^N)$ can be obtained in a similar way. $p(x_n, u_n | y_1^n)$ and $p(x_n, u_n | y_1^N)$ are the probabilities if known the observation y_1^n and y_1^N, commonly they are called filtering probability and smoothing probability, respectively. Similar meaning can be deduced for $p(x_n | y_1^n)$, $p(x_n | y_1^N)$,

$p(u_n|\boldsymbol{y}_1^n)$, $p(u_n|\boldsymbol{y}_1^N)$. Then, the estimated hidden state will be obtained via MPM (Maximum Posterior Mode) criterion using the smoothing probability:

$$
\begin{aligned}
\hat{x}_n &= \arg\max_{x_n \in \Lambda} p\left(x_n|\boldsymbol{y}_1^N\right), \\
\hat{u}_n &= \arg\max_{u_n \in \Gamma} p\left(u_n|\boldsymbol{y}_1^N\right).
\end{aligned}
\tag{5}
$$

3.2. TMC Embedding a Gaussian Mixture Model

When extending TMC to TMC-GMM, it needs to introduce Gaussian mixture density into the conditioned observation probability. In fact, embedding GMM in TMC can be regarded as introducing a new statistic process $\boldsymbol{H} = (H_1, \cdots, H_N)$ into TMC, where H_n takes its value h_n in a finite set $K = \{1, \cdots, \kappa\}$ and κ is the number of Gaussian components in the mixture. Please remind that $\boldsymbol{H}$ has no realistic meaning, it is just a latent variable in the model to introduce the mixtures. Let c_{ij} be the weight of jth Gaussian mixture component when $v_n = i$, with the constraint $\sum_{j=1}^{\kappa} c_{ij} = 1$. $\boldsymbol{\mu}_{ij}$ and Σ_{ij} are the mean value and covariance of the Gaussian mixture component. Denoting $\boldsymbol{Z} = (\boldsymbol{T}, \boldsymbol{H})$ and assuming $p(h_1^N|v_1^N) = \prod_{n=1}^{N} p(h_n|v_n)$, $\boldsymbol{Z}$ is Markovian with transitions:

$$
p(z_{n+1}|z_n) = p(v_{n+1}|v_n)p(h_{n+1}|v_{n+1})p(\boldsymbol{y}_{n+1}|v_{n+1}, h_{n+1}),
\tag{6}
$$

where

$$
p(\boldsymbol{y}_n|v_n = i, h_n = j) \sim \mathcal{N}\left(\boldsymbol{\mu}_{ij}, \Sigma_{ij}\right), \quad i \in \Lambda \times \Gamma, j \in K,
\tag{7a}
$$

$$
p(\boldsymbol{y}_n|v_n) = \sum_{j=1}^{\kappa} c_{ij} \cdot p(\boldsymbol{y}_n|v_n = i, h_n = j),
\tag{7b}
$$

with $p(h_n = j|v_n = i) = c_{ij}$. We can see that Equations (6) and (7a) are extensions of Equations (2) and (3), by introducing a new process $\boldsymbol{H}$. The dependency graph of TMC-GMM is shown in Figure 1b.

Please notice that the only difference between TMC and TMC-GMM is the Gaussian densities in TMC are replaced with Gaussian mixtures, all the other calculations remain the same. Then estimating the individual x_n and u_n in TMC-GMM follows the same as in TMC, by using Equations (4) and (5).

3.3. Semi TMC-GMM

In the Markov model considered in our previous work [7], the remaining time of the sojourn of the hidden state v_n is of geometric distribution. While considering $\boldsymbol{V}$ as semi-Markovian seems to better suited to our problematic, as in general, v_n has no geometric remaining sojourn time. For example, the gait phase has a minimum duration, while in geometric distribution the maximal probability is for null duration. In real applications of classic hidden semi-Markov model (HSMM) [26,27], there is a fixed maximum sojourn time for each possible value of v_n. When v_n switches to a new value, the maximal possible random sojourn time is shorter than a fixed value M. Once the sojourn time has elapsed at time n, the hidden state must change to a different value, i.e., $p(v_{n+1} = v_n) = 0$. This implies that the maximum sojourn time should be large enough to cover the largest possible sojourn time, which appears as a drawback in our application. In another semi-Markov approach described in [28] that we adopt here, the random sojourn time (just after having switched) is not the exact duration of the state, but the minimum sojourn time. This means that once the sojourn time elapsed, the next hidden state is possible to stay the same. This character allows make the maximum value M significantly smaller than the one in classic HSMM, which accelerates the entire method since the dimension of transition matrix is reduced.

To be more precise, consider a new stochastic process $D = (D_1, \cdots, D_N)$ that represents the minimum remaining sojourn time in a given hidden state v_n, and the possible realization of each D_n (denoted by d_n) takes its value in $L = \{0, 1, \cdots, \ell\}$. Thus for $V_n = v_n$ and $D_n = d_n$, we have $v_n = v_{n+1} = \cdots = v_{n+d_n}$. And v_{n+d_n} is obtained *w.r.t.* $p(v_{n+d_n+1}|v_{n+d_n})$, which is a transition similar to the ones in the TMC and TMC-GMM. Thus, v_{n+d_n+1} is possible to be the same as v_{n+d_n}. Once v_{n+1} is set, a new minimum sojourn time d_{n+1} is obtained in $L = \{0, 1, \cdots, \ell\}$. Please notice that for $D_n = d_n \neq 0$, there is $D_{n+1} = d_{n+1} = d_n - 1$, which is specified in Equation (10).

Finally, SemiTMC-GMM is extended from TMC-GMM Z via the couple (Z, D), which follows the transition probabilities:

$$p(z_{n+1}, d_{n+1}|z_n, d_n) = p(v_{n+1}|z_n, d_n)p(h_{n+1}|v_{n+1})p(d_{n+1}|v_{n+1}, d_n)p(y_{n+1}|v_{n+1}, h_{n+1}), \qquad (8)$$

$$p(v_{n+1}|z_n, d_n) = \begin{cases} \delta_{v_n}(v_{n+1}), & d_n > 0 \\ p^*(v_{n+1}|v_n), & d_n = 0 \end{cases}, \qquad (9)$$

$$p(d_{n+1}|v_{n+1}, d_n) = \begin{cases} \delta_{d_n-1}(d_{n+1}), & d_n > 0 \\ p(d_{n+1}|v_{n+1}), & d_n = 0 \end{cases}, \qquad (10)$$

where δ is the Kronecker function ($\delta_a(b) = 1$ for $a = b$ and $\delta_a(b) = 0$ for $a \neq b$).

The properties of the four terms on the right side of Equation (8) are clarified in the following:

1. In Equation (9), when $d_n = 0$, the transition $p^*(v_{n+1}|v_n)$ behaves the same as the state transition of TMC and TMC-GMM, which means that v_{n+1} can be different from or same as v_n, depending on the distribution of $p^*(v_{n+1}|v_n)$.
2. $p(d_{n+1}|v_{n+1}, d_n)$ is the probability of the minimal remaining sojourn time of v_{n+1}, conditioned on v_{n+1} and d_n.
3. $p(h_{n+1}|v_{n+1})$ and $p(y_{n+1}|v_{n+1}, h_{n+1})$ are same as the ones in TMC-GMM, shown in Equation (7a).

Now, the Equations (9) and (10) together describe how the hidden states, V_n and D_n, transfer in SemiTMC-GMM.

The dependency graphs of the three models, i.e., TMC, TMC-GMM, and SemiTMC-GMM, are shown in Figure 1. The couple $V = (X, U)$ is regarded as one hidden state for reducing the complexity of the graphs, as well as reminding that the total number of processes involved in the three models are 3, 4, and 5, respectively.
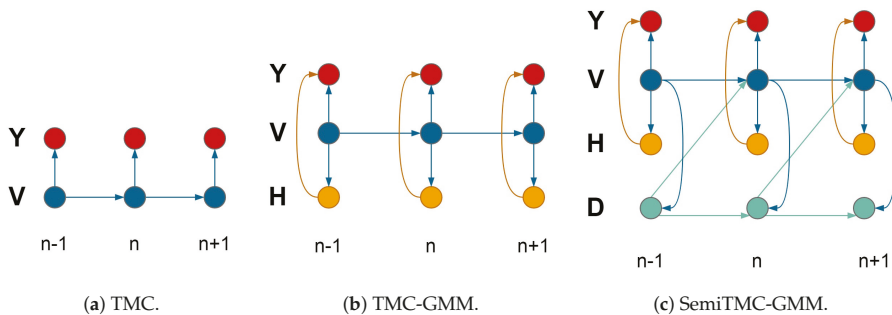

(a) TMC.   (b) TMC-GMM.   (c) SemiTMC-GMM.

Figure 1. Dependency graphs. TMC = triplet Markov chain; GMM = Gaussian mixture model.

Estimating the individual x_n and u_n is different from both TMC and TMC-GMM, for the sense of introducing the sojourn state D_n. The probabilities of x_n can be obtained by

$$p(x_n|\boldsymbol{y}_1^n) = \sum_{u_n}\sum_{d_n} p(x_n, u_n, d_n|\boldsymbol{y}_1^n),$$

$$p(x_n|\boldsymbol{y}_1^N) = \sum_{u_n}\sum_{d_n} p(x_n, u_n, d_n|\boldsymbol{y}_1^N). \tag{11}$$

The probabilities $p(x_n, u_n, d_n|\boldsymbol{y}_1^n)$ and $p(x_n, u_n, d_n|\boldsymbol{y}_1^N)$ are the filtering and smoothing probability of the hidden state in SemiTMC-GMM, respectively. Likewise, the probabilities $p(u_n|\boldsymbol{y}_1^n)$ and $p(u_n|\boldsymbol{y}_1^N)$ are obtained in a similar way. Finally, the estimated hidden state $\hat{x}_n$ and $\hat{u}_n$ can be obtained by Equation (5).

To summarize, the proposed SemiTMC-GMM is a model contains five stochastic processes X, U, D, H, Y, with Markov distribution of $\boldsymbol{Z}^* = (X, U, D, H, Y)$. The process X models the activities we are looking for, Y models the observation, U models the introduced gait or leg phase, D models the semi-Markovianity of $V = (X, U)$, and H models the presence of Gaussian mixtures. Thus, $\boldsymbol{Z}^* = (V, W, Y)$ can be regarded as a classic TMC with hidden state V, observed Y, and an additional latent $W = (D, H)$.

3.4. Application of SemiTMC-GMM

The question is now how to apply the proposed model to recognise lower limb locomotion activities. In our previous work [7], gait cycle was introduced into the estimation of four locomotion activities, and the results show that it can improve the accuracy. As introduced in [29], one gait cycle can be divided into four gait phases, i.e., stance, push-up, swing, and step down. In this work, we are pursuing a method that does not require the sensor to be placed on the feet only. On the contrary, it can be placed on different places of the lower limb, such as thigh, shank, or foot. The segmentation of gait cycle is based on the motion of foot, so similarly we can define 'leg cycle' based on the motion of leg. One leg cycle can be segmented into four leg phases, which are low position, lifting, high position, and dropping.

Let assume the hidden state X represents the activity, and U be the gait cycle or leg cycle. Thus, the dimension of $\Lambda(r)$ depends on the number of activities; while for Γ, τ is equal to 4. The transition of X and U follows a specific order because the feet move from attaching on the ground to swinging in the air alternately, or the legs switch between lifting to dropping. Therefore, we define a specific transition graph for X and U. As shown in Figure 2, the numbers 1–4 represent the hidden state U, the four gait and leg phases. We can see that U transfers from phase 1 to phase 4 and back to phase 1 again cyclically if the activity does not change. While the activity is switching, U transfers from phase 1 of the previous activity to phase 2 of the current activity.
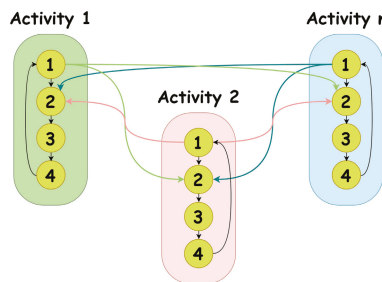

Figure 2. Hidden state transition graph. The activities represent X, and the numbers 1–4 represent U and stand for the four gait phases, or leg phases.

The hidden states H and D are not the final goals of the recognition, and they have no physical meaning neither. For simplification, the dimension of L (ℓ) is set to 9. This value was determined by our experience, a too small value will make the results of SemiTMC-GMM no difference from that of TMC-GMM, while a too large value will cost too much time for running the code. The performance of different GMM components number (κ) is evaluated on two datasets, as depicted in Section 5.

The observation Y is the feature extracted from the sensor readings. The utilized features are the sliding mean value and standard deviation, by calculating the mean value and standard deviation of acceleration and angular rate within a sliding window length. Since IMUs measure 3-dimensional acceleration and angular rate, then the dimension of the observation Y (w) equals to 12. The initialization of the hidden states is the same as the one in our previous work [7], so it will not be repeated here. Afterward, based on the initial hidden states v_1^N and features y_1^N, the initial conditioned GMM density $p(y_n|v_n, h_n)$ can be easily obtained if the mixture number κ is known. When the initialization is done, batch mode EM algorithm can be applied to train the model, which will be described in details in Section 4. Then, the trained model can be used for the batch mode testing, or, as the initial model of on-line EM algorithm.

4. Parameter Estimation

From the previous section, it is now clear how the hidden state transfers and how to compute the observation probability. In this section, we focus on how to obtain the filtering and smoothing probabilities, and to apply parameter updating based on the on-line EM algorithm.

Before starting the explanation, we need to introduce the parameter set first. As described in the previous Section, the parameter set can be defined as $\theta = \{\zeta_k, a_{lk}, c_{ij}, \mu_{ij}, \Sigma_{ij}\}$, in which ζ_k is the initial probability of hidden state, and a_{lk} is the l-th row and k-th column element in the transition matrix A. Because GMM density only depends on v_n, then $i \in \Lambda \times \Gamma$, $j \in K$. In SemiTMC-GMM, the entire hidden state is (V, D), and then $l, k \in \Lambda \times \Gamma \times L$, and l, k is equal to the couple of (i, d_n). Therefore, the initial probability becomes $\zeta_k = p((v_1, d_1) = k)$, and $a_{lk} = p((v_{n+1}, d_{n+1}) = k|(v_n, d_n) = l)$. For simplification, the indices i, j, l, k will keep the same meaning and will no longer be specified in the remaining.

4.1. Batch Mode EM Algorithm

The batch mode parameter restoration using EM algorithm is quite simple and has been utilized in many researches. A dominated way to do this is using the well-known Baum-Welch algorithm. This is an algorithm that makes the expectation step and maximization step recursively. Here we simply describe how to extend the expectation and maximization steps to SemiTMC-GMM model, within one iteration of the EM algorithm. A simple recall of the forward and backward calculations [7] are displayed below:

$$
\begin{aligned}
\alpha_1(k) &= p\left((v_1, d_1) = k, y_1\right), \\
\alpha_n(k) &= \sum_{(v_{n-1}, d_{n-1}) \in \Lambda \times \Gamma \times L} \alpha_{n-1}(k) \cdot p\left((v_n, d_n)|(v_{n-1}, d_{n-1})\right) \cdot p\left(y_n|(v_n, d_n)\right), \\
\beta_N(k) &= 1, \\
\beta_n(k) &= \sum_{(v_{n+1}, d_{n+1}) \in \Lambda \times \Gamma \times L} p\left((v_{n+1}, d_{n+1})|(v_n, d_n)\right) \cdot p\left(y_{n+1}|(v_{n+1}, d_{n+1})\right) \cdot \beta_{n+1}(k).
\end{aligned}
\tag{12}
$$

In the above equations, the $\alpha_n(k)$ and $\beta_n(k)$ are the forward and backward calculations, $p\left((v_n, d_n)|(v_{n-1}, d_{n-1})\right)$ is the state transition probability that has been described in Equations (8)–(10), $p\left(y_n|(v_n, d_n)\right) = p(h_n|v_n) p(y_n|v_n, h_n)$ is the conditioned observation density.

Then, the algorithm requires the following probabilities:

$$\gamma_n(k) = p((\boldsymbol{v}_n, d_n) = k | \boldsymbol{y}_1^N) = \frac{\alpha_n(k)\beta_n(k)}{\sum\limits_{k' \in \Lambda \times \Gamma \times L} \alpha_n(k')\beta_n(k')}, \tag{13}$$

$$\tilde{\gamma}_n(i) = \sum_{d_n} \gamma_n((i, d_n)) = \sum_{d_n} p(\boldsymbol{v}_n = i, d_n | \boldsymbol{y}_1^N), \tag{14}$$

$$\tilde{\gamma}_n(i, j) = \tilde{\gamma}(i) \cdot \frac{c_{ij} p(\boldsymbol{y}_n | \boldsymbol{v}_n = i, h_n = j)}{\sum\limits_{j' \in K} c_{ij'} p(\boldsymbol{y}_n | \boldsymbol{v}_n = i, h_n = j')}, \tag{15}$$

$$\xi_n(l, k) = \frac{\alpha_n(l) \cdot p(\boldsymbol{y}_{n+1}, h_{n+1}, (\boldsymbol{v}_{n+1}, d_{n+1}) = k | \boldsymbol{y}_n, h_n, (\boldsymbol{v}_n, d_n) = l) \cdot \beta_{n+1}(k)}{\sum\limits_{l', k' \in \Lambda \times \Gamma \times L} \left\{ \alpha_n(l') \cdot p(\boldsymbol{y}_{n+1}, h_{n+1}, (\boldsymbol{v}_{n+1}, d_{n+1}) = k' | \boldsymbol{y}_n, h_n, (\boldsymbol{v}_n, d_n) = l') \cdot \beta_{n+1}(k') \right\}}. \tag{16}$$

$\gamma_n(k)$ is the probability of $(\boldsymbol{v}_n, d_n)$ conditioned on all observed data $\boldsymbol{y}_n^N$. $\tilde{\gamma}_n(k)$ is the marginal probability of $\gamma_n(k)$ over d_n, this probability is the one that we are looking for to estimate the concerning hidden state $\boldsymbol{v}_n$. $\tilde{\gamma}_n(i, j)$ is the probability of each Gaussian component *w.r.t.* $\tilde{\gamma}_n(k)$; this probability helps to compute the parameters related to Gaussian mixture, i.e., c_{kj}, $\boldsymbol{\mu}_{kj}$, Σ_{kj}. $\xi_n(l, k)$ is the joint probability of $(\boldsymbol{v}_n, d_n) = l$ and $(\boldsymbol{v}_{n+1}, d_{n+1}) = k$ conditioned on $\boldsymbol{y}_1^N$. Here we give the formula of parameter update by using Equations (13)–(16):

$$\zeta_k = \gamma_1(k), \tag{17}$$

$$a_{lk} = \frac{\sum\limits_{n=1}^{N-1} \xi_n(l, k)}{\sum\limits_{n=1}^{N-1} \gamma_n(l)}, \tag{18}$$

$$c_{ij} = \frac{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i, j)}{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i)}, \tag{19}$$

$$\boldsymbol{\mu}_{ij} = \frac{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i, j) \boldsymbol{y}_n}{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i, j)}, \tag{20}$$

$$\Sigma_{ij} = \frac{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i, j) (\boldsymbol{y}_n - \boldsymbol{\mu}_{ij})^{\mathsf{T}} (\boldsymbol{y}_n - \boldsymbol{\mu}_{ij})}{\sum\limits_{n=1}^{N} \tilde{\gamma}_n(i, j)}. \tag{21}$$

In fact, Equations (13)–(16) are the expectation step in one iteration of EM algorithm, while Equations (17)–(21) are the maximization step. Then, the parameter can be learned by recursively performing the two steps until the iteration number exceeds a pre-defined value, maximally 100 iterations, for example.

4.2. Sufficient Data Statistics

Since Gaussian Markov models belong to the exponential family, the likelihood function of SemiTMC-GMM can be written in the form of [30]

$$p_\theta(\boldsymbol{z}_n, d_n) = f(\boldsymbol{z}_n, d_n) \exp\left(\langle s(\boldsymbol{z}_n, d_n), \psi(\theta) \rangle - J(\theta)\right), \tag{22}$$

where $s(z_n, d_n)$ is a vector of complete-data sufficient statistics belonging to convex set S, $\langle \cdot, \cdot \rangle$ denotes the scalar product, function $\psi(\cdot)$ maps θ to the natural parametrization, and $J(\cdot)$ is the log-partition function. For SemiTMC-GMM, the definition of statistics is

$$s^{(1)}_{n',lk} = \mathbb{1}\{(v_{n'}, d_{n'}) = l, (v_{n'+1}, d_{n'+1}) = k\}, \tag{23}$$

$$s^{(2)}_{n',k} = \mathbb{1}\{(v_{n'}, d_{n'}) = k\}, \tag{24}$$

$$s^{(3)}_{n',ij} = \mathbb{1}\{v_{n'} = i, h_{n'} = j\}, \tag{25}$$

$$s^{(4)}_{n',ij} = \mathbb{1}\{v_{n'} = i, h_{n'} = j\} \boldsymbol{y}_{n'}, \tag{26}$$

$$s^{(5)}_{n',ij} = \mathbb{1}\{v_{n'} = i, h_{n'} = j\} \boldsymbol{y}_{n'}^{\mathsf{T}} \boldsymbol{y}_{n'}, \tag{27}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, $n' = 1, \ldots, N$. Then, the statistics vector at time n' is of the form $s_{n'} = \left\{ s^{(1)}_{n',lk}, s^{(2)}_{n',k}, s^{(3)}_{n',ij}, s^{(4)}_{n',ij}, s^{(5)}_{n',ij} \right\}$. Consequently, the sufficient statistics S_n is the expectation of $s_{n'}$ conditioned on $\boldsymbol{y}_1^n$:

$$S_n = \frac{1}{n} \mathbf{E}_\theta \left(\sum_{n'=1}^{n} s_{n'} \right) \Big| \boldsymbol{y}_1^n. \tag{28}$$

Denote $S_n = \left\{ S^{(1)}_{n,lk}, S^{(2)}_{n,k}, S^{(3)}_{n,ij}, S^{(4)}_{n,ij}, S^{(5)}_{n,ij} \right\}$, in which the elements are the expectation of the ones with respect to $s_{n'}$. Now, comparing the Equation groups (13)–(21) and (23)–(28), we can reform the parameter update Equations (17)–(21) with sufficient statistics:

$$\tilde{S}^{(2)}_{n,i} = \sum_{d_n} S^{(2)}_{n,(i,d_n)}, \tag{29}$$

$$\zeta_k = S^{(2)}_{1,k}, \tag{30}$$

$$a_{n,lk} = S^{(1)}_{n,lk} / S^{(2)}_{n,k}, \tag{31}$$

$$c_{n,ij} = S^{(3)}_{n,ij} / \tilde{S}^{(2)}_{n,i}, \tag{32}$$

$$\boldsymbol{\mu}_{n,ij} = S^{(4)}_{n,ij} / S^{(3)}_{n,ij}, \tag{33}$$

$$\Sigma_{n,ij} = S^{(5)}_{n,ij} / S^{(3)}_{n,ij} - \boldsymbol{\mu}_{n,ij}^{\mathsf{T}} \boldsymbol{\mu}_{n,ij}, \tag{34}$$

in which $a_{n,lk}$, $c_{n,ij}$, $\boldsymbol{\mu}_{n,ij}$, $\Sigma_{n,ij}$ are the updated parameters at time n.

Remark 1. *Replacing n with N in Equation (28), which means all the observed data $\boldsymbol{y}_1^N$ are used, S_N is then called as complete sufficient statics. Therefore, using S_N to compute the parameters in Equations (30)–(34) will be exactly the batch mode parameter learning that is given in Equations (17)–(21).*

4.3. On-Line Estimation

In the previous section, we have discussed about how to use sufficient statistics to learn θ in batch mode. In order to apply the on-line estimation, a common way [30] is to update the sufficient statistics when a new observed data come in:

$$S_{n+1} = (1 - \rho_{n+1}) \cdot S_n + \rho_{n+1} \cdot \mathbf{E}_{\theta_n} \left(s_{n+1} | y_{n+1} \right), \tag{35}$$

where ρ_n is the stepsize sequence that satisfies $\sum_{n=1}^{\infty} \rho_n = \infty$, $\sum_{n=1}^{\infty} \rho_n^2 < \infty$. Normally it is set to $\rho_n = 1/n$. Then, the new parameter θ_{n+1} is available by Equations (29)–(34). The estimation of x_{n+1}, u_{n+1} can be obtained by Equations (5) and (11).

In this paper, we do not update θ at every sampling time. Instead, we set a window length W_l and accumulate the latest W_l observed data first. Then use Equations (13)–(16) to get the smoothed result, compute the sequenced statistics $s_n|_1^{W_l}$ for all the W_l data by Equations (23)–(27). Afterward, update the sequenced sufficient statistics $S_n|_1^{W_l}$ and $\theta_n|_1^{W_l}$ by Equations (35) and (29)–(34), respectively. It should be noticed that in on-line mode, the initial probability ζ_k is not necessary.

After describing the batch mode and on-line parameter learning, a diagram of the training and testing is displayed in Figure 3. In the training stage, the block of model parameter learning is the Baum-Welch algorithm. The trained model is used in both the two kinds of testing, the model is updated in on-line testing, but not in the batch mode testing. Besides, the estimated hidden state of batch mode is from the smoothed probability, whereas the one of on-line is from the filtered probability.
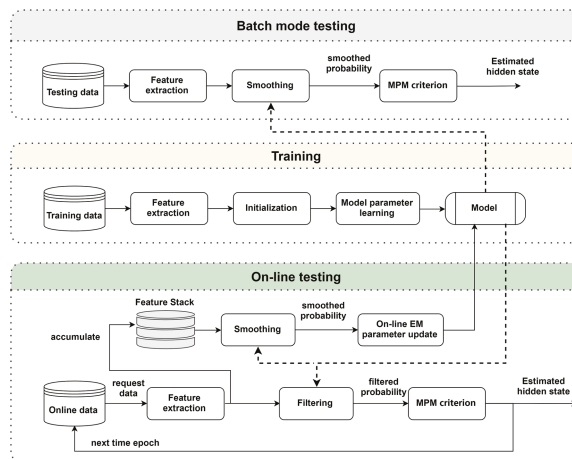

Figure 3. Diagram of the training stage, and the testing stage for both batch mode and on-line testing.

5. Experimental Results

Two datasets are used to validate the proposed algorithm. The first dataset is the Sports and Daily Activities (SDA) dataset [10], in which eight subjects were enrolled to perform 19 daily and sports activities while wearing five Xsens MTx (Details of Xsens MTx can be found in http://www. xsens.com/en/general/mtx)) IMUs on their torso, left arm, right arm, left thigh, and right thigh, all the sensors measure the acceleration and angular rate of the body parts where the sensors are placed. The five sensor placements are determined in this dataset because the involved activities are not only lower limb locomotion activities but also include static activities and upper limb activities, such as sitting, lying, rowing, and playing basketball... So they used the 5 sensors to collect the motion data from different parts of the body. However, in this paper, the proposed algorithm is designed for recognizing the lower limb locomotion activity with periodic gait or leg cycle performed by healthy people. Since we only care about the movement of the lower limbs and healthy people have a symmetric motion of the two legs, then it is possible to use only one sensor placed on either the left or right leg to recognize the considered activities. Therefore, we only use the sensor placed on the right thigh to validate our algorithm. The sensor sampling rate was set to 25 Hz, the acceleration sensing range was set to ± 18 g, the angular rate sensing range was set to $\pm 1200°/s$. Because the objective of the proposed algorithm is to detect lower limb locomotion activities that have gait cycle or leg cycle, while the 19 activities consists of both lower limb locomotion activities with and without the cycles, then only 11 suitable activities out of the total are selected in this work: walk in parking lot, walk on treadmill with incline, walk on treadmill on flat, stair descent, stair ascent, run on treadmill, jump, exercise on stepper, exercise of cycling in vertical position, exercise of cycling in horizontal

position, exercise on cross trainer. These 11 locomotion activities of SDA dataset are referred to as D1A1 to D1A11 in the remaining of this paper. In the dataset, the subjects performed each activity for about 5 min separately, and each activity was divided into 60 segments of 5 s. Therefore, there are $60 \times 8 = 480$ segments for each activity. In order to make the dataset available for our algorithm, we firstly combined the 60 segments of one activity from one subject to recover the 5 min activity, then combined the data of the same activity from all the subjects to form 40 min for each activity, the final data was obtained by combining the 11 activities. Thus, the duration of the data is $11 \times 40 = 440$ min.

There are only 7500 samplings for each experiment of SDA, and the data length is not long enough to use on-line EM recognition. Therefore, we utilize the second dataset for the validation of the proposed on-line EM algorithm. This second dataset, described in [7], is called Locomotion of Foot-mounted IMU (LMFIMU) dataset (the dataset and its details are available on the website: https://github.com/unilee/TMC_LowerLimbActs). Ten subjects were enrolled to perform a specific experiment that lasts nearly 30 min with a Shimmer3 (Shimmer3 GSR+, details at the manufacturer's site http://www.shimmersensing.com/images/uploads/docs/ConsensysPRO_Spec_Sheet_v1.1.0. pdf) IMU mounted on the right shoe. The sensor is placed on the shoe in LMFIMU but not the thigh as in SDA dataset is because both feet and thighs have a periodic pattern when people are performing the lower limb locomotion activities. In fact, the sensor can be placed at anywhere of the lower limbs in this study. So foot is chosen in the LMFIMU dataset to show that the proposed algorithm is not restricted to only one sensor position. Because of the enrolled 10 healthy subjects, left foot and right foot have the same behavior, then the sensor is placed on the right shoe of each subject. Each subject was asked to perform one experiment, which contained two identical sections of a sequence of four locomotion activities: walking, running, stair ascent and stair descent. Therefore, the performance of the second section will be improved compared to the first section, if the on-line algorithm can gradually learn the activity pattern of the subject. The four locomotion activities are referred to as D2A1 to D2A4 in the rest of this paper. The sensor sampling rate was set to 100 Hz, so the data length is long enough for the on-line EM algorithm. The sensing range of the acceleration and angular rate are ± 8 g and $\pm 1000°$/s, respectively.

The proposed SemiTMC-GMM model is compared with TMC-GMM to see the advancement of semi-Markov structure in recognizing lower limb locomotion activities. GMM is implemented by different κ to see the impact of the GMM components number that has on recognition accuracy.

5.1. SDA Dataset

The batch mode recognition is tested by a leave-one-out cross-validation (LOOCV) strategy, i.e., taking one subject for testing and the others for training, then make the test for all the subjects. The sliding window length of feature extraction is set to 5. The length is set based on our experience; normally it depends on the shortest gait or leg phase of the activities, and a window length larger than the shortest duration may conceal some information of the phase, while too small length may cause a badly calculated standard deviation. Both the SemiTMC-GMM and TMC-GMM model are involved in the validation, and the GMM mixture number κ is set to 1, 3, 6, and 9, respectively. Particularly when $\kappa = 1$, the conditioned observation density yields to the conventional Gaussian distribution.

The overall accuracy of batch mode recognition on SDA dataset is shown in Figure 4a. As it can be seen, SemiTMC-GMM achieves an accuracy improvement of about 2–3% compared to TMC-GMM. The proposed model reaches the highest accuracy of 86.00% when $\kappa = 6$, while the one of TMC-GMM is 83.76%. Meanwhile, TMC-HIST obtains the lowest accuracy of 77.91%. Figure 4b is the plot of Receiver Operating Characteristics (ROC) *w.r.t.* different κ. The ROC space takes the true positive rate (TPR) as vertical axis and false positive rate (FPR) as horizontal axis, a better recognition performance is more closer to the upper left corner. So we can see that SemiTMC-GMM (red markers) obtains a better performance than the one of TMC-GMM (cyan markers). Table 1 shows the sensitivity, specificity, F1 score, and Matthews correlation coefficient (MCC) of each individual activity. Particularly for the sensitivity of each individual activity, it equals to the accuracy of corresponding activity. Activities

D1A1 to D1A3 are recognized with relatively poor performance, it is because that these three activities are all walking and are very easily misclassified.
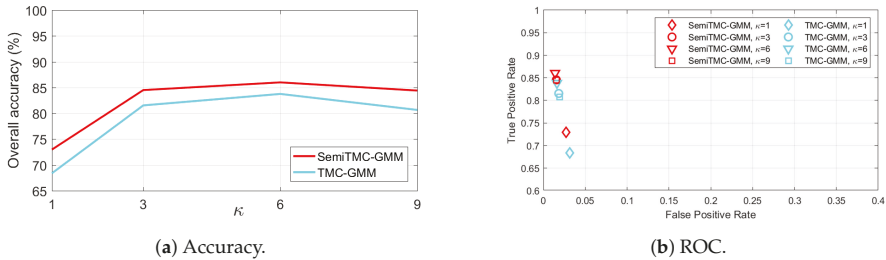

(**a**) Accuracy.

(**b**) ROC.

Figure 4. The batch mode recognition performance of the Sports and Daily Activities (SDA) dataset, of the SemiTMC-GMM, and TMC-GMM models, according to different GMM mixture number κ. ROC = Receiver Operating Characteristics.

Table 1. The sensitivity, specificity, F1 score, Matthews correlation coefficient (MCC) value of the batch mode recognition, for each activity of SDA dataset, using the sensor placed on right thigh. **Up**: TMC-HIST; **middle**: TMC-GMM when $\kappa = 6$; and **down**: SemiTMC-GMM when $\kappa = 6$.

		Activity					
		D1A1	**D1A2**	**D1A3**	**D1A4**	**D1A5**	**D1A6**
	Sensitivity	0.4900	0.5463	0.6997	0.9017	0.7885	1.0000
	Specificity	0.9392	0.9883	0.9649	0.9839	0.9222	0.9939
	F1 Score	0.4687	0.6574	0.6837	0.8708	0.6057	0.9709
	MCC	0.4128	0.6461	0.6511	0.8587	0.5781	0.9684
TMC-HIST		**D1A7**	**D1A8**	**D1A9**	**D1A10**	**D1A11**	**Total**
	Sensitivity	0.8308	0.7116	0.9489	0.9972	0.6618	0.7797
	Specificity	0.9911	0.9924	1.0000	1.0000	0.9813	0.9779
	F1 Score	0.8654	0.7966	0.9737	0.9986	0.7168	0.7826
	MCC	0.8535	0.7854	0.9715	0.9985	0.6936	0.7652
		D1A1	**D1A2**	**D1A3**	**D1A4**	**D1A5**	**D1A6**
	Sensitivity	0.6784	0.6797	0.5483	0.9146	0.8980	1.0000
	Specificity	0.9322	0.9993	0.9866	0.9689	0.9465	0.9995
	F1 Score	0.5777	0.8059	0.6525	0.8164	0.7305	0.9978
	MCC	0.5353	0.8067	0.6382	0.8025	0.7151	0.9975
TMC-GMM		**D1A7**	**D1A8**	**D1A9**	**D1A10**	**D1A11**	**Total**
	Sensitivity	0.8843	0.8917	0.8602	0.9876	0.8784	0.8383
	Specificity	0.9961	0.9940	0.9987	0.9998	0.9999	0.9838
	F1 Score	0.9197	0.9140	0.9184	0.9930	0.9348	0.8419
	MCC	0.9129	0.9059	0.9132	0.9923	0.9309	0.8319
		D1A1	**D1A2**	**D1A3**	**D1A4**	**D1A5**	**D1A6**
	Sensitivity	0.6672	0.7247	0.6182	0.9638	0.8767	0.9990
	Specificity	0.9457	0.9972	0.9860	0.9773	0.9563	0.9990
	F1 Score	0.6054	0.8273	0.7039	0.8752	0.7509	0.9944
	MCC	0.5644	0.8223	0.6862	0.8666	0.7327	0.9939
SemiTMC-GMM		**D1A7**	**D1A8**	**D1A9**	**D1A10**	**D1A11**	**Total**
	Sensitivity	0.9025	0.9410	0.8561	0.9956	0.9215	0.8606
	Specificity	0.9936	0.9922	0.9996	0.9994	1.0000	0.9860
	F1 Score	0.9175	0.9324	0.9208	0.9948	0.9590	0.8620
	MCC	0.9096	0.9255	0.9165	0.9943	0.9560	0.8516

As reported in [20], the classifiers of kNN, SVM and decision tree are tested on SDA dataset using all the five sensors. The accuracies are 78.97%, 84.03%, and 84.63%, respectively. In [21], the authors used SDA dataset and showed single sensor recognition accuracy of four classifiers: kNN, decision tree, discriminant analysis, and Naive Bayes. Specifically for the right leg sensor that is used in our paper, the four classifiers obtained accuracies of 81.72%, 78.78%, 87.03%, and 76.93%. Therefore, we can state that SemiTMC-GMM outperforms the generic classifiers like kNN, SVM, decision tree, and Naive Bayes, and obtains a similar performance of discriminant analysis. On the other hand,

the authors in [31] used CNN to recognize human daily activities in OPPORTUNITY dataset [32], which contains activities, such as open (close) door, open (close) drawer, clean table, and drink cup. They obtained an accuracy of 85.8% by using 23 body-worn sensors, 12 object sensors, and 21 ambient sensors. In addition, for the OPPORTUNITY dataset, [18] used CNN to obtain an accuracy of 77.99% by using the body-worn sensors only, while in [33], CNN obtained an accuracy of 93.75% on six activities: walking, stair ascent, stair descent, sitting, standing, and laying. Because of the prevalent CNNs can generate high dimensional features that suit for the recognition task, then CNNs may probably be suited for sophisticated activities. But it requires a huge quantity of data to train the network, and it is difficult to make CNN work for an adaptive on-line scenario. So, maybe CNN could obtain higher accuracy than SemiTMC-GMM, but we still believe that our proposed model is competent in some scenarios.

5.2. LMFIMU Dataset

For this dataset, the size of sliding window for computing features is set to 15, and it is also determined by our experience. Firstly, the batch mode recognition is performed using LOOCV strategy. Figure 5a shows the recognition accuracy *w.r.t.* different κ. The accuracy of SemiTMC-GMM when $\kappa = 9$ is 95.16%, while the one of TMC-GMM is 92.57%. Meanwhile, the choice of κ has less impact on accuracy for SemiTMC-GMM. The recognition accuracy obtained by TMC-HIST is 80.42%, which is lower than the ones of TMC-GMM and SemiTMC-GMM when $\kappa > 1$. The ROC shown in Figure 5b implies that SemiTMC-GMM outperforms the TMC-GMM in a general view. Table 2 shows the sensitivity, specificity, F1 score, and MCC of each individual activity. By comparing the batch mode recognition shown in Tables 1 and 2, both TMC-GMM and SemiTMC-GMM outperform TMC-HIST. It means that considering the observation correlation improves the recognition performance.
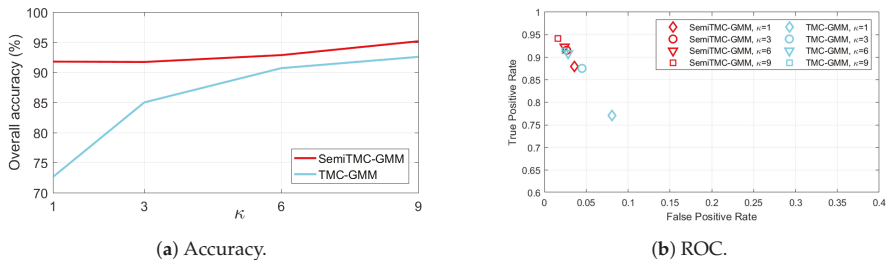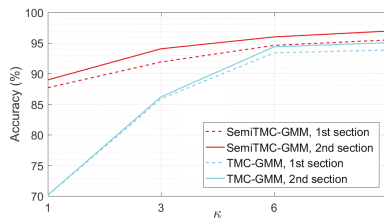

(a) Accuracy. **(b)** ROC.

Figure 5. The batch mode recognition performance of the Locomotion of Foot-mounted IMU (LMFIMU) dataset, SemiTMC-GMM, and TMC-GMM models, according to different GMM mixture number κ.

As a matter of fact, Figures 4a, 5a and 6a show that introducing semi-Markov structure into the TMC model can improve the accuracy. Meanwhile, using GMM with $\kappa > 1$ also improves the recognition significantly. But it does not mean that using a larger κ allows higher accuracy to be achieved. In Figure 4a, the accuracy when $\kappa = 9$ is slightly lower than that obtained when $\kappa = 6$, it is because the observation of SDA dataset is more closer to a GMM mixture of 6 densities. A too large value of κ may probably lead to an overfitting problem. It is sure that κ can be automatically acquired through the methods, such as BIC [34] and AIC [35], to make κ consistent with different activities. For simplification in this paper, we manually set κ to 6 for all the activities based on the experimental results.

Then, the on-line EM algorithm is performed to validate the adaptive on-line recognition performances. The proposed algorithm is implemented in MATLAB code, running on a 64-bit system computer with 3.2GHz CPU and 32G RAM. In the dataset, the average experiment time is 32.33 min, while the computing time of SemiTMC-GMM when $\kappa = 1, 3, 6, 9$ are 9.72, 14.72, 21.53, and 27.65 min, respectively. Thus, using on-line EM is applicable in on-line scenarios. The window length W_l for updating the parameters is set to 1000, which means that parameters are updated every 10 s.

Table 2. The sensitivity, specificity, F1 score, MCC value of the batch mode recognition, for each activity of LMFIMU dataset, using the sensor placed on right shoe. **Up**: TMC-HIST; **middle**: TMC-GMM when $\kappa = 9$; and **down**: SemiTMC-GMM when $\kappa = 9$.

		Activity				
		D2A1	**D2A2**	**D2A3**	**D2A4**	**Total**
TMC-HIST	Sensitivity	0.7007	0.9721	0.7705	0.9385	0.8454
	Specificity	0.9858	0.8931	0.9174	0.9595	0.9389
	F1 Score	0.8169	0.8258	0.6885	0.8596	0.7977
	MCC	0.7194	0.7833	0.6317	0.8382	0.7431
TMC-GMM	Sensitivity	0.9399	0.9475	0.9105	0.8590	0.9142
	Specificity	0.9720	0.9996	0.9512	0.9787	0.9754
	F1 Score	0.9547	0.9723	0.8327	0.8641	0.9060
	MCC	0.9130	0.9654	0.8044	0.8419	0.8812
SemiTMC-GMM	Sensitivity	0.9608	0.9829	0.9483	0.8749	0.9417
	Specificity	0.9831	0.9987	0.9634	0.9910	0.9841
	F1 Score	0.9713	0.9891	0.8799	0.9071	0.9368
	MCC	0.9445	0.9861	0.8600	0.8932	0.9210

(**a**) Accuracy.
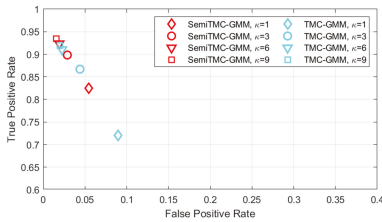
(**b**) ROC of first section.
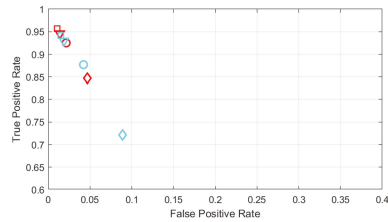
(**c**) ROC of second section.

Figure 6. The on-line mode recognition performance of the two experiment sections in LMFIMU dataset, according to different GMM mixture number κ.

Figure 6a shows the recognition accuracy obtained by LOOCV strategy. The solid lines are higher than the dashed lines which means that the on-line EM algorithm can improve the recognition performance. In addition, the GMM with $\kappa > 1$ can significantly improve the accuracy. When $\kappa = 9$, SemiTMC-GMM has an accuracy improved from 95.48% in the first section to 96.93% in the second section, TMC-GMM achieves an improvement from 93.83% to 95.04%. By contrast, in the adaptive on-line algorithm using TMC-HIST in our previous work, the accuracy was improved from 95.32% to 96.93%. However, this high accuracy is mainly because of the gait cycle complete detection in the adaptive on-line algorithm, which manually set the activity of all the samplings in one gait cycle to be identical. If the gait cycle complete detection is not used, TMC-HIST will fail in the on-line recognition, with the accuracies of 78.32% in the first section and 65.20% in the second section. Comparing SemiTMC (when $\kappa = 1$) and TMC-HIST, we can conclude that semi-Markov structure is more robust for recognizing the hidden states which have sojourn time. Figure 6b,c show the ROC of the two experiment sections, SemiTMC-GMM obtains better performance than TMC-GMM in both the two sections, and SemiTMC-GMM obtains a bigger improvement from the first to second section than the

other one. Therefore, the results indicate that both GMM density and semi-Markov structure improve the on-line recognition, and the combination the two improves the performance the most.

By comparing the two different sensor placements in SDA and LMFIMU dataset, the proposed algorithm shows that the sensor is not necessary to be placed at a specific place of the lower limb. In fact, the sensor can be placed in any position that implies the introduced gait phase and leg phase.

In order to understand the dynamic performance of the parameter updating, Figure 7 shows the recognition accuracy computed during the latest 10 s. Notice that the accuracies when $\kappa = 1$ are not displayed in Figure 7a,e because TMC obtains accuracies lower than 70% for D2A1 and D2A3. SemiTMC-GMM obtains a relatively fast convergence rate when κ equals to 6 and 9. The activities D2A1 and D2A2 reach high accuracy within 20 s in the first section of the experiment, 97.77% and 99.02%, respectively. By contrast, D2A3 and D2A4 are slower (take about 50 s) than the former two activities, and obtain lower accuracies of 92.04% and 89.48%, respectively. The main reason of this phenomenon is that the activity patterns of D2A3 and D2A4 vary much more differently among the subjects. But in a general view, we can still state that the on-line EM algorithm can dynamically improve the recognition accuracy to a reasonable level.
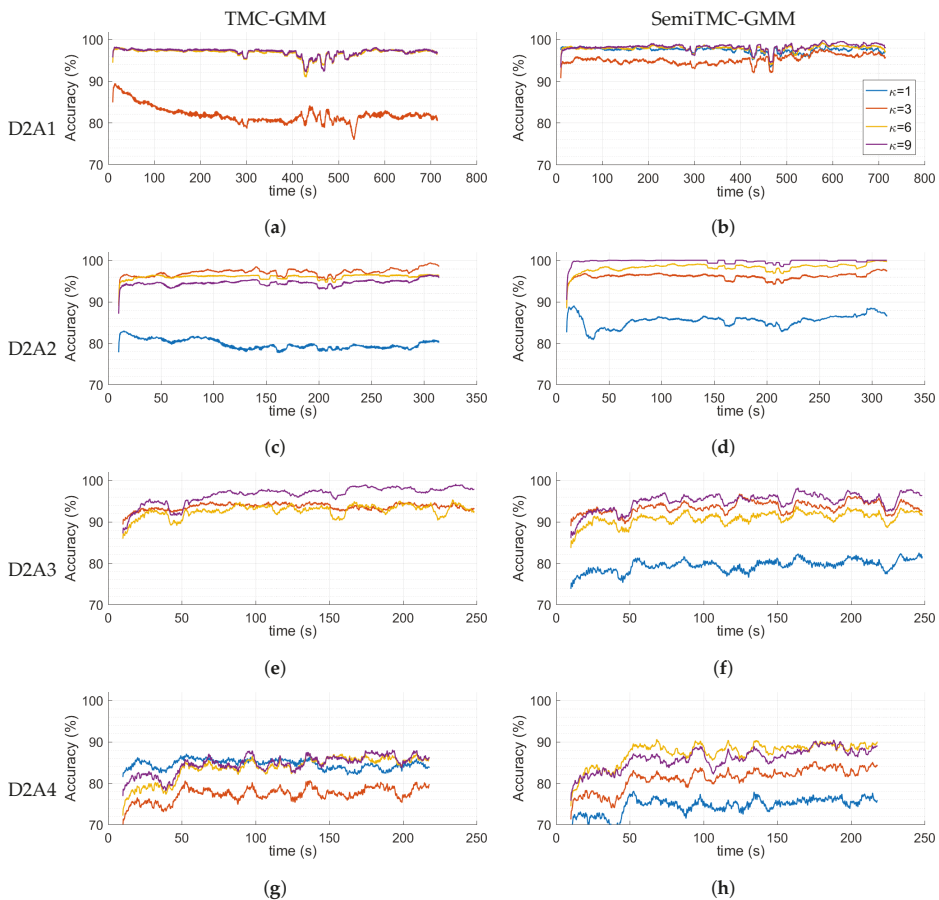

Figure 7. Recognition accuracy computed in the latest 10 s *w.r.t.* each activity of LMFIMU dataset. (**Left column**) TME-GMM; (**right column**) SemTMC-GMM.

Figure 8 displays the estimated gait cycles of each activity, when the model converged, obtained by TMC-GMM and SemiTMC-GMM, κ is set to 1 and 6. ω^x, ω^y and ω^z are the sliding mean of angular rate along the three axes of sensor. The features are 12-dimensional, but here we only display the acceleration of along the three axes to show how the gaits proceed. Hence, the estimated gait cycles are displayed *w.r.t.* four models, i.e., TMC, SemiTMC, TMC-GMM, and SemiTMC-GMM. In fact, the gait phases or leg phases are introduced in the model to improve the recognition accuracy of the lower limb locomotion activity. The figure shows that SemiTMC-GMM obtains the most regular gait cycle, with no fluctuation in short period and no missing detection. As a consequence, the well estimated gait or leg cycle obtained from SemiTMC-GMM leads to a higher activity recognition performance.

(**a**) D2A1.

(**b**) D2A3.

(**c**) D2A2.

(**d**) D2A4.
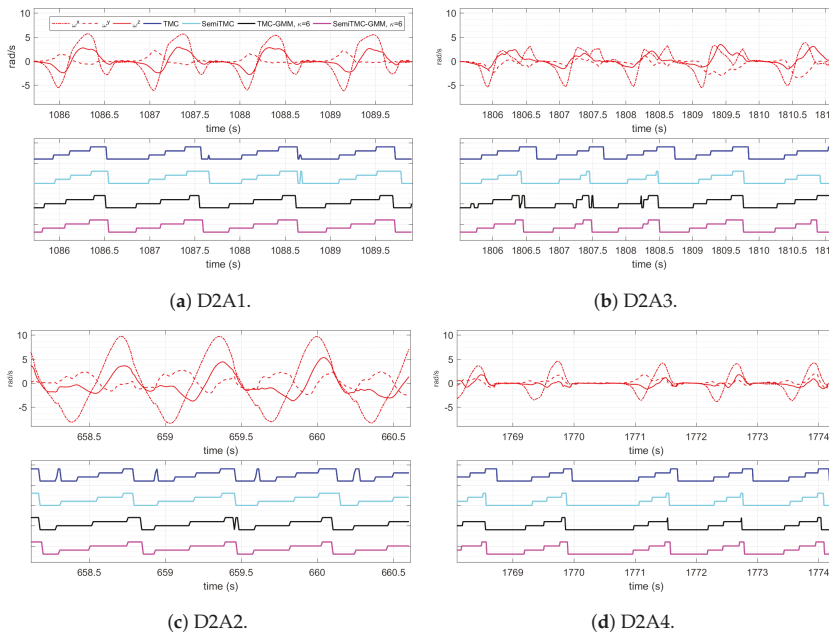
Figure 8. Estimated gait cycle of each activity. The blue, cyan, black, and magenta represent the gait obtained by TMC, SemiTMC, TMC-GMM, and SemiTMC-GMM, respectively.

6. Conclusions

In this paper, we proposed a wearable IMU-based algorithm for recognizing lower limb locomotion activities, with the help of introducing gait cycle or leg cycle into the model. The algorithm is based on the developed SemiTMC-GMM model, which better replicate the real motion. Our experiments show that semi-Markov structure and GMM density can better recover gait or leg cycles, which in return improve the activity recognition significantly. The adopted on-line EM algorithm can gradually improve the accuracy to a high level. The proposed algorithm is not only developed for the applications which require run-time activity recognition, but is also helpful to those applications that require gait cycles. For example, if using two sensors placed on both left and right legs of impaired people, then it is possible to develop a special SemiTMC-GMM model that detects their imbalanced gait phases. This can be beneficial for exoskeleton equipment to better assist impaired people in performing locomotion activities, by providing precise information to the equipment.

While there are still some limitations, the proposed algorithm only takes periodic lower limb locomotion into consideration; neither the static activity nor non-periodic lower limb locomotion activity is involved in our current work, such as standing and making turn. To distinguish static and motion activities, it is possible to include specific features into the observations. For example,

standard deviation will be close to zero when a person is in static; otherwise, it will vary according to the motion. Distinguishing periodic and non-periodic can be accomplished by periodic pattern mining method, such as fast Fourier transform-based [36] and principle component analysis-based [37] approaches. Our future work will focus on adopting more types of activities, including static activity and non-periodic locomotion activities.

Author Contributions: Conceptualization, H.L.; Methodology, H.L., S.D., W.P.; Software, H.L.; Validation, H.L., S.D. and W.P.; Formal analysis, H.L., S.D.; Investigation, H.L., S.D.; Resources, S.D., W.P.; Data curation, H.L.; Writing—original draft preparation, H.L.; Writing—review and editing, S.D., W.P.; Visualization, H.L.; Supervision, S.D., W.P.; Project administration, S.D.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, Z.; Lu, Y.; You, C.; Chiang, C. Deep learning for sensor-based rehabilitation exercise recognition and evaluation. *Sensors* **2019**, *19*, 887. [CrossRef] [PubMed]
2. Susnea, I.; Dumitriu, L.; Talmaciu, M.; Pecheanu, E.; Munteanu, D. Unobtrusive monitoring the daily activity routine of elderly people living alone, with low-cost binary sensors. *Sensors* **2019**, *19*, 2264. [CrossRef] [PubMed]
3. Elhoushi, M.; Georgy, J.; Noureldin, A.; Korenberg, M.J. A survey on approaches of motion mode recognition using sensors. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1662–1686. [CrossRef]
4. Waheed, S.A.; Khader, P.S.A. A novel approach for smart and cost effective IoT based elderly fall detection system using Pi camera. In Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 14–16 December 2017; pp. 1–4.
5. Hsu, Y.; Yang, S.; Chang, H.; Lai, H. Human daily and sport activity recognition using a wearable inertial sensor network. *IEEE Access* **2018**, *6*, 31715–31728. [CrossRef]
6. Xie, L.; Tian, J.; Ding, G.; Zhao, Q. Human activity recognition method based on inertial sensor and barometer. In Proceedings of the IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), Moltrasio, Italy, 26–29 March 2018; pp. 1–4.
7. Li, H.; Derrode, S.; Pieczynski, W. An adaptive and on-line IMU-based locomotion activity classification method using a triplet Markov model. *Neurocomputing* **2019**, *362*, 94–105. [CrossRef]
8. Gorynin, I.; Gangloff, H.; Monfrini, E.; Pieczynski, W. Assessing the segmentation performance of pairwise and triplet Markov models. *Signal Process.* **2018**, *145*, 183–192. [CrossRef]
9. Pieczynski, W.; Hulard, C.; Veit, T. Triplet Markov chains in hidden signal restoration. In *Image and Signal Processing for Remote Sensing VIII*; International Society for Optics and Photonics: Crete, Greece, 2003; Volume 4885, pp. 58–68.
10. Barshan, B.; Yüksek, M.C. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput. J.* **2014**, *57*, 1649–1667. [CrossRef]
11. Parri, A.; Yuan, K.; Marconi, D.; Yan, T.; Crea, S.; Munih, M.; Lova, R.M.; Vitiello, N.; Wang, Q. Real-time hybrid locomotion mode recognition for lower limb wearable robots. *IEEE/ASME Trans. Mechatron.* **2017**, *22*, 2480–2491. [CrossRef]
12. Chen, Z.; Zhu, Q.; Soh, Y.C.; Zhang, L. Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3070–3080. [CrossRef]
13. Wannenburg, J.; Malekian, R. Physical activity recognition from smartphone accelerometer data for user context awareness sensing. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 3142–3149. [CrossRef]
14. Zhao, H.; Wang, Z.; Qiu, S.; Wang, J.; Xu, F.; Wang, Z.; Shen, Y. Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion. *Inf. Fusion* **2019**, *52*, 157–166. [CrossRef]
15. Van Kasteren, T.; Englebienne, G.; Kröse, B.J. Activity recognition using semi-Markov models on real world smart home datasets. *J. Ambient Intell. Smart Environ.* **2010**, *2*, 311–325.
16. Sathyanarayana, A.; Joty, S.; Fernandez-Luque, L.; Ofli, F.; Srivastava, J.; Elmagarmid, A.; Taheri, S.; Arora, T. Impact of physical activity on sleep: A deep learning based exploration. *arXiv* **2016**, arXiv:1607.07034.

17. Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of deep learning in biomedicine. *Mol. Pharm.* **2016**, *13*, 1445–1454. [CrossRef] [PubMed]
18. Ordóñez, F.; Roggen, D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]
19. Ramasamy Ramamurthy, S.; Roy, N. Recent trends in machine learning for human activity recognition—A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, 1254. [CrossRef]
20. Bao, Y.; Chen, W. Automatic model construction for activity recognition using wearable devices. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Athens, Greece, 19–23 March 2018; pp. 806–811.
21. Rokni, S.A.; Ghasemzadeh, H. Autonomous training of activity recognition algorithms in mobile sensors: A transfer learning approach in context-invariant views. *IEEE Trans. Mob. Comput.* **2018**, *17*, 1764–1777. [CrossRef]
22. Schneider, T.; Helwig, N.; Schütze, A. Automatic feature extraction and selection for condition monitoring and related datasets. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 14–17 May 2018; pp. 1–6.
23. Rezaie, H.; Ghassemian, M. An adaptive algorithm to improve energy efficiency in wearable activity recognition systems. *IEEE Sens. J.* **2017**, *17*, 5315–5323. [CrossRef]
24. Dao, M.S.; Nguyen-Gia, T.A.; Mai, V.C. Daily human activities recognition using heterogeneous sensors from smartphones. *Procedia Comput. Sci.* **2017**, *111*, 323–328. [CrossRef]
25. Martindale, C.F.; Sprager, S.; Eskofier, B.M. Hidden Markov model-based smart annotation for benchmark cyclic activity recognition database using wearables. *Sensors* **2019**, *19*, 1820. [CrossRef] [PubMed]
26. Barbu, V.S.; Limnios, N. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: Their Use in Reliability and DNA Analysis*; Springer Science & Business Media: New York, NY, USA, 2009; Volume 191.
27. Yu, S.Z. *Hidden Semi-Markov Models: Theory, Algorithms and Applications*; Morgan Kaufmann: Amsterdam, The Netherlands, 2015.
28. Lapuyade-Lahorgue, J.; Pieczynski, W. Unsupervised segmentation of hidden semi-Markov non-stationary chains. *Signal Process.* **2012**, *92*, 29–42. [CrossRef]
29. Shetty, N.; Bendall, S. Understanding the gait cycle, as it relates to the foot. *Orthop. Trauma* **2011**, *25*, 236–240. [CrossRef]
30. Cappé, O. Online EM algorithm for hidden Markov models. *J. Comput. Graph. Stat.* **2011**, *20*, 728–749. [CrossRef]
31. Yang, J.; Nguyen, M.N.; San, P.P.; Li, X.; Krishnaswamy, S. Deep convolutional neural networks on multichannel time series for human activity recognition. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
32. Roggen, D.; Calatroni, A.; Rossi, M.; Holleczek, T.; Förster, K.; Tröster, G.; Lukowicz, P.; Bannach, D.; Pirkl, G.; Ferscha, A.; et al. Collecting complex activity datasets in highly rich networked sensor environments. In Proceedings of the 2010 Seventh International Conference on Networked Sensing Systems (INSS), Kassel, Germany, 15–18 June 2010, pp. 233–240.
33. Ronao, C.A.; Cho, S. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **2016**, *59*, 235–244. [CrossRef]
34. Bhat, H.S.; Kumar, N. *On the Derivation of the Bayesian Information Criterion*; School of Natural Sciences, University of California: Berkeley, CA, USA, 2010.
35. Arnold, T.W. Uninformative parameters and model selection using Akaike's Information Criterion. *J. Wildl. Manag.* **2010**, *74*, 1175–1178. [CrossRef]
36. Li, Z.; Ding, B.; Han, J.; Kays, R.; Nye, P. Mining periodic behaviors for moving objects. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 1099–1108.
37. He, Z.; Wang, X.S.; Lee, B.S.; Ling, A.C. Mining partial periodic correlations in time series. *Knowl. Inf. Syst.* **2008**, *15*, 31–54. [CrossRef]

Identification of Gait Motion Patterns Using Wearable Inertial Sensor Network

Kee S. Moon [1,*], Sung Q Lee [2,*], Yusuf Ozturk [3], Apoorva Gaidhani [1] and Jeremiah A. Cox [1]

1   Department of Mechanical Engineering, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA; apoorva2501@gmail.com (A.G.); cox_j_a@hotmail.com (J.A.C.)
2   Electronics and Telecommunications Research Institute, ICT, 218 Gajeong-ro, Yuseong-gu, Daejeon 34129, Korea
3   Department of Electrical and Computer Engineering, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA; yozturk@sdsu.edu
*   Correspondence: kmoon@sdsu.edu (K.S.M.); hermann@etri.re.kr (S.Q.L.); Tel.: +1-619-594-8660 (K.S.M.)

Abstract: Gait signifies the walking pattern of an individual. It may be normal or abnormal, depending on the health condition of the individual. This paper considers the development of a gait sensor network system that uses a pair of wireless inertial measurement unit (IMU) sensors to monitor the gait cycle of a user. The sensor information is used for determining the normality of movement of the leg. The sensor system places the IMU sensors on one of the legs to extract the three-dimensional angular motions of the hip and knee joints while walking. The wearable sensor is custom-made at San Diego State University with wireless data transmission capability. The system enables the user to collect gait data at any site, including in a non-laboratory environment. The paper also presents the mathematical calculations to decompose movements experienced by a pair of IMUs into individual and relative three directional hip and knee joint motions. Further, a new approach of gait pattern classification based on the phase difference angles between hip and knee joints is presented. The experimental results show a potential application of the classification method in the areas of smart detection of abnormal gait patterns.

1. Introduction

Early detection of symptoms enables better chances of treatment and sometimes complete cure of many diseases. Many health-related symptoms can be diagnosed in their primitive stages by monitoring body functions continuously. With the advancement in wearable sensor technology, many new pervasive monitoring systems are available in the market that can track physiological signals and wirelessly transmit them [1]. Many issues are related to gait disorders such as asynchronous body movement, asymmetry, loss of smoothness, and slowing of gait speed, etc. The understanding and identification of the human gait have generated much interest in many research fields beyond medical sciences such as biomechanics, etc. For the study of human gait, researchers have used various techniques in lab environments [2].

A gait cycle is described as the sequence and rhythm of foot/limb movements in the walking pattern of a person to achieve the motive of locomotion by propelling the body's center of gravity forward. A gait cycle comprises of a stance phase and a swing phase. The stance phase corresponds to the instance while walking when the body balances on both the feet, whereas the swing phase is the one in which the body balances on one foot while the other foot is off the ground to take the next step. Normal or healthy individuals have a normal gait, but certain neurological and musculoskeletal

disorders cause abnormality in gait pattern [3,4]. Notably, mobility and disability problems, including the gait disorder, tend to increase within the elderly population. Some gait cycle patterns generally change with age, such as gait velocity, double stance time, walking posture, and joint motion [5].

Gait assessment is one of the essential steps when diagnosing neurological conditions, orthopedic problems and medical conditions. [6]. Manifestations of an abnormal gait are the loss of symmetry of motion and timing between the left and right sides, which can cause difficulty in initiating or maintaining pace. The three-dimensional gait analysis enables the detection of angular deflections during gait; this is important for clinical purposes to be able to identify different disorders accurately [7]. A standard method for measuring gait cycle is to use optical tracking systems such as the Vicon motion analysis system. A set of cameras track the markers attached to the human body to measure and calculate the three-dimensional positions of the body segments. However, this tracking approach needs controlled environments and limits the area of operation in a laboratory room, where the camera and the equipment are located [8].

As an alternative solution, the application of the accelerometer provides a simple means for analyzing gait in real-time for monitoring physical activity levels and classifying movements [9–12]. Gyroscopes were applied in the measurement of angular rate and the angle of various joints by mounting multiple devices on the lower or upper extremities [12–15]. In this paper, a wireless IMU sensor network (i.e., two IMUs) is used to enable a three-dimensional angular assessment of the thigh and the knee for gait analysis [16]. The differential changes of the rotation angles are also used to reduce the influences of the drift and offset noises that are common in IMU raw data. In particular, this paper presents gait phase differences using the sagittal, frontal, and transverse directional angles of the knee joint as a novel parameter to classify normal and abnormal gait patterns. The preliminary experimental results demonstrate that the proposed parameter is promising and can carry out gait pattern identification.

2. Wearable Inertial Sensor Network

2.1. Sensor System

This paper describes the method to acquire and analyze the IMU data so that an abnormal gait from a combination of hip and knee motion can be detected. The main system configuration is a pair of IMU sensors to create a network of joints and links of the human leg (Figure 1). The study uses Bluetooth technology for wireless transmission, and an Android device or a PC acts as a host receiver for the data transmitted using these sensors. They have a host application installed to allow the collection of data in the receiver. A MATLAB program involving mathematical calculations provides the analysis of the sensor data.

The specifications of the wireless IMU system are as follows:

- The angular motion data from the IMU is in the form of quaternions.
- The wireless IMU sensor is small in size with a diameter of 23 mm.
- The wireless communication unit provides a low power communication interface between the data acquisition system and PC or any host where Bluetooth 4.0 is supported.
- A Nordic Semiconductor nRF51822 chip is used as a communication module.
- The nRF51822 communication system on chip (SoC) supports three kinds of data rates—2 Mbps, 1 Mbps, and 250 kbps.
- The transmission of raw signals has a sampling rate of 250 data points per seconds.
- The reconfigurable system interfaces the components using API (application programming interfaces) that can interface with 3rd party applications.
- An onboard Li-Polymer charge management controller MCP73831 recharges the sensor module.
- With Bluetooth Low Energy (BLE), two devices communicate only at each connection interval, so the minimum time between transfers will be the connection interval (between 7.5 ms to 4.0 s).

- The NRF24L01+ is a 2.4 GHz Radio-Frequency (RF) transceiver integrates an RF transceiver, a full-speed USB 2.0 compliant device controller, an 8-bit microcontroller, and flash memory.
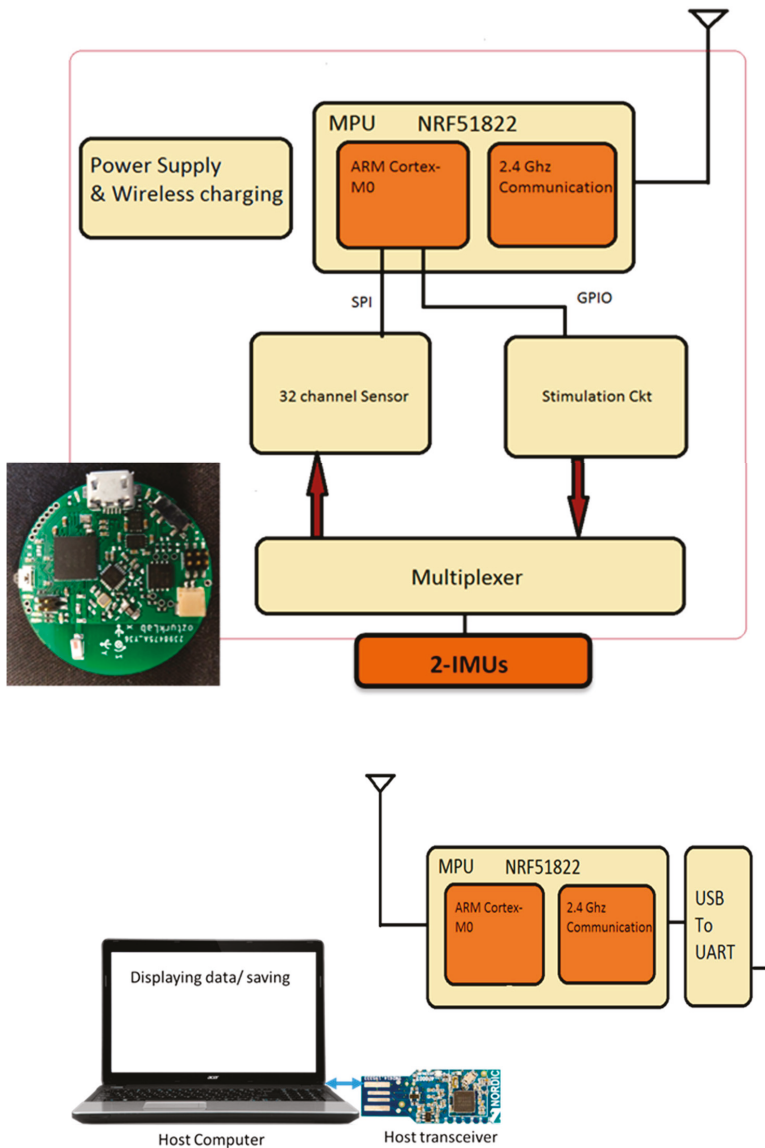

Figure 1. A custom designed IMU network to record leg movements: the sensor system diagram and the manufactured wireless IMU sensor circuit.

The motion sensor module consists of a 9-axis sensor made by 'Invensense', MPU9150. The motion processor acquires the data collected by the accelerometer, gyroscope, and magnetometer, to generate the data in the format of quaternions. The MPU-9150 features three 16-bit analog-to-digital converters (ADCs) for digitizing the gyroscope with a full-scale range of ± 250, ± 500, ± 1000, and ± 2000°/s (dps).

The noise rate is 0.005 (∘/s/√ Hz). However, our test showed about 5°/h stability of the sensor assembly at the 20 Hz sampling rate.

2.2. Gait Cycle

The body segments are referenced to the anatomical position and are described as occurring in three planes (Figure 2a). The body is divided into left and right halves by the sagittal plane. The frontal plane divides the body into anterior (front) and posterior (rear) portions. The body is divided into superior (upper) and inferior (lower) portions of the transverse plane. Joint motions during walking in the sagittal reference plane are shown in Figure 2b. The motion of the hip and knee in the sagittal plane are referred to as flexion and extension. Ankle movement in the sagittal plane is referred to as dorsiflexion and plantar flexion. Motions in the frontal plane are abduction and adduction, and motions in the transverse plane are internal and external rotations. Translating the center of mass of the human body is the primary task of walking. The sagittal plane is the parallel plane to the progression of walking. The left and right step during walking is a single gait cycle, which is a pattern of repeatable movements.
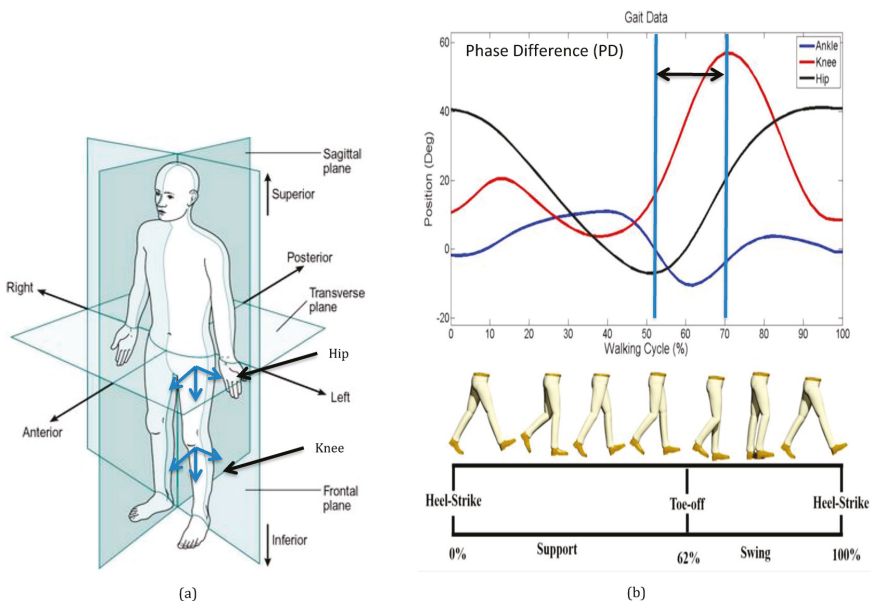

Figure 2. One gait cycle of human walking: (**a**) description of the hip and knee local coordinate system that are referenced to the anatomical position in three reference planes; (**b**) description of the phase difference motion of the hip and knee in the sagittal plane [17].

One gait cycle of human walking consists of two consecutive steps. A step is a period from right heel contact (RHC) to left heel contact (LHC). Two symmetric steps define a typical gait cycle containing single support (SS) and double support (DS) phase. During walking, the leg moves in a repeatable motion giving each person a unique walk. The leg moves by the multiple degrees of angular movements of the hip and knee joints to achieve a periodic pattern that enables the forward movement of the body [18,19]. Figure 2b shows that there is a time difference between the two peak swing motions of the hip and knee in the sagittal direction. It is assumed that phase differences exist in all the three-dimensional reference planes, which can characterize a unique gait cycle pattern.

2.3. 3-Dimensional Hip and Knee Joint Angles

As shown in Figure 3, a human leg, as a kinematic manipulator, consists of two physical links (i.e., thigh and shin) and joints (i.e., hip and knee), each having three rotational joints. Both the physical joints are imaginary. Physical joint A is located at the hip (hip joint), and physical joint B is situated in the knee. Each physical joint has three Degree of Freedom (DOF) imparting a robot system. Links A and B are also imaginary and connect physical joint A to link A and physical Joint B to link B, respectively. Sensor A is placed at the end of Link A and gives the orientation of link A to the base frame. Sensor B is situated at the end of link B and provides the orientation of link B to the base frame. It should be noted that the direction of both the sensors is maintained the same.

By defining the transformation matrices T_A^0 as the mapping of sensor A in the base frame 0 and T_B^A as the mapping of sensor B in the sensor A frame, they can be described as follows based on the Denavit and Hartenberg (D–H) convention. The four parameters of each link give a homogeneous transformation matrix A_i. As per D-H convention law, every A_i can be represented as a product of four basic transformations as shown below.

$$A_i = Rot_{z,\theta_i}\, Trans_{z,d_i}\, Trans_{x,a_i}\, Rot_{x,a_i} \tag{1}$$

$$A_i = \begin{bmatrix} c_{\theta_i} & -s_{\theta_i} & 0 & 0 \\ s_{\theta_i} & c_{\theta_i} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \alpha_i \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c_{\alpha_i} & -s_{\alpha_i} & 0 \\ 0 & s_{\alpha_i} & c_{\alpha_i} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$A_i = \begin{bmatrix} c_{\theta_i} & -s_{\theta_i}c_{\alpha_i} & s_{\theta_i}s_{\alpha_i} & a_i c_{\theta_i} \\ s_{\theta_i} & c_{\theta_i}c_{\alpha_i} & -c_{\theta_i}s_{\alpha_i} & a_i s_{\theta_i} \\ 0 & s_{\alpha_i} & c_{\alpha_i} & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Here, 's' stands for sine and 'c' stands for cosine of the respective angles. Equation (3) is the general form of the homogeneous transformation matrix, which varies with each link, based on the D-H parameters of that link.

The transformation from frame 0 to frame 1 of the model can now be obtained by simply substituting the D–H parameters of link 1 in the general form of the homogeneous transformation matrix. Likewise, the homogeneous transformation matrix can be derived for link 1 through 6. Let A_1^0 be the transformation from frame 0 to frame 1 or the mapping of frame 1 in frame 0. Then,

$$A_1^0 = \begin{bmatrix} c_{\theta_1} & 0 & -s_{\theta_1} & 0 \\ s_{\theta_1} & 0 & c_{\theta_1} & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

Similarly,

$$A_2^1 = \begin{bmatrix} c_{\theta_2} & 0 & s_{\theta_2} & 0 \\ s_{\theta_2} & 0 & -c_{\theta_2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A_3^2 = \begin{bmatrix} c_{\theta_3} & -s_{\theta_3} & 0 & 0 \\ s_{\theta_3} & c_{\theta_3} & 0 & 0 \\ 0 & 0 & 1 & d_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

$$A_4^3 = \begin{bmatrix} c_{\theta_4} & 0 & -s_{\theta_4} & 0 \\ s_{\theta_4} & 0 & c_{\theta_4} & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, A_5^4 = \begin{bmatrix} c_{\theta_5} & 0 & s_{\theta_5} & 0 \\ s_{\theta_5} & 0 & -c_{\theta_5} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6}$$

$$A_6^5 = \begin{bmatrix} c_{\theta_6} & -s_{\theta_6} & 0 & 0 \\ s_{\theta_6} & c_{\theta_6} & 0 & 0 \\ 0 & 0 & 1 & d_6 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

Since there are only 2 physical joints, A and B, each having 3 rotational joints whose origins coincide, the number of homogeneous transformation matrices can also be reduced from 6 to only 2. The mapping of the 1st physical joint to frame 0 is T_3^0. Let T_3^0 be the transformation from frame 0 to frame 3 or, analogously, the mapping of frame 3 in frame 0. Since the model distinctly consists of two physical links and joints, each having three rotational joints, the number of transformation matrices can be reduced from 6 to 2 as follows:

$$T_A^0 = A_1^0 \, A_2^1 \, A_3^2, \tag{8}$$

$$T_A^0 = \begin{bmatrix} c_1 c_2 c_3 - s_1 s_3 & -c_1 c_2 s_3 - s_1 c_3 & c_1 s_2 & d_A(c_1 s_2) \\ s_1 c_2 c_3 + c_1 s_3 & -s_1 c_2 s_3 + c_1 c_3 & s_1 s_2 & d_A(s_1 s_2) \\ -s_2 c_3 & s_2 s_3 & c_2 & d_A(c_2) \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

$$T_B^A = A_4^3 \, A_5^4 \, A_6^5, \tag{10}$$

$$T_B^A = \begin{bmatrix} c_4 c_5 c_6 - s_4 s_6 & -c_4 c_5 s_6 - s_4 c_6 & c_4 s_5 & d_B(c_4 s_5) \\ s_4 c_5 c_6 + c_4 s_6 & -s_4 c_5 s_6 + c_4 c_6 & s_4 s_5 & d_B(s_4 s_5) \\ -s_5 c_6 & s_5 s_6 & c_5 & d_B(c_5) \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

where $c_1, c_2, c_3, c_4, c_5, c_6$ stand for $\cos(\theta_1)$, $\cos(\theta_2)$, $\cos(\theta_3)$, $\cos(\theta_4)$, $\cos(\theta_5)$, $\cos(\theta_6)$, and $s_1, s_2, s_3, s_4, s_5, s_6$ are $\sin(\theta_1)$, $\sin(\theta_2)$, $\sin(\theta_3)$, $\sin(\theta_4)$, $\sin(\theta_5)$, $\sin(\theta_6)$, respectively.
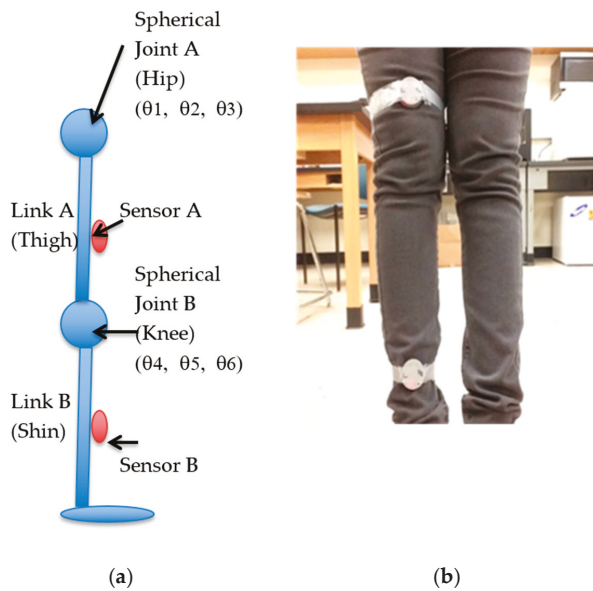

Figure 3. Three-dimensional joint angles: (**a**) description of a human leg model as a kinematic manipulator that consists of two physical links (i.e., thigh and shin) and joints (i.e., hip and knee), each having three rotational joints, the transformation; (**b**) an experimental sensor network setup.

The global homogeneous transformation matrix T_0^G is assumed to be an identity matrix,

$$T_0^G = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

The forward kinematic equation to obtain the sensor B orientation in/with respect to base frame (frame 0) is given as follows:

$$T_B^G = T_0^G * T_A^0 * T_B^A. \tag{13}$$

Rearranging the above equation to get orientation of sensor B with respect to the sensors A [11] or equivalently the orientation of frame 6 with respect to frame 3:

$$T_B^A = \left(T_A^0\right)^{-1} * \left(T_0^G\right)^{-1} * T_B^0 \tag{14}$$

where T_B^0 is the initialized version of $T_{B\,(original)}^0$.

The initialization of rotation matrix obtained from the sensors A and B is done in order to perfectly align sensor B (T_B^0) with sensor A (T_A^0) in the sagittal plane. To initialize $T_{A\,(original)}^0$ and $T_{B\,(original)}^0$ in the mid-sagittal plane, the rotation matrices of sensors A and B at the initially aligned position of the leg in the sagittal plane are calculated. Thus, the calibrated rotation matrix of sensor A (T_A^0) is given by

$$T_{A\,(inv)}^0 = \left[T_{A\,(original)}^0(at\ initial\ position)\right]^{-1}, \tag{15}$$

$$T_A^0 = T_{A\,(original)}^0 * T_{A\,(inv)}^0. \tag{16}$$

Similarly, the calibrated rotation matrix of sensor B (T_B^0) is given by

$$T_{B\,(inv)}^0 = \left[T_{B\,(original)}^0(at\ initial\ position)\right]^{-1}, \tag{17}$$

$$T_B^0 = T_{B\,(original)}^0 * T_{B\,(inv)}^0. \tag{18}$$

Note that, in Equations (16) and (18), only the inverse of the first matrices of $T_{A\,(original)}^0$ and $T_{B\,(original)}^0$ are taken so that the first matrices of initialized T_A^0 and T_B^0 become an identity matrix always.

Next, by applying inverse kinematics to T_A^0, the 3-D hip joint angles $\theta_1, \theta_2, \theta_3$ are obtained.

$$r_{33} = c_2, \quad \theta_2 = \cos^{-1}(r_{33}) \tag{19}$$

$$r_{31} = -s_2 c_3, \quad \theta_3 = \cos^{-1}\left(\frac{-r_{31}}{s_2}\right) \tag{20}$$

$$r_{13} = c_1 s_2, \quad \theta_1 = \cos^{-1}\left(\frac{r_{13}}{s_2}\right) \tag{21}$$

Similarly, by applying inverse kinematics to T_B^A, the 3-D knee joint angles $\theta_4, \theta_5, \theta_6$ (Figure 3b) can by calculated by

$$r_{33} = c_5, \quad \theta_5 = \cos^{-1}(r_{33}), \tag{22}$$

$$r_{31} = -s_5 c_6, \quad \theta_6 = \cos^{-1}\left(\frac{-r_{31}}{s_5}\right), \tag{23}$$

$$r_{13} = c_4 s_5, \quad \theta_4 = \cos^{-1}\left(\frac{r_{13}}{s_5}\right). \tag{24}$$

3. Experimental Results

Normal gait requires sophisticated interaction of many systems of the body, such as strength, coordination, and sensation. For example, aging is an increasing global public health issue. When aging causes weakening strength and sensation, the interacting systems may result in abnormal gait or walking abnormality. In this section, a preliminary experimental result for the recognition of normal and abnormal gait types from the calculated sets of 3-D hip and knee joint angles (Equations (12)–(17)) are described. Hemiplegic gait is one of the results of post-stroke patients who have weak flexor muscle movement during the gait cycle. The knee is stiff, hyperextends during stance, and does not regularly flex during the swing. The contralateral step often advances to meet the position of the paralyzed limb, instead of periodically moving forward beyond it. These movements include raising the pelvis to clear the paralyzed leg and circumduction, an abnormal shift. As a result, the toe traces a semicircle on the floor, first moving outward and then inward as it advances, instead of a normal straightforward movement [20–23] (Figure 4).
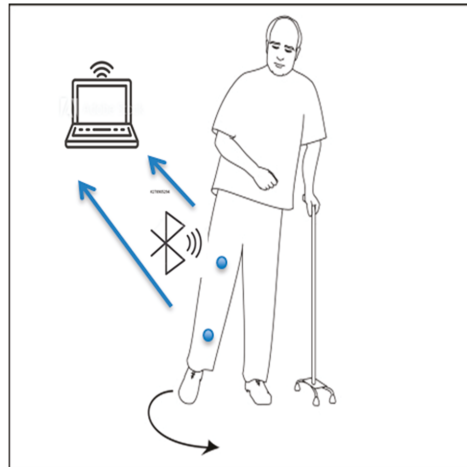

Figure 4. Description of a human leg movement during hemiplegic gait experiment.

This experiment provides the results of three-dimensional gait analysis performed on a single participant (i.e., an author). The experiments were performed multiple times to ensure the repeatability of the results and to obtain a statistically significant data set. A PC was then used to record and process the data. The experiments were completed the same way each time but on different days. There was only a single individual that performed the test, and thus, the sampling size was not statistically significant. However, for proof of concept purposes, the test was successful and serves as motivation to repeat the tests on a larger sample size in the future. Figure 5 shows the two kinds of gait patterns that were considered: hemiplegic and normal gaits. Three-dimensional kinematics of the hip and knee joint angles were recorded and calculated. The figure shows the relative angles of the hip and knee during the gait cycle along with extension and flexion in the sagittal plane, the frontal plane, and the transverse plane. From the figure, it can be noticed the changed coordination phase of the hip and knee movements of the Hemiplegic gait.
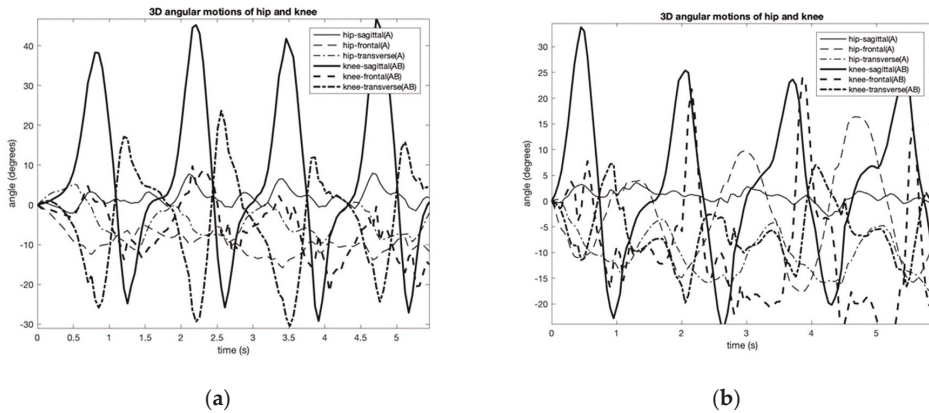
(**a**)   (**b**)

Figure 5. An example of calculated 3-D hip and knee joint angle patterns associated with normal (**a**) (periods of 1.3 s) and simulated hemiplegic walking (**b**) (periods of 1.8 s).

In this paper, a new method to calculate the parameter of the three-dimensional phase difference angles between the hip and knee movements is described to identify biomechanical changes in gait patterns. The phase difference angle analysis aims to establish the distinct conditions for mapping of the characteristic patterns of the joint motions in the performing of specific leg movements. The phase difference angles at time t and the *i*th-directional plane (i.e., sagittal, frontal, and transverse) is given by

$$\varphi_t^i = 360° \, f \, \Delta t, \tag{25}$$

where f is gait cycle frequency and Δt is the time difference (PD) between the hip and the knee angles.

This study uses the instantaneous signal (i.e., the temporal derivative of the oscillation) since it is useful for describing the non-monochromatic nature of the joint angle signals. Further, it also helps to reduce the effect on noises in IMU signal measurement, such as the signal shift and offset errors. Figure 6 shows the normalized gait phase diagrams of the hip and knee joint angles of Figure 5. From the diagram, the phase deference angles (i.e., hip-sagittal vs. knee-sagittal; hip-sagittal vs. knee-frontal; hip-sagittal vs. knee-transverse) are calculated using Equation (25). All phase deference angles of the two gait patterns are summarized in Figure 7. The figure indicates that all the sagittal, frontal, and transverse directions show a different pattern between the normal and the hemiplegic gait of the subject. Table 1 summarizes the experimental results from multiple trials of the subject.

Table 1. The experimental results of the normalized gait phase angles with normal and simulated hemiplegic walking.

	Gait	Hip-Sagittal VS. Knee-Sagittal (deg.)	Hip-Sagittal VS. Knee-Frontal (deg.)	Hip-Sagittal VS. Knee-Transverse (deg.)
Normal	AVG [1]	101.41	91.70	204.37
Normal	STD [1]	9.39	21.90	6.69
Hemiplegic	AVG [2]	−62.61	−27.22	−10.26
Hemiplegic	STD [2]	45.62	54.20	7.04

[1] The gait cycle frequency is 0.77 Hz. [2] The gait cycle frequency is 0.56 Hz.
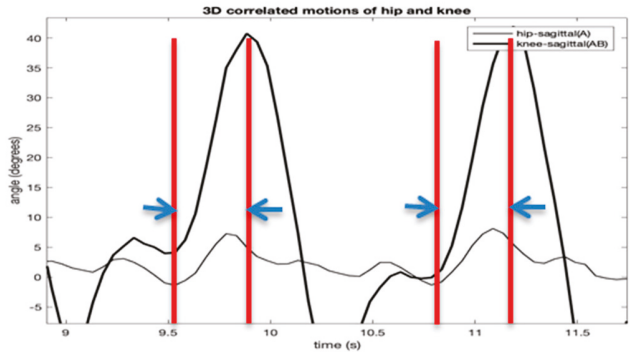
(**a**)
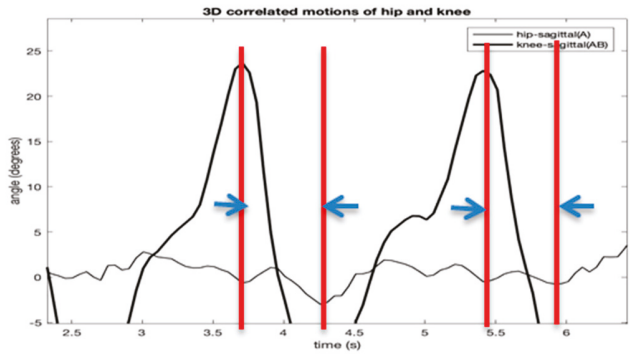
(**b**)

Figure 6. The normalized gait phase diagrams of the hip and knee joint angles of Figure 5 with normal (**a**) and simulated hemiplegic walking (**b**) (hip-sagittal vs. knee-sagittal).
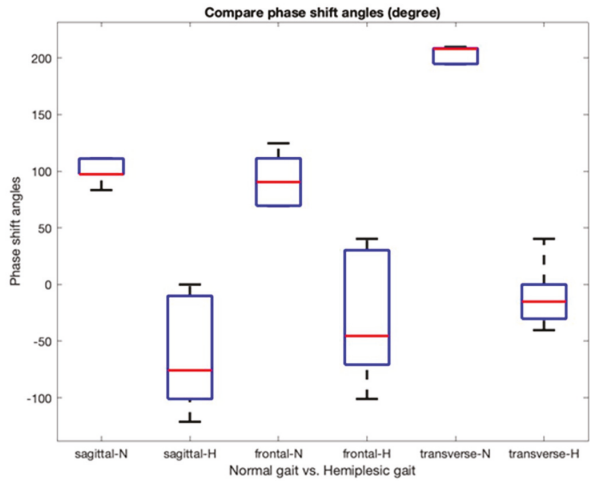

Figure 7. The normalized gait phase diagrams of the hip and knee joint angles of Figure 5 (i.e., hip-sagittal vs. knee-sagittal).

4. Discussion

The gait experiment with the two-IMU network system showed promising results in measuring three-dimensional joint angles. One of the main contributions of this paper is to provide a procedure to calculate the three-dimensional relative angles of the knee joint with respect to the hip joint using a pair of IMUs. It seems that the IMU-based method has lower accuracy compared to an optical camera-based method, such as the Vicon system [24]. However, the preliminary experimental results showed similar waveforms of hip and knee angles with comparable peak heights [24]. Unknown initial joint angles and error accumulation in the integral value of the gyroscope were still a significant problem. Further, there were some minor differences when observing the results, such as some abrupt changes at peaks or valleys during motion. There could be several reasons for the differences: human error or sensitivity of the sensors. Human error could be from mounting the sensors. If there was some error from the mounting, it did not appear to be significant.

In the identification of the characteristics of human gait, the study proposes using phase difference angle patterns. Figure 7 and Table 1 show evidence of the phase difference angle changes between the hip and joint movements during hemiplegic walking. Table 1 shows that the phase difference angles in all the directions were decreasing consistently.

It is known that a stroke commonly results in hemiplegic gait cycle impairments that are associated with decreased hip and knee coordination and limited leg muscle strength. These impairments often result in biomechanical changes during walking. The figures reflect the effects of reduced control of the leg swing that results in the decreased phase difference angles. The three-dimensional phase difference angle patterns are suitable for training machine learning methods to detect gait changes during walking [25,26].

5. Conclusions

The work contributes to classifying the characteristic joint movement patterns of the hip and the knee joints during the walking. The study conducted the following three steps to recognize the phase difference angle changes: i) calculation of 3-D joint angles of the hip and knee from the two-IMU sensor network system; ii) calculation of the normalized gait phase diagrams; iii) estimation of the gait phase difference angles for classifying abnormality in gait patterns. The work contributes to IMU signal analysis by introducing the following techniques: instantaneous signal normalization to reduce noise from the acquisition and the gait phase difference angle as a technique for the decomposition of signals. The methods proposed in the paper are suitable for establishing the patterns of different gait movements. The proposed techniques contribute in many ways to state-of-the-art identification, modeling, decomposition, and prediction of IMU signal analysis for various body motion applications.

Author Contributions: Conceptualization, K.S.M.; methodology, K.S.M. and S.Q.L.; software, K.S.M. and Y.O.; validation, A.G., and J.A.C.; investigation, K.S.M. and S.Q.L.; resources, K.S.M. and S.Q.L.; data curation, K.S.M. writing—original draft preparation, K.S.M.; writing—review and editing, Y.O. and S.Q.L.; supervision, K.S.M.; project administration, K.S.M.; funding acquisition, K.S.M. and S.Q.L.

References

1. Taj-Eldin, M.; Ryan, C.; O'Flynn, B.; Galvin, P. A Review of Wearable Solutions for Physiological and Emotional Monitoring for Use by People with Autism Spectrum Disorder and Their Caregivers. *Sensors (Basel)* **2018**, *18*, 4271. [CrossRef] [PubMed]

2. Muro-De-La-Herran, A.; Garcia-Zapirain, B.; Mendez-Zorrilla, A. Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications. *Sensors* **2014**, *14*, 3362–3394. [CrossRef]

3. Mannini, A.; Trojaniello, D.; Cereatti, A.; Sabatini, A.M. A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington's Disease Patients. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4732167/ (accessed on 21 January 2016).

4. Tao, W.; Liu, T.; Zheng, R.; Feng, H. Gait Analysis Using Wearable Sensors. *Sensors* **2012**, *12*, 2255–2283. [CrossRef] [PubMed]

5. Gait Disorders in the Elderly Geriatrics. Available online: https://www.msdmanuals.com/professional/geriatrics/gait-disorders-in-the-elderly/gait-disorders-in-the-elderly (accessed on 15 November 2019).

6. Creaby, M.W.; Bennell, K.L.; Hunt, M.A. Gait Differs Between Unilateral and Bilateral Knee Osteoarthritis. *Arch. Phys. Med. Rehabil.* **2012**, *93*, 822–827. [CrossRef] [PubMed]

7. Tadano, S.; Takeda, R.; Miyagawa, H. Three Dimensional Gait Analysis Using Wearable Acceleration and Gyro Sensors Based on Quaternion Calculations. *Sensors* **2013**, *13*, 9321–9343. [CrossRef]

8. Seel, T.; Raisch, J.; Schauer, T. IMU-Based Joint Angle Measurement for Gait Analysis. *Sensors* **2014**, *14*, 6891–6909. [CrossRef]

9. Godfrey, A.; Conway, R.; Meagher, D.; Ólaighin, G. Direct measurement of human movement by accelerometry. *Med. Eng. Phys.* **2008**, *30*, 1364–1386. [CrossRef] [PubMed]

10. Karantonis, D.; Narayanan, M.; Mathie, M.; Lovell, N.; Celler, B. Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 156–167. [CrossRef] [PubMed]

11. Anwary, A.; Yu, H.; Vassallo, M. An Automatic Gait Feature Extraction Method for Identifying Gait Asymmetry Using Wearable Sensors. *Sensors* **2018**, *18*, 676. [CrossRef] [PubMed]

12. Coste, C.A.; Sijobert, B.; Pissard-Gibollet, R.; Pasquier, M.; Espiau, B.; Geny, C. Detection of Freezing of Gait in Parkinson Disease: Preliminary Results. *Sensors* **2014**, *14*, 6819–6827. [CrossRef] [PubMed]

13. Filippeschi, A.; Schmitz, N.; Miezal, M.; Bleser, G.; Ruffaldi, E.; Stricker, D. Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion. *Sensors* **2017**, *17*, 1257. [CrossRef] [PubMed]

14. Leardini, A.; Lullini, G.; Giannini, S.; Berti, L.; Ortolani, M.; Caravaggi, P. Validation of the angular measurements of a new inertial-measurement-unit based rehabilitation system: Comparison with state-of-the-art gait analysis. *J. Neuroeng. Rehabil.* **2014**, *11*, 136. [CrossRef]

15. Nguyen, H.; Lebel, K.; Boissy, P.; Bogard, S.; Goubault, E.; Duval, C. Auto detection and segmentation of daily living activities during a Timed up and Go task in people with Parkinson's disease using multiple inertial sensors. *J. Neuroeng. Rehabil.* **2017**, *14*, 2. [CrossRef]

16. Gaidhani, A.; Moon, K.; Ozturk, Y.; Lee, S.; Youm, W. Extraction and Analysis of Respiratory Motion Using Wearable Inertial Sensor System during Trunk Motion. *Sensors* **2017**, *17*, 2932. [CrossRef]

17. Cestari Soto, M. Variable-Stiffness Joints with Embedded Force Sensor for High-Performance Wearable Gait Exoskeletons. Ph.D. Thesis, Universidad Politécnica de Madrid, Madrid, Spain, January 2017. [CrossRef]

18. Cimolin, V.; Galli, M. Summary measures for clinical gait analysis: A literature review. *Gait Posture* **2014**, *39*, 1005–1010. [CrossRef] [PubMed]

19. The Gait Cycle. *Orthopaedic Surgery Review*; Thieme: New York, NY, USA, 2009.

20. Moseley, A.; Wales, A.; Herbert, R.; Schurr, K.; Moore, S. Observation and analysis of hemiplegic gait: Stance phase. *Aust. J. Physiother.* **1993**, *39*, 259–267. [CrossRef]

21. Moore, S.; Schurr, K.; Wales, A.; Moseley, A.; Herbert, R. Observation and analysis of hemiplegic gait: Swing phase. *Aust. J. Physiother.* **1993**, *39*, 271–278. [CrossRef]

22. Kim, Y.; Yi, J. Gait Analysis in Normal and Hemiplegic Patients Using Accelerometers (Gait & Motion Analysis). *Proc. Asian Pac. Conf. Biomech.: Emerg. Sci. Technol. Biomech.* **2004**, *1*, 113–114. [CrossRef]

23. Rossi, S.; Colazza, A.; Petrarca, M.; Castelli, E.; Cappa, P. Gait analysis of hemiplegic and healthy subjects in different leg loading conditions. *Gait Posture* **2012**, *35*. [CrossRef]

24. Tugui, R.D.; Antonescu, D.; Nistor, M.I.; Bucur, D.; Kostrakievici, S. Gait Analysis in Cerebral Palsy Using Vicon System. *Proc. Manuf. Syst.* **2012**, *7*, 99–104.

25. Das, D.; Chakrabarty, A. Human gait based gender identification system using Hidden Markov Model and Support Vector Machines. In Proceedings of the International Conference on Computing, Communication & Automation, Noida, India, 15–16 May 2015. [CrossRef]

26. Yoo, J.-H.; Hwang, D.; Nixon, M.S. Gender Classification in Human Gait Using Support Vector Machine. In *Advanced Concepts for Intelligent Vision Systems Lecture Notes in Computer Science*; Springer: New York, NY, USA, 2005; pp. 138–145. [CrossRef]