Detection of semi-standardized gait tests from free-living inertial sensor recordings in Parkinson's disease

Bachelor's Thesis in Medical Engineering

submitted
by

Annika Mücke

born 12.10.1997 in Schlema

Written at

Machine Learning and Data Analytics Lab (CS 14)
Department of Computer Science
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Advisors:
Martin Ullrich M. Sc., Arne Küderle M. Sc., Dr.-Ing. Felix Kluge, Prof. Dr. Björn Eskofier
(Machine Learning and Data Analytics Lab, FAU Erlangen-Nürnberg)
Dr. phil. Heiko Gaßner
(Department of Molecular Neurology, University Hospital Erlangen)

Started: 04.05.2020

Finished: 02.10.2020

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 2. Oktober 2020

Übersicht

Ein auffälliges Gangbild ist ein charakteristisches Symptom für neurodegenerative Erkrankungen des Motorsystems, wie beispielsweise die Parkinson-Krankheit. Um Gang objektiv zu analysieren, können inertiale Messeinheiten verwendet werden. Da diese portabel sind, können sie auch zur Langzeitüberwachung von Patienten im häuslichen Umfeld verwendet werden, was eine umfassende Einschätzung des Krankheitsstadiums erlaubt. Um die Auswertung von Sensordaten, die im häuslichen Umfeld aufgenommen wurden, zu erleichtern und die Ergebnisse besser vergleichen zu können, sind standardisierte Testprotokolle, wie der 4x10m-Test, hilfreich. Solche Tests werden normalerweise bei klinischen Untersuchungen angewendet, lassen sich aber auf das Heimumfeld übertragen.
Es gibt bisher keinen Algorithmus, der eine automatische Erkennung solcher Gangtests in Aufnahmen aus dem häuslichen Umfeld ermöglicht, weswegen die Tests manuell von den Patienten annotiert werden müssen. Um den dadurch generierten Aufwand zu verringern, war das Ziel dieser Arbeit die Entwicklung von Algorithmen, die 4x10m-Tests automatisch in Ganztagesdatensätzen erkennen. Dabei wurden die 4x10m-Tests stets serienweise in verschiedenen, von Patienten selbst gewählten, Geschwindigkeitsstufen durchgeführt. Folglich mussten zuerst diese 4x10m-Testserien im Gesamtsignal erkannt und anschließend unter Zuordnung der jeweiligen Geschwindigkeitsstufen in einzelne 4x10m-Tests aufgeteilt werden.
Für den ersten Teil wurde ein auf Dynamic Time Warping basierender Ansatz gewählt, der sich an einem bereits existierenden und für klinische Daten validierten Algorithmus orientiert und eine Weiterentwicklung dessen darstellt. Verschiedene Methoden für Dynamic Time Warping wurden verglichen und Eingabeparameter optimiert. Dafür wurden Daten aus insgesamt 153 Tagesaufnahmen von 12 Patienten verwendet, die 418 4x10m-Testserien enthielten. Eindimensionales Dynamic Time Warping unter separater Verwendung der Daten zweier Fußsensoren sowie eines Hüftsensors erreichte mit einem schrittbasierten F1-Score von 80,5% und einem gebietsbasierten F1-Score von 92,8% die besten Ergebnisse.
Für den zweiten Teil wurden für jede 4x10m-Testerie die Punkte der stärksten Änderung in der Gangeschwindigkeit berechnet. Diese wurden als Grenzen der einzelnen 4x10m-Tests definiert und anschließend die jeweilige Geschindigkeitsstufe zugeordnet. Als Testdaten wurden insgesamt 298 4x10m-Testserien von 12 Patienten verwendet. Für die optimalen Parameter ergab sich ein F1-Score von 91,5%.

Die entwickelten Algorithmen tragen zu einer vereinfachten automatischen Auswertung von Gangdaten, die im Kontext von Langzeitüberwachung aufgenommen wurden, bei. Ein Vergleich der resultierenden Gangparameter mit den aus den manuellen Annotationen Berechneten wird weiterhin zeigen, ob die Genauigkeit für eine vollautomatische Erkennung von 4x10m-Tests ausreichend ist. Nichtsdestotrotz können die Algorithmen im Rahmen eines intelligenten Annotationstools angewendet werden.

Abstract

Pathological gait is a characteristic symptom for neurodegenerative disorders affecting the motor system, such as Parkinson's disease (PD). For objective mobile gait analysis, inertial sensors can be utilized. Their portability and unobtrusive character enables the application for long-term monitoring of patients in their home environment. Inertial sensor recordings from free-living settings provide a comprehensive insight into the patient's disease progression and support for example fall risk assessment. In order to obtain a consistent evaluation of these free-living recordings and to achieve comparability and reproducibility of results across studies, standardized protocols usually performed during clinical visits, such as the 4x10 meter walking test (4x10MWT), can be transferred to the home environment.

So far, there is no method for automatic detection of standardized gait tests in free-living recordings. Therefore, manual annotations need to be set by the patients. To remove the burden of the annotations from the patients and at the same time simplify the analysis of free-living recordings, the goal of this thesis was the development of a tool for automatic detection of 4x10MWTs from full-day inertial sensor recordings. In the underlying data, these 4x10MWTs were performed in the form of so-called *gait test series*, including three repetitions of the 4x10MWT in different self-selected gait speed levels. For this purpose, a pipeline consisting of two parts was developed. The first part was focused on the automatic detection of *gait test series* within free-living inertial sensor data, whereas the second part aimed at the separation of single 4x10MWTs from detected *gait test series* and the assignment to the corresponding gait speed levels.

For the first part, a detection algorithm based on dynamic time warping (DTW), which has previously been validated for clinical recordings, was adapted. Different DTW approaches were compared and relevant hyperparameters were optimized using a data set consisting of 153 full-day recordings of 12 patients, including 418 *gait test series*. Single-dimensional subsequence dynamic time warping (ssDTW) using both feet and hip sensor signals as input achieved the best performance with a stride-based *F1-score* of 80.5% and an area-based *F1-score* of 92.8%.

For the second part, the steepest changes of gait velocity were detected within each *gait test series*. Based on that, 4x10MWT borders were determined and speed levels assigned. The algorithm was optimized using a subset of 298 *gait test series* and achieved an *F1-score* of 91.5%.

The developed algorithms will contribute to a simplified automatic analysis of free-living inertial sensor-based gait recordings. Whether the accuracy is sufficient for fully automatic recognition requires further investigations. Calculating the gait parameter outcome will reveal, if relevant information is lost compared to the manual labels, or if the accuracy of the proposed algorithm is sufficient. Nevertheless, the pipeline can be used as the basis for a smart labeling tool.

Contents

Chapter 1

Introduction

Age-related neurodegenerative diseases have recently attracted a lot of attention in research due to their irreversibility and a lack of effective treatment [Hun10]. Parkinson's disease (PD) is one of the most common examples of such neurodegenerative diseases [Buc19]. It mainly affects the motor system [Jan08] and has a prevalence of 0.5-1% for people aged between 65 and 69 and up to 3% for people older than 80 [Tan96]. Therefore, by 2030 between 8.7 and 9.3 million people are expected to be affected by PD [Dor07]. Since it is an incurable disease, treating the symptoms and preventing side effects as much as possible is an important target in order to improve the quality of life of the affected patients.

The cardinal symptoms of PD are tremor, rigidity, bradykinesia, and postural instability [Wal06, Jan08]. Hence, patients show movement impairments in general and a pathological gait in particular [Jan08].

There are different ranking scales applied in clinical practice for classifying patients' symptoms, such as the Unified Parkinson's Disease Rating Scale (UPDRS) or the Hoehn & Yahr (H&Y) scale [Sko17]. They are implemented according to standardized protocols, allowing a physician to assess the subject's movement impairments [Sko17, Wal06]. However, a drawback of those rating methods is the absence of quantitative measurements to evaluate motor symptoms, as the rating can vary with the assessor and common impairments such as shuffling gait cannot be categorized into a discrete scale [Sch17]. Furthermore, detailed information such as small within-person changes might be missed [Buc19].

To provide a more objective monitoring for PD patients and in order to be able to allow individual treatment, additional technical measurement tools can be used. Gold standard systems for gait analysis, e.g. camera-based motion capture systems [Car13] or instrumented walkways [McD01], require a costly motion laboratory with limited relation to real-world gait. Sensor-based gait

analysis using wearable inertial measurement units (IMU) is an unobtrusive and less expensive alternative [Bam08, Klu13, Liu09, Sch17], which has been patented already (Appendix A.2). IMUs include an accelerometer and a gyroscope recording 3-d acceleration and 3-d angular velocities, respectively. The resulting signals enable a precise estimation of spatio-temporal parameters, such as stride time and gait velocity, and inter-stride variation [Sch17]. The accuracy of sensor-based gait parameter estimations has been validated against the gold standard systems for healthy subjects of different age groups [Kan15, Mar10] as well as subjects with movement impairments [Ber18, Ram15].
Using sensor-derived gait characteristics it is possible to differentiate between healthy and parkinsonian gait [Bar11], objectively classify a patients stage of disease [Klu13], evaluate therapeutic measures [Pul18] or predict the risk of falling [Del17]. To achieve intra- as well as inter-subject comparability when investigating gait in the clinic, standardized gait tests are a commonly used auxiliary tool. They can either be evaluated by simple measures, such as the required time for a certain task [Ste08, Blo16], or by utilizing IMUs for objective gait parameter estimation [Dew14, Dib17, For11, Gha19, Gre10, Sal10, Smi16, Zam11] (section 1.1). Additionally, due to their portability and mobility, IMUs enable long-term monitoring of patients in their home environment and therefore provide a comprehensive impression of the patient's condition instead of a momentary insight as it would be the case during a short clinical examination [Del16, Zam11]. The feasibility and benefits of continuous home monitoring, as well as the evaluation and interpretation of the large amounts of data generated, are currently highly frequented research topics [Bro16, Bro17, Del16, Del17, Gal19, Hil19, Wei14, Zam11] (section 1.2).

1.1 Standardized Gait Tests

There are various established gait tests for assessing PD symptoms focusing on different aspects, e.g. balance, mobility, turning behavior, or endurance [Blo16, Ste08].
One commonly used test is the Timed Up and Go Test (TUG). It was applied in several studies to investigate balance, mobility, and the effects of environmental interventions at the same time [Dew14, For11, Sal10, Smi16, Zam11]. For the TUG, the subject is sitting on a chair, standing up and walking a straight walkway of defined length in comfortable speed, turning around at the end of the walkway, walking back to the chair, and sitting down again. The TUG without auxiliary tools is evaluated by measuring the time a subject needed for completion [Blo16, For11, Ste08]. The test-retest reliability of a gait test was estimated using the intra-class correlation coefficient (ICC) ranging from 0 to 1 for no test-retest reliability and perfect test-retest reliability, respectively.

For the TUG, ICC was reported to range from 0.70 to 0.99 [Smi16, Ste08]. There are indications that the TUG duration results enable discrimination of different disease stages [Dew14], and fallers and non-fallers [For11].
Salarian et al. developed and patented the so-called instrumented Timed Up and Go Test (iTUG) (Appendix A.1), including five inertial sensors and an algorithm for automatic detection, separation, and analysis of different components of the TUG, such as sit-to-stand transitions, turning, and walking [Sal10]. Discriminative features for PD patients and healthy controls were found to be stride velocity and stride length with an average ICC larger than 0.75, as well as cadence, peak arm swing velocity, and turning velocity with an average ICC larger than 0.90 [Sal10]. Using iTUG, compared to traditional TUG, reliability increased especially for the discrimination of early-to-moderate stage of disease [Sal10]. Dibilio et al. utilized the iTUG output to differentiate between patients in "on"- and "off"-state [Dib17]. Greene et al. used and patented a similar sensor-based TUG analysis system for fall risk prediction (Appendix A.3), achieving a sensitivity of 77.3% and a specificity of 75.9% [Gre10].
Furthermore, to assess a patients ability to vary gait velocity, the 10 meter walking test (10MWT) can be used. In this test, the patient is covering a 10-meter long straight walkway. The 10 meter walking test (10MWT) can be repeated in different self-selected speeds [Blo16]. Using a stopwatch, the required time for each walkway is measured and the velocity is calculated. The ICC is ranging from 0.75 to 0.98 [Blo16, Mil13, Ste08]. Minimal detectable change in gait speed, i.e. the minimal amount of change which is not caused by measurement variation [Ste08], was reported between 0.09 $\frac{m}{s}$ and 0.13 $\frac{m}{s}$ [Mil13]. When additionally using foot-mounted inertial sensors for gait parameter estimation during 10MWT, PD can be detected in an early stage [Bar11]. Furthermore, the stage of motor impairment can be objectively assessed [Klu13].
However, short walkways like a 10-meter path do not reflect real-life conditions optimally. The attention of a patient is higher for very short intermittent walks than for continuous walking, which has shown to negatively impact the predictive power of gait parameter based classification models [Ur 19]. Galna et al. investigated the influence of the amount of recorded steps to reliably determine the gait variability, which was calculated as the standard deviation of stride velocity, length and width, and stance and swing duration [Gal13]. The biggest improvement of accuracy was observed across the first 30 strides. Consequently, Galna et al. recommend gait test protocols that include at least 30 strides [Gal13].
Schlachetzki et al. extended the 10 meter walking test (10MWT) to the 4x10 meter walking test (4x10MWT) and demonstrated its clinical relevance. The 4x10MWT includes walking along a straight walkway with the length of 10-meter and performing a 180° turn afterwards followed by

returning the walkway. This procedure is then repeated. They showed that parameters obtained from inertial sensor recordings during 4x10MWTs objectively track gait changes associated with PD [Sch17].
Another important aspect besides the analysis of continuous walking is the observation of turnings. PD patients tend to perform turnings slower and with an increased amount of steps [EG14]. Ghassemi et al. analyzed turning-related parameters, such as the number of strides per turn and turning duration, additionally to gait parameters [Gha19]. They used the 4x10MWT to investigate both continuous walking at different speeds and turning behavior. A correlation of turning parameters with different disease stages and levels of motor impairment was reported [Gha19]. 4x10MWTs do not only contain straight strides, but also deceleration, acceleration and turning segments. These distinct clusters (constant, non-constant and turning) were automatically detected using a machine learning based algorithm by Nguyen et al. [Ngu19]. Their study revealed that non-constant gait segments may provide additional information concerning clinically relevant PD-related motor symptoms. An algorithm for detecting 4x10MWTs within inertial sensor recordings recorded in the clinic automatically instead of requiring manual labeling, which serves as a basis for this thesis, was developed by Fischer et al. [Fis20].

1.2 Free-living Gait Analysis

The availability of inexpensive, unobtrusive wearable devices opens the possibility for personalized long-term monitoring. Although there is still a lack of algorithms and evaluation methods to analyze the data in a way that makes it applicable for widespread clinical application [Del16], ongoing research demonstrates the potential of mobile gait analysis [Bro16, Bro17, Del16, Del17, Gal19, Hil19, Wei14, Zam11]. Free-living gait data was shown to provide a more comprehensive insight into the patient's condition than clinical rating scales like the UPDRS [Del16, Gal19] or sensor based clinical gait parameter recordings [Bro16, Gal19].
Galperin et al. reported that free-living gait assessment reveals additional information, as the variance in free-living gait parameters is not completely predictable using the UPDRS score and laboratory-recorded parameters [Gal19]. Brodie et al. noticed a significant difference between mean cadence and step time variability recorded at home and in the clinic. Even though the maximum values of those two parameters were comparable for both clinical and free-living recordings, mean cadence was lower at home, whereas mean step time variability was higher [Bro16]. This outcome suggested that clinical measurements only provide a snapshot of an individual's best performance rather than a comprehensive insight [Bro16].

To prove the necessity of home monitoring, Hillel et al. tried to simulate free-living conditions in the clinic. They compared gait parameters during dual-task walking at the clinic with the ones recorded during walking a straight walkway at home [Hil19]. Even though gait speed, step regularity, and stride regularity worsened significantly compared to single-task walking and tended toward the free-living parameters, there was no reliable correlation. They concluded that clinical measurements, even if adapted more toward real-life conditions, do not reflect real-life recordings sufficiently [Hil19].

Another example directly comparing clinical and home recordings is a study by Zampieri et al. [Zam11]. They investigated the differences between two iTUGs performed within a time span of 24 hours, once in the clinic and once in the patient's home environment. A tendency for PD patients to perform worse in their home environment compared to the clinic was observed. A significant correlation between location and measured gait parameter was found for the walking speed. Participants suffering from PD walked slower at home, whereas healthy controls performed similarly at home and the clinic. Reasons were discussed to be that on one hand, people feel more comfortable at home causing the dopamine level to be lower than in the clinic, where the dopamine level might increase due to anxiety and the ambition to impress the physician. On the other hand, the home environment might be more cluttered and distracting than the clinic triggering symptoms like freezing of gait (FOG) [Zam11].

The additional information gained from free-living data enables monitoring of falls and fall risk assessment. Hausdorff et al. invited an inertial sensor-based system for detection and tracking of falls and near-fall events at home (Appendix A.4). Furthermore, various studies [Bro17, Del17, Wei14] tried to distinguish between fallers and non-fallers with the main aim to predict a patient's probability of falling. Brodie et al. investigated a correlation between a subject's overall quality of walking, measured by the total amount of steps and the mean cadence, and their tendency to fall [Bro17]. Del Din et al. conducted a study with participants with and without PD and reported that fallers suffering from PD showed a significantly higher variance in step length than non-fallers. Fallers walked with shorter bouts and slower, shorter and less variable strides than non-fallers [Del17]. All together, free-living measurements outperform traditional clinical scales in fall prediction accuracy [Wei14].

1.3 Purpose of the thesis

A persisting problem of most home-monitoring studies is that they are still controlled by protocols with detailed specifications. These protocols differ considerably between individual studies [Del16]. In order to enable a consistent evaluation and achieve comparability and reproducibility of results across studies, usage of standardized protocols and validation procedures is recommended [Del16, Zam11]. This can be implemented in practice by integrating standardized gait tests, as established for clinical applications, into free-living study protocols.
An interesting aspect now is the question how sensor-based gait tests can be integrated into patients' everyday life over a longer period of time and to what extent they provide a usable outcome compared to fully unsupervised free-living recordings. So far, no completed study has been reported in literature combining completely unsupervised free-living recordings and standardized gait tests. Nevertheless, a current project is a joint study (*FallRiskPD*) of the Machine Learning and Data Analytics Lab (Friedrich-Alexander-Universität Erlangen-Nürnberg) and the University Hospital Erlangen. The goal is to investigate machine learning algorithms for the prediction of fall risk from long-term IMU gait data. For a period of two weeks, patients were instructed to wear three inertial sensors, attached to both feet and the lower back, during their daily activities. Additionally, they were asked to perform standardized gait tests, namely the TUG, and the 4x10MWT, where the latter was executed in different self-chosen velocities, in the patients' home environment every day.
To enable separate analyses of free-living gait and standardized tests, the gait test recordings need to be extracted from the long-term data, i.e. the corresponding timestamps are required. This can for instance be accomplished with a smartphone application used by the patient to track start and stop time of the gait tests. However, especially elderly, motor-impaired people experience difficulties handling mobile touch devices, which causes imprecise, incorrect or missing labels [Naf14, Har13]. As a consequence, before the automatic, algorithm-based determination of gait parameters can be executed, manual review and refinement of all data are necessary. Thus, to simplify the handling of mobile gait analysis systems and enable a feasible integration into the everyday life of patients, it is desirable to automatically recognize standardized gait tests within long-term data recordings.

The purpose of this thesis is therefore to develop an algorithm for automatic detection of standardized gait tests from free-living inertial sensor data. The focus was set on the detection of 4x10MWT series and subsequent segmentation into the different self-chosen speed levels. Therefore, the previous approach of Fischer et al. [Fis20] is taken as a basis, extended for the applicability to home monitoring data as recorded in the *FallRiskPD* study, and validated against the expert-labeling of the respective data. The questions to be addressed are:

- Is it possible to detect series of 4x10MWTs within free-living inertial sensor data?
- Is it possible to differentiate separate 4x10MWTs and assign the corresponding gait speed levels?

A gait test detection algorithm contributes to the automatic processing of free-living data and enables better comparability across studies based on gait parameters determined from standardized tests.

Chapter 2

Fundamentals

In the following, the theoretical background of this thesis is explained. First, a basic understanding of human gait is given (section 2.1). Then, medical aspects about PD are outlined (section 2.2). Finally, the basic working principle of IMUs utilized for sensor-based gait analysis is clarified (section 2.3).

2.1 Human Gait

Human gait is a complex series of movements, that requires an interplay of the nervous system and various muscle groups of trunk and extremities [Pir16]. The human gait pattern consists of a periodically repeated sequence of certain gait events. The time span in which both limbs are performing all these events is called a gait cycle (Figure 2.1). A gait cycle can be subdivided in different phases such as *swing* and *stance phase*, or *single* and *double support phase*. It is defined as the interval between two consecutive repetitions of a gait event.
Usually, the initial contact (IC) of one foot with the ground is chosen as the beginning and end of a gait cycle [Whi07]. It describes the point in time when the foot initially strikes the ground. For a healthy person's gait, the heel is the first part of the foot that is placed to the ground. Therefore, IC is also referred to as heel strike. For pathological, shuffling gait this might not be the case [Tun17], thus IC is a more precise term.
With IC, the *stance phase* starts, which describes the part of the gait cycle, where the respective foot is in contact to the ground. The body weight is then continuously shifted until it entirely remains on this foot, allowing the opposite foot to take off from the ground. At this point, the complete foot is resting on the ground, which is defined as mid stance (MS). After IC of the opposite foot, body weight is shifted back to the opposite side. The released foot is then lifted

beginning with the heel taking off the ground. When the last part of the foot that is still touching the ground is raised, the *swing phase* starts. This gait event is described as terminal contact (TC), foot-off, or toe-off. After TC, by flexion of hip and knee, the leg is raised. At the same time, leg and trunk muscles are activated to control the body weight transition [Pir16]. The *swing phase* and thus one gait cycle end with the next IC.
Simultaneously, the opposite foot is performing the same cycle of gait events with a displacement of half a cycle [Whi07]. Hence, the *stance phases* of both feet are overlapping for a short period of time. Therefore a differentiation between a *double support phase*, where both feet touch the ground, and a *single support phase*, where only one foot has contact to the ground while the opposite one is swinging, can be made. For healthy gait, the ratio of double support time and single support time during one gait cycle is 20% to 80%, respectively [Tun17]. The stance phase, however, constitutes about 60% of the gait cycle, whereas the swing phase takes up 40% [Pir16].

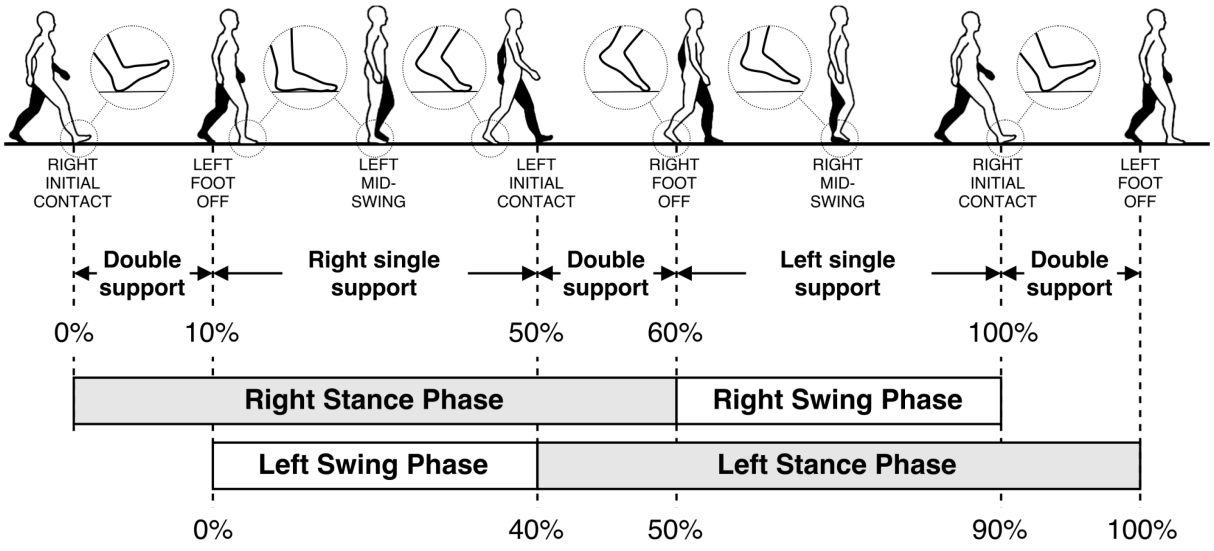

Figure 2.1: **Schematic Overview of Phases and Events of the Human Gait Cycle.** [Tun17]

2.2 Parkinson's Disease

Parkinson's disease (PD) is a progressive neurodegenerative disorder causing a variety of motor and non-motor symptoms (section 2.2.1). Although PD was first described by James Parkinson in *"An essay on the shaking palsy"* in 1817 [Par02], it took more than 100 years to determine that the disease is linked with a loss of cells in the substantia nigra [Jan08]. In 1960, it was further

discovered that PD patients show a decreased concentration of the neurotransmitter dopamine in the brain. Hence, the symptoms of PD can be drug-treated with levodopa, a precursor for the biosynthesis of dopamine [Hor06]. In contrast to dopamine, levodopa is able to cross the blood-brain barrier and can therefore directly impact the affected areas of the brain [Har80]. Nevertheless, medical treatment can suppress the symptoms and slow down the disease progression, but not revert or stop it. Therefore, in order to guide therapeutic decisions and achieve the best possible quality of life for patients, it is important to be aware of the manifold symptoms of PD [Jan08], assess its severity [Klu13], and continuously monitor its progression [Sch17].

2.2.1 Cardinal Symptoms

The cardinal motor-related signs of PD are *tremor*, *rigidity*, *bradykinesia*, and *postural instability* [Wal06, Jan08].

Tremors occur permanently at a frequency between $4\,\mathrm{Hz}$ and $6\,\mathrm{Hz}$. Mostly, extremities are affected asymmetrically during disease onset, worsening with increasing disease severity [Sam04]. Additionally, a tremor at rest that disappears when performing any actions, can occur for parts of the face or the legs [Jan08].

Rigidity refers to the stiffness of limbs during passive joint movements as a consequence of permanent muscle contraction [Sam04]. It is also often associated with pain [Jan08].

Bradykinesia describes the slowness of movement that is characteristic for all basal ganglia diseases. Patients experience difficulties in the context of movement planning and execution and while performing multi-tasking actions [Jan08] or fine motor tasks [Sam04].

Postural instability includes a loss of postural reflexes resulting in poor balance and is mostly developing in an advanced stage of disease. When pulled backwards, a patient with *postural instability* needs several steps to regain balance. This leads to an increased fall risk [Sam04]. Another characteristic phenomenon accompanying *postural instability* is freezing of gait (FOG), which refers to the sudden inability to move. Typical situations when FOG occurs are when walking is initiated, during walking when passing an obstacle or a narrow passage, or during turning [Jan08]. FOG does not occur universally, but for around 50% of all PD patients [Mac07]. At an advanced disease stage and accompanied by long-term drug treatment, an additional problem that arises are motor fluctuations. Patients suffering from motor fluctuations are experiencing two unpredictably alternating phases, the so-called "on"-state and "off"-state. During "on"-state, symptoms are reduced as the patient is responding well to the medication, whereas during "off"-state, the patient has severe, uncontrollable symptoms [Wal06].

Besides the visible motor limitations associated with PD, there is a wide range of additional impairments to be mentioned. Examples are dysfunctions of the autonomic nervous system, such as orthostatic hypotension, cognitive disorders, such as dementia or depression, sleep disorders, and sensory abnormalities, such as loss of olfactory functions [Jan08, Sam04].

2.2.2 Impact on Gait

Resulting from the motor-related clinical features explained above, PD patients show a characteristic gait pattern. When observed visually, it includes slow, shuffling gait with short strides, reduced arm swing during walking, and sudden hesitation especially after turnings due to FOG [Sam04].
This gait behavior can also be described objectively using spatiotemporal gait parameters. Shuffling gait is determined by a decreased foot clearance as well as lower IC and TC angles. With advanced stage of disease, these parameters decrease further [Sch17]. When analyzing turning behavior, significant differences to healthy gait can be observed in terms of turning duration and required number of steps [Man15].
Severe *postural instability* leads to increased swing time variation [Sch17]. When FOG is present, the variability of all gait parameters, such as the inter-stride variability of stride time and length, is increased [Hau03].

2.3 Inertial Measurement Units

Inertial sensors are the most common wearable device for tracking motion and orientation parameters. Because of their unobtrusive and lightweight character combined with the capability to perform precise measurements, they are used in various fields of technology, such as inertial navigation [Woo07], smartphone applications, or human movement analysis [Ben15]. The built-in components vary depending on the intended applications. The most important parts for human gait analysis are accelerometer (section 2.3.1) and gyroscope (section 2.3.2).

2.3.1 Accelerometer

According to Newton's Second Law of Motion (Equation 2.1), the acceleration $\boldsymbol{a}$ of an object is determined by its mass m an the force $\boldsymbol{F}$ applied to it.

$$\boldsymbol{F} = m \cdot \boldsymbol{a} \tag{2.1}$$

When the acceleration is known, integration yields the speed and double integration the displacement of an object and therefore important information concerning its movement.
Modern accelerometers are realized as micro-electro-mechanical systems (MEMS), which means they include an electrical component transforming the mechanical into an electrical signal. Compared to mechanical systems, MEMS bring several advantages.They are sensitive also for small accelerations, have low power consumption, and are insensitive to temperature fluctuations. [Saz14]
There are different working principles existing, most suitable for body motion measurement is the so-called beam-type accelerometer. It consists of an elastic beam that is fixed on one side and has a seismic mass attached to the other end (Figure 2.2).

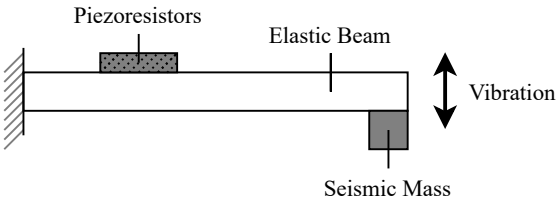

Figure 2.2: **Schematic construction of a beam-type accelerometer.** Adapted from Albarbar et al. [Alb09].

Following Newton's Second Law, when the object is moving, a force is applied to the seismic mass of the beam proportional to the acceleration. To avoid resonance after deflection of the beam, a mechanical damping is included in the system. The applied force is then determined using an electrical component, such as a piezoresistive element, that is changing its resistance as a consequence of the deflection and therefore causing a polarization voltage change proportional to the displacement. [Saz14]
Alternatively, piezoelectric, capacitive, or extensimetric components are commonly used [Bro19]. For a comprehensive investigation of movement, all directions of space need to be considered. Therefore, three of the described uniaxial accelerometers are commonly combined orthogonal into a single 3-d-accelerometer [Bro19].

2.3.2 Gyroscope

Gyroscopes are used to determine the angular velocity of a system. A traditional mechanical gyroscope can be conceived of as a system consisting of a spinning wheel attached to perpendicular gimbals. When the system is inclined, the wheel's orientation remains unchanged whereas the gimbals tilt around the corresponding angle. Thus, by determining the inclination angles of all gimbals relatively to one another, the orientation change caused by an angular acceleration can be estimated [Woo07].

Nowadays, MEMS-based gyroscopes are commonly used, as they are more robust and more cost-effective in the manufacturing process [Woo07]. They are measuring Coriolis forces, that are generated within a rotating reference system with angular velocity $\boldsymbol{\omega}$. If a mass m is moving within the reference system with constant velocity $\boldsymbol{v}$, a Coriolis force $\boldsymbol{F}_c$ is acting on it (Equation 2.2), resulting in a deflection in vertical direction. Combining this definition with Newton's Second Law, it can be found that the acceleration of the mass is related to the angular velocity (Equation 2.3).

$$\boldsymbol{F}_c = -2m(\boldsymbol{v} \times \boldsymbol{\omega}) \tag{2.2}$$

$$\boldsymbol{a}_c = -2(\boldsymbol{v} \times \boldsymbol{\omega}) \tag{2.3}$$

As a consequence, accelerometers as described above can be utilized for angular velocity estimation. For Coriolis forces to arise, they need to be constantly in motion. Therefore, the seismic mass is electronically forced to permanent vibration with a velocity $\boldsymbol{v}$ in the direction of the MEMS surface. If a turning movement is then applied, a deflection in the perpendicular direction can be recorded [Saz14] (Figure 2.3). This principle is again applied for all dimensions to capture all rotations in 3-d-space.

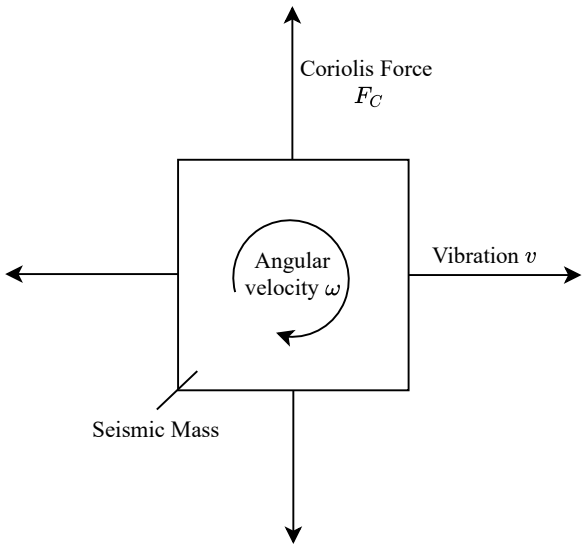

Figure 2.3: **Working Principle of a Gyroscope.** When the seismic mass of an accelerometer is vibrating along the indicated direction, and an angular velocity $\boldsymbol{\omega}$ is applied, the seismic mass is accelerated in perpendicular direction by the resulting Coriolis force. Adapted from Woodman [Woo07].

2.3.3 Coordinate Frames

Since the signal of both gyroscope and accelerometer consists of data recorded from three different perpendicular axes, it can be represented as a vector. The underlying reference axes and therefore the coordinate frame for the signal components can be chosen in different ways.
The unprocessed recordings gathered by IMUs describe the world from the sensor's view. That means, they refer to the coordinate system defined by the sensor, i.e. the *sensor frame*. When measuring human gait signal, the inertial sensors are attached to and therefore moving with the human body. Thus, the *sensor frame* is constantly moving. To achieve a physically meaningful outcome, for different use cases different coordinate system transformations can be applied to the *sensor frame* signal. Essentially, there are two distinct reference coordinate systems the inertial sensor signal can be converted to.
On one hand, there is the fixed *world frame*, which is defined relative to the space the human is moving in. This frame is required to calculate global displacements during walking. To map the *sensor frame* values of the body-mounted IMUs to the *world frame*, a trajectory reconstruction is required.
On the other hand, there is the *body frame* using the human body as a reference system. If the position of the sensor relative to the body is known, a rotation matrix can be set up for the conversion from *sensor frame* to *body frame*. This transformation can be applied to the signal of a hip mounted sensor for instance. However, an exception must be made for sensors mounted to opposite-sided extremities such as left and right foot. Since the same movement is desired to cause the same signal independent of the observed extremity, the axis crossing the sagittal plane needs to be mirrored. The resulting signal consists of an axis in medial to lateral (ml), posterior to anterior (pa), and superior to inferior (si) direction (Equation 2.4). This convention causes the local *body frame* of one of the mirrored sensor positions to loose its right-handedness. Therefore, some physical calculations can not be performed using the *body frame* coordinate system, especially if they involve the use of several coordinate axes.

$$gyr = \begin{pmatrix} \boldsymbol{g}_{pa} \\ \boldsymbol{g}_{ml} \\ \boldsymbol{g}_{si} \end{pmatrix}; \quad acc = \begin{pmatrix} \boldsymbol{a}_{pa} \\ \boldsymbol{a}_{ml} \\ \boldsymbol{a}_{si} \end{pmatrix} \tag{2.4}$$

Chapter 3

Data Acquisition

In this chapter, the *FallRiskPD* study desgin is described (section 3.1). More detailed, the standardized gait tests performed in this study, which are particularly relevant for this thesis, are emphasized (section 3.2). Furthermore, basic information about the technology used for data collection (section 3.3) are outlined. Additionally, an overview of the participants is given and characteristics of the underlying dataset for development and evaluation of the proposed gait test detection pipeline are explained (section 3.4).

3.1 Study Design

The aim of the *FallRiskPD* project is to determine a PD patient's individual risk of falling before the actual fall happens. For this purpose, large and continuous gait recordings are required. They were recorded using *Mobile GaitLab* (Portabiles HealthCare Technologies GmbH, Erlangen, Germany), which is a wearable sensor system for gait analysis. It can be integrated into a patient's everyday life as a long-term home monitoring system. Using the obtained data, the objective is to extract specific gait patterns and allow a prediction of a patient's fall risk by using machine learning algorithms.

The *FallRiskPD* project consists of two study phases. The first phase was implemented as a retrospective, multi-center study to record gait data of several PD patients. The patients wore the sensor systems continuously from 8:00 a.m. until early evening over a period of approximately two weeks. After putting the sensors off, they were placed at their charging station, where the recorded data was downloaded from the internal sensor storage to an accompanying smartphone. All full-day recordings like this are referred to as a session in the following.

Three months after this first recording phase, the patient's frequency of falls was determined again in a follow-up examination. Based on this data, crucial features for fall prediction are determined, extracted, and used for development and training of an algorithm for fall risk estimation. In the second phase of the project, this algorithm is validated and tested comprehensively from a clinical and technical viewpoint within a prospective, multi-center study.

3.2 Standardized Gait Tests

In addition to the recording of completely unsupervised free-living data, gait parameters were also acquired during semi-standardized gait tests, similar to those performed during clinical examinations. For that purpose, the subjects were instructed to perform multiple defined gait tests three times a day during the recording period in the morning around 8:00 a.m, at noon between 11:00 a.m. and 12:00 a.m., and in the afternoon around 4:00 p.m., respectively. These tests included three 4x10MWTs performed subsequently with short resting periods in between. Every test was intended to be executed in a different, self-selected speed. One 4x10MWT should thereby be performed at a speed that is comfortable for the patient, one deliberately slow and one comparably fast. Although in the laboratory patients were always instructed to execute the tests in this order, for the free-living recordings there was no particular order specified. As it was difficult to maintain a distance of exactly $10\,\mathrm{m}$ due to different spatial conditions depending on the subject's current environment, the patients were instructed to choose a straight path available at the time of test execution, whose length came closest to 10 meters, as their walkway. After the 4x10MWTs, a TUG test was performed. For this test, the patients needed to sit down on a chair, get up, walk around approximately three meters straight, turn around, walk back, and sit down again.
For tracking the beginning and end times of the daily tests, a smartphone application was provided (Figure 3.1). Additionally, the test persons had to answer questions about their subjective motor condition and eventual fall events eyery day with the help of an electronic diary which was also included in the smarthphone application.

3.3 Sensors

All observed sensor measurements used in this thesis were performed with the *Mobile GaitLab* system including three *NilsPod* (Portabiles GmbH, Erlangen, Germany) inertial sensor units. The *NilsPod* sensor consists of a 3-d accelerometer (range $\pm 16\,\mathrm{g}$), a 3-d gyroscope (range $\pm 2000\,^\circ/\mathrm{s}$), and features wireless synchronization [Rot18]. Data was recorded at a sampling rate of $102.4\,\mathrm{Hz}$.

The recordings were automatically transferred and stored to the accompanying smartphone when the sensor was placed on the charging station. Two sensors were mounted to the shoes, where all patients wore the same orthopedic shoe model (Figure 3.2). The third sensor was attached to the lower back with a hip belt.

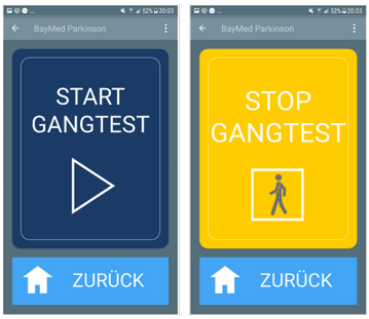

Figure 3.1: **GUI of the Gait Test Annotation App.**

Figure 3.2: **Shoe Sensor Placement.** The IMU sensors were attached to both feet by mounting them to the shoe as shown.

3.4 Dataset

3.4.1 Participants

For development, training and testing of the proposed gait test detection pipeline, a dataset consisting of 12 participants with a mean age of 64.3 years and a mean H&Y disease stage of 2.6 (Table 3.1) diagnosed with idiopathic PD was used. The participants were recruited by the University Hospital Erlangen, the Hospital Rummelsberg, and the Ernst von Bergmann Hospital Potsdam. All data was recorded in the participant's home environment during the first phase of the *FallRiskPD* study.

Age [years] ± SD	64.3 ± 8.4
Height [cm] ± SD	171.9 ± 9.4
Weight [kg] ± SD	76.7 ± 15.2
UPDRS-III ± SD	17.8 ± 6.2
H&Y ± SD	2.6 ± 0.4
Disease Duration [years] ± SD	8.2 ± 4.3
Sex (m/f)	9 / 3

Table 3.1: **Subject Characteristics.**

3.4.2 Inclusion Criteria for Recordings

In total, for all observed patients, 164 sessions were recorded. As a prerequisite for declaring a recorded session as valid, in this thesis, three conditions were defined to simplify and standardize the evaluation:

1. For all three sensors, recordings were available.
2. The recorded data of all three sensors was synchronized.
3. Smartphone annotations of performed gait tests were available.

Three sessions were excluded because of missing sensor recordings, eight sessions were excluded because of missing smarthphone annotations, resulting in 153 valid full day recordings with a length of $9.77\,\mathrm{h} \pm 3.66\,\mathrm{h}$.

3.4.3 Manual Labels

To achieve a precise ground truth for evaluating the algorithm, all 4x10MWTs in these sessions were labeled manually using the timestamps set with the smartphone application as benchmarks. That means, an area of $\pm 1000\,\mathrm{s}$ around the beginning and end timestamp was investigated to prevent 4x10MWTs from being missed because of smartphone annotations that were accidentally set too early or too late. Every 4x10MWT in this area was labeled and a corresponding speed level was assigned. Each label was set in such a manner that the longest possible resting periods before the first and after the last stride of a test were included.
Consecutive 4x10MWTs within an interval shorter than $15\,\mathrm{s}$ were expected to occur in sets of three according to the *FallRiskPD* study protocol (section 3.1), one in preferred, one in slow, and one in fast speed. These sets are referred to as *gait test series* in the following. A differentiation

was made between *gait test series* performed in compliance with the protocol and *gait test series* that did not follow the protocol exactly. Examples for non-compliant *gait test series* are such, which included too many or too few repetitions of the straight walkway resulting in 6x10MWTs or 2x10MWTs for instance. Furthermore interpreted as non-compliant were such, which included too many or too few repetitions of 4x10MWTs within one *gait test series* (Figure B.1). 1404 4x10MWTs were labeled in the investigated valid sessions, forming 418 *gait test series*. 298 (71.3%) of these *gait test series* were compliant with the given protocol (Table B.1).

Chapter 4

Methods

Within this chapter, methods for detection and decomposition of *gait test series* based on free-living gait data recordings are presented. A previously introduced pipeline by Fischer et al. [Fis20] served as a basis for this. Refinements and changes that were required for the application to free-living data are going to be outlined and described in detail. For finding *gait test series*, some general assumptions regarding their ideal appearance were made:

(i) A *gait test series* consists of three succeeding 4x10MWTs of different velocities.

(ii) Within every 4x10MWT, a $180°$ turn is performed after each 10-meter walk.

(iii) A $180°$ turn was performed between every 4x10MWT to return the same way, resulting in a total of twelve turns.

(iv) The 4x10MWTs can be categorized by the patient's self-selected gait speed into slow, preferred, and fast.

(v) Resting periods can occur between succeeding 4x10MWTs, but are shorter than $15\,\mathrm{s}$.

These presumptions were oriented towards the *FallRiskPD* study protocol (section 3.2) with the aim to cover as many recorded 4x10MWTs as possible. To find the points in time of 4x10MWTs in full-day recordings, the gait recordings were undertaken three major processing steps (Figure 4.1).

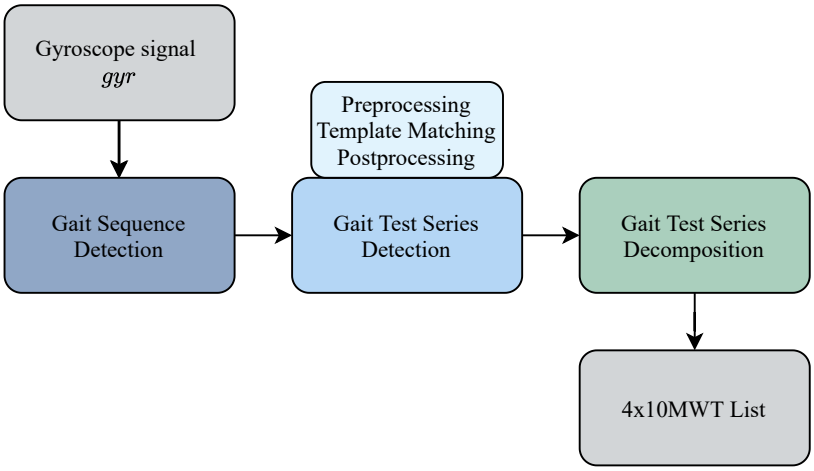

Figure 4.1: **Pipeline Overview.** Three main processing steps were required to segment*4x10MWT*s from free-living movement recordings and achieve a test list including timestamps and speed labels.

As a preprocessing step, a gait sequence detection was performed (section 4.1). Then, the first research goal was adressed by investigating the discovered gait sequences for the characteristic pattern that is created by periodically occurring turns in the *gait test series* (section 4.2). Therefore, template matching was applied (section 4.2.2) after signal preprocessing for pattern enhancement (section 4.2.1). Afterwards, matches were filtered by postprocessing (section 4.2.3). As a result, a list of detected *gait test series* was received for every session. Within the last step, the focus was set on answering the second research question by decomposition of the detected *gait test series* into separate 4x10MWTs, and assignment to their corresponding speed levels (section 4.3).

4.1 Gait Sequence Detection

Within home-monitoring data, many other activities of daily living are recorded next to walking, such as sitting, standing, driving or cycling. This results in a large data volume. To increase the pipeline efficiency by only considering relevant parts of the recorded free-living data containing gait signal, a gait sequence detection was applied first.
For this purpose, an algorithm proposed by Ullrich et al. was used [Ull20]. It is based on the gyroscope signal gyr in *body frame* coordinates. The main part of the gait sequence detection consists of two steps: First of all, signal parts containing general activity were detected. Therefore, sliding windows were utilized covering $10s$ of the gyroscope signal. Succeeding windows had a 50% overlap each. Offset was corrected by calculating the mean of gyr and subtracting it.

Afterwards, the gyroscope signal norm was computed for each window sample-wise (Equation 4.1). The average window norm was then compared to a global activity threshold of $50°/s$. Windows with an underlying norm were rejected.

$$|gyr[i]| = \sqrt{g_{pa}[i]^2 + g_{ml}[i]^2 + g_{si}[i]^2} \tag{4.1}$$

The second step investigates the detected active areas using frequency analysis of the g_{ml} signal to distinguish between gait and non-gait movements. The algorithm exploits the phenomenon that acceleration and angular velocity of the feet during walking are mainly composed of a main frequency and its harmonics [Mac05]. This is therefore also reflected in the IMU signal. Consequently, when the frequency spectrum of the IMU recordings was analyzed, peaks occured along multiples of the fundamental frequency of human gait, which was investigated to be around $1\,\mathrm{Hz}$ [Mac05]. The dominant frequency was estimated using cross correlation. Then, a peak detection was performed on this signal to detect the most prevalent frequencies.
If the dominant frequency was equivalent to the fundamental frequency of gait, and at least two peaks were detected around its harmonics, a window was classified as gait signal. Finally, all detected overlapping gait windows were concatenated to gait sequences.
As this pipeline was applied to synchronized sensor data only, it can be assumed that gait sequences are identical for all sensors. Therefore, to simplify the following processing steps all intervals detected separately from right and left foot sensor are aligned in an additive manner. That means, each sample that has been classified as part of a gait sequence within at least one window is added to a global gait sequence list for all sensors.
As stated in assumption (v), it was expected that short resting periods can occur between single 4x10MWTs. To perform template matching independently for each of the detected gait sequences, every *gait test series* was required to be completely included within a gait sequence. Therefore, in this particular application short resting periods were desired not to be rejected. Hence, as a last step of the gait sequence detection, every pair of adjacent gait sequences was concatenated under the constraint that the beginning of the succeeding interval was not more than $15\,\mathrm{s}$ apart from the end of the preceding interval (Figure 4.2).

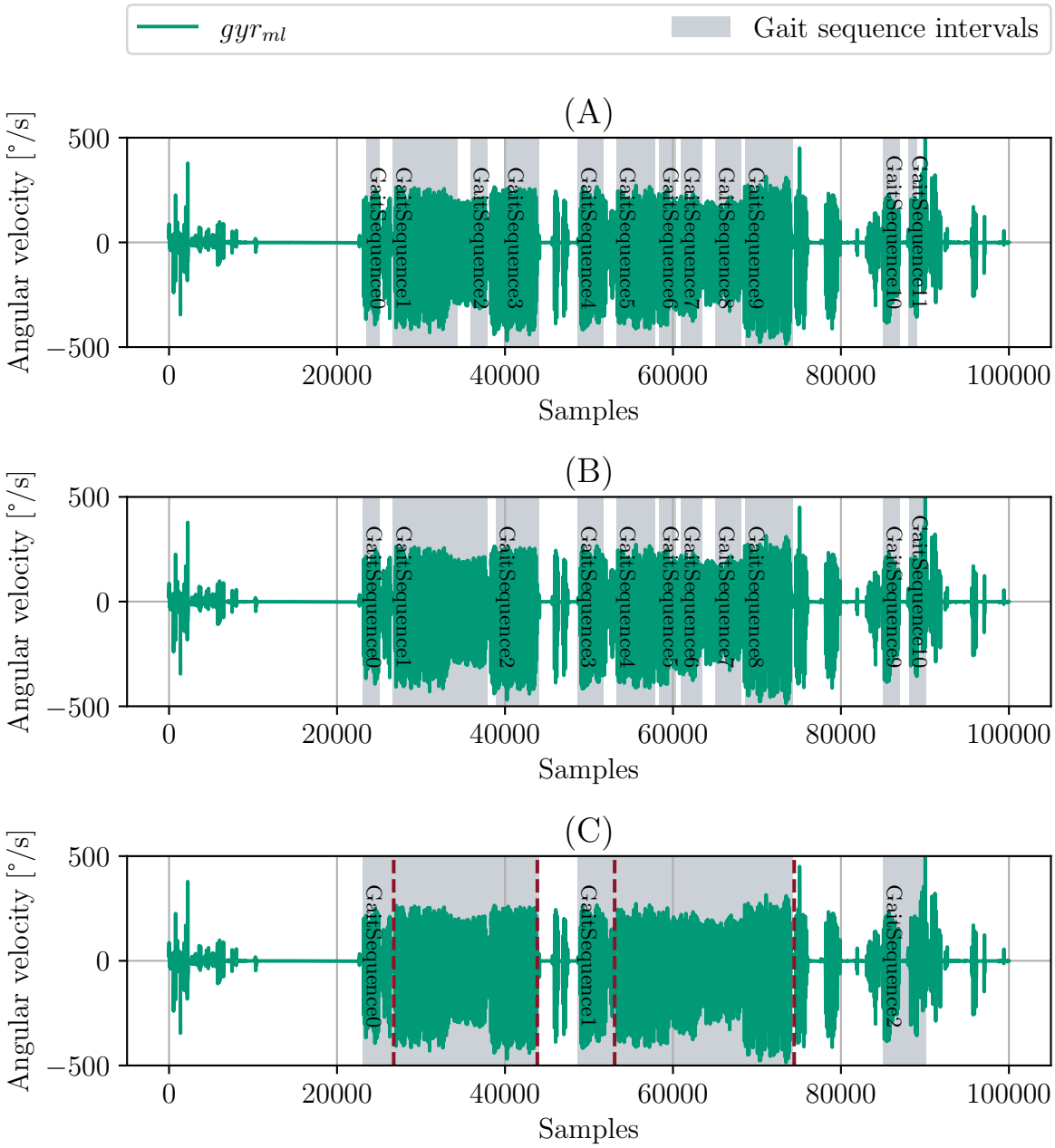

Figure 4.2: **Gait Sequence Detection.** *(A)* shows the labels resulting from the gait sequence detection algorithm [Ull20]. The displayed signal is an excerpt from a free-living dataset, recorded by the left foot mounted sensor. In *(B)*, the gait sequences are shown after combining the results of both foot sensors. In *(C)*, the final gait sequences which were created after concatenating labels with short pauses in between can been seen. After this step it can also be observed that the included *gait test series*, whose beginning and end were highlighted with red lines in *(C)*, are completely covered by one gait sequence each. In the following, only the recordings within these sequences will be investigated further.

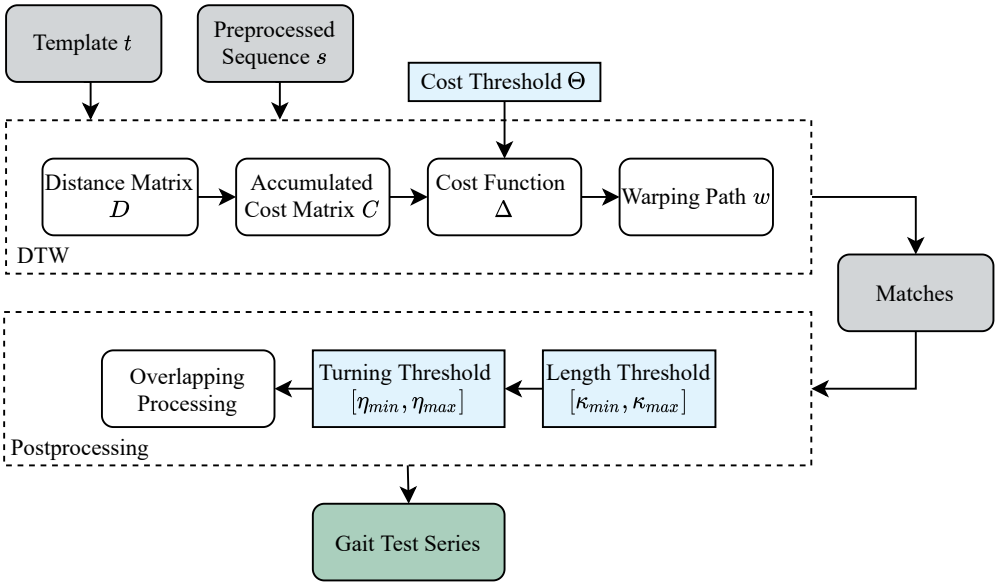

Figure 4.3: **Flowchart for the proposed Gait Test Series Detection Pipeline starting from the Preprocessed Signal.**

4.2 Gait Test Series Detection

Considering 4x10MWTs, important characteristic features differentiating them from real life movements were found to be the evenness and regularity of turnings followed by gait sequences with uniform lengths. Since the length of the walkway can not be assumed to be exactly $10\,\mathrm{m}$ within the home environment of a patient, the turns were chosen as the decisive criterion.
Thus, the resulting angular velocities of interest were recorded by $\boldsymbol{g}_{si}$. Consequently, the $\boldsymbol{g}_{si}$ axis was extracted from the raw IMU signal. This signal was preprocessed to enable *gait test series* detection using template matching. Finally, matches were refined within a postprocessing step (Figure 4.3).

4.2.1 Preprocessing

To enhance the turnings of interest and eliminate noise as well as slight shuffling from the IMU signal, a preprocessing pipeline was introduced by Fischer et al. before the actual template matching step [Fis20]. In the following, an adapted version of that is briefly explained and modifications, including normalization and downsampling, are outlined (Figure 4.5).

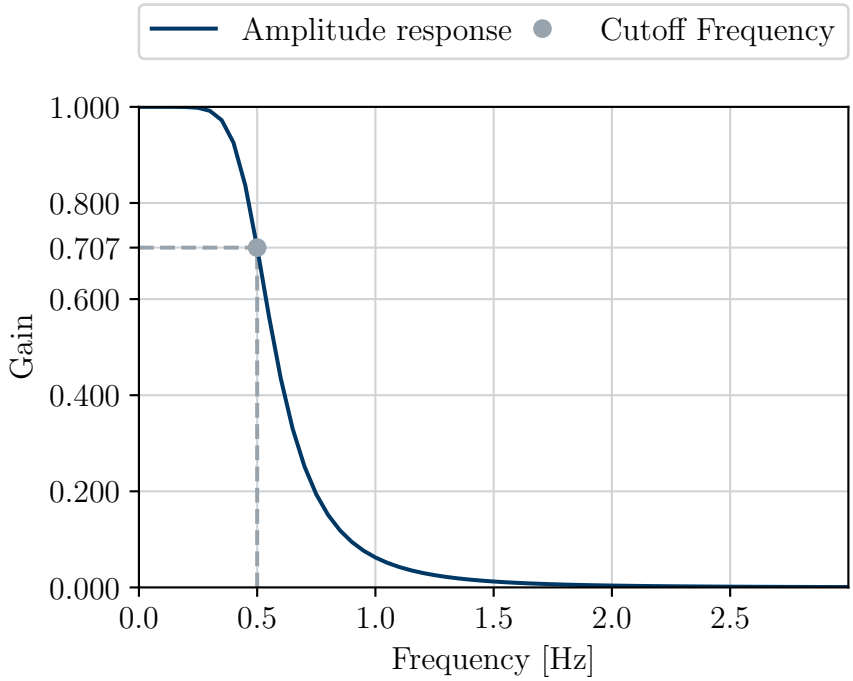

Figure 4.4: **Butterworth Lowpass Filter Gain.** All frequencies below the cutoff frequency are transmitted unattenuated, whereas frequencies above are damped by more than 90% and frequencies above 2 Hz are completely eliminated.

Filtering, Smoothing, Squaring

To achieve physically meaningful units, first of all the raw data of $\boldsymbol{g}_{si}$ was calibrated according to Ferraris et al. [Fer95]. Afterwards, lowpass filtering with a *Butterworth filter* h_{butter} with order $n = 4$ and cutoff frequency $f_c = 0.5\,\text{Hz}$ was applied to only retain frequencies beneath the human locomotor band [Ilu13] resulting in the modified signal $\boldsymbol{g}_{filt}$ (Equation 4.2).

$$\boldsymbol{g}_{filt} = \boldsymbol{g}_{si} * h_{butter} \tag{4.2}$$

A *Butterworth filter* is characterized by its amplitude response in the passband, which is almost constant. For the cutoff frequency, the attenuation is exactly $3dB$, i.e. $\frac{1}{\sqrt{2}} \approx 0.707$ of the amplitude of the input signal. The steepness of the gain function's gradient is increased with a higher filter order [But30] (Figure 4.4).

Afterwards, to eliminate small, impulsive peaks caused by single steps, whilst preserving large peaks resulting from complete turnings, the filtered signal g_{filt} was smoothed using a *median*

filter [Fis20]. This nonlinear filter is replacing each sample with the median of its neighboring values. The investigated area was chosen to be two seconds, which means that for a signal with the sampling frequency f_s a window width $w = 2s \cdot f_s$ was used.

$$\boldsymbol{g}_{smooth} = median(\boldsymbol{g}_{filt}) \tag{4.3}$$

Finally, the resulting signal $\boldsymbol{g}_{smooth}$ (Equation 4.3) was squared for every sample $\boldsymbol{g}_{smooth}[n]$ (Equation 4.4) to enhance the resulting turning peaks and to obtain only positive values that were regardless of the direction of rotation.

$$\boldsymbol{g}_{sq}[n] = (\boldsymbol{g}_{smooth}[n])^2 \text{ for } n = 0, 1, \ldots, N-1 \tag{4.4}$$

Normalization

Within the implementation of Fischer et al., the raw data was normalized to a maximum value range of $[-1, 1]$ by its sensor range as a first preprocessing step [Fis20]. The used gyroscope sensors covered a value range of $\pm 2000°/s$, whereas the peak turning velocity within a $180°$ turn was measured to be $162.3°/s$ on average within PD patients [Sal09]. Thus, when normalizing the gyroscope signal to the sensor range, the results of Fischer et al. did not exhaust the full spectrum of values. Consequently, to achieve more intuitive values and increase the comparability of different subsections of the recordings, a uniform scaling of each data subset was desirable. Therefore, in this thesis, *min-max normalization* was performed after squaring to map all values of the preprocessed signal to $[0, 1]$. For this linear transformation, minimum and maximum of the considered subsequence $\boldsymbol{g}_{sq}$ were estimated and used to calculate the new value $\boldsymbol{g}_{norm}[n]$ for every sample $\boldsymbol{g}_{sq}[n]$ of the squared signal:

$$\boldsymbol{g}_{norm}[n] = \frac{\boldsymbol{g}_{sq}[n] - min(\boldsymbol{g}_{sq})}{max(\boldsymbol{g}_{sq}) - min(\boldsymbol{g}_{sq})} \text{ for } n = 0, 1, \ldots, N-1 \tag{4.5}$$

Downsampling

In order to reduce the runtime of the following template matching (section 4.2.2), the pipeline by Fischer et al. was extended by an additional preprocessing step. The fundamental idea was to reduce the number of points in the preprocessed signal $\boldsymbol{g}_{norm}$ without losing information.
The Nyquist–Shannon sampling theorem states that the highest frequency that can be accurately recorded within a discrete signal is as high as half its sampling frequency [Osh06]. Hence, it determines a lower border for the minimal amount of samples per second to be stored if the

maximal frequency is known. As described above in section 4.2.1, a *Butterworth filter* was applied to the preprocessed gyroscope signal already. Consequently, the filtered signal contains a negligible amount of frequencies greater than $1\,\mathrm{Hz}$ (Figure 4.4). A sampling rate of $2\,\mathrm{Hz}$ was therefore decided to be sufficient for storing all necessary information. Consequently, the signal $\boldsymbol{g}_{norm}$ was downsampled to $2\,\mathrm{Hz}$ resulting in the final preprocessed result $\boldsymbol{g}_{resampled}$.
To determine the new values of this resampled signal, a signal function was generated using *linear interpolation*. For this method, first the distance between two samples was calculated for the reduced sampling rate. Based on that result, a new list of sampling points was generated. The corresponding function values were determined from a continuous signal function, which was generated by connecting adjacent signal values with a straight line. This procedure was implemented using the SciPy function `interpolate.interp1d` [Vir20].

4.2.2 Template Matching

After preprocessing, the aim was to detect the pattern of rotations described in assumptions (ii) and (iii) using the enhanced turning peaks around the $\boldsymbol{g}_{si}$ axis, which can be achieved by template matching. This is a pattern recognition method that involves comparing a known reference pattern, the so-called template, with a unknown test pattern, aiming at measuring their similarity [The09]. There are different possibilities to compare a template with the investigated signal. For the application in this thesis, sequences with different lengths needed to be matched, since 4x10MWTs were performed in different, self-selected velocities and thus vary in duration according to assumption (iv). Furthermore, the peaks in the signal caused by turnings were not distributed evenly along the signal, since the breaks between single 4x10MWTs and the ratio of different walking speeds and thus the distance between the individual rotations differed between individual subjects.
Therefore, a non-linear template matching method that allows the calculation of similarity between two time series of different lengths was required. Subsequence dynamic time warping (sDTW) is suitable for this use case, as it allows pattern matching including parts of the template being stretched or squeezed to achieve an optimal fit [Mü07]. In the following, the template creation procedure is elucidated followed by an outline of the basic working principle of sDTW along with possible variations.

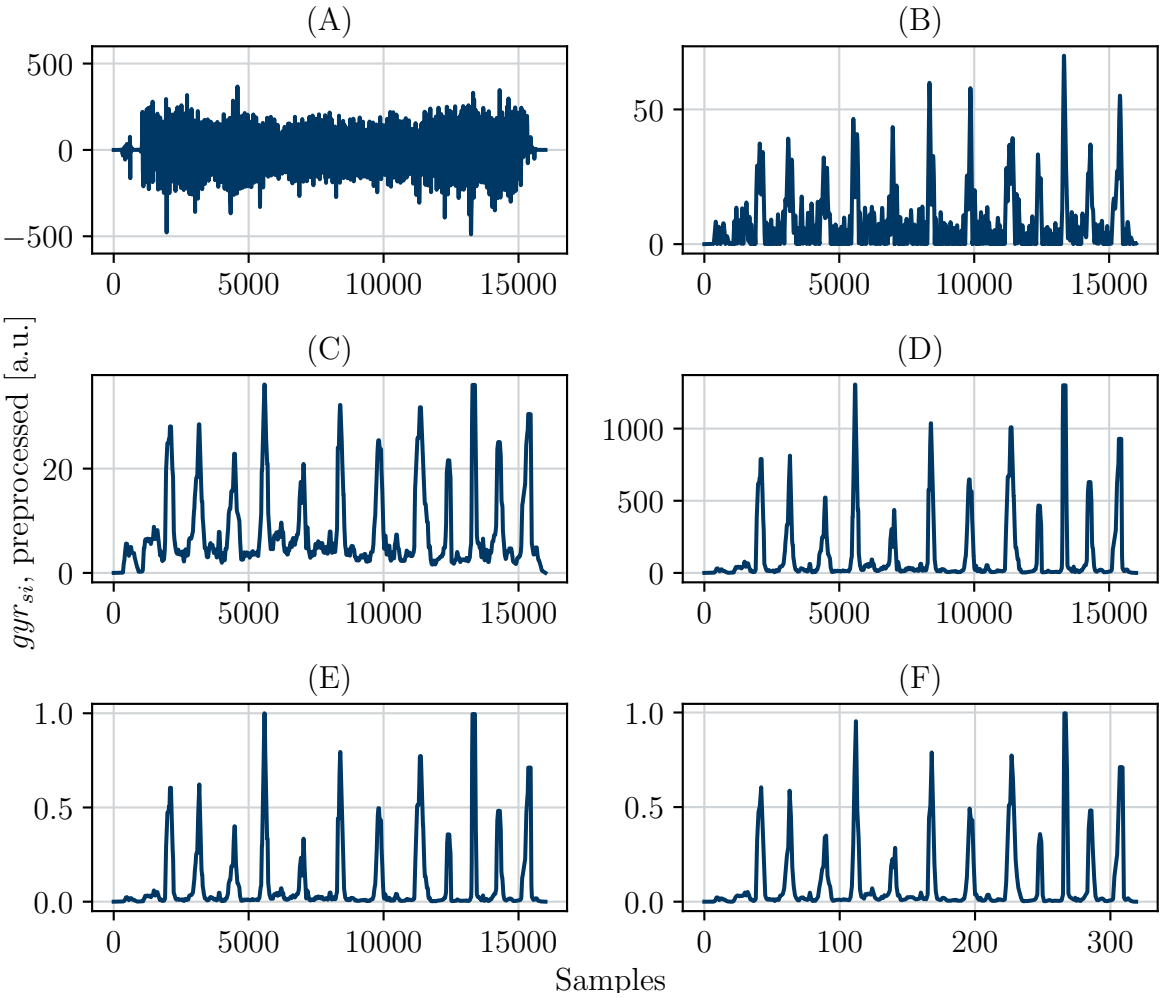

Figure 4.5: **Preprocessing.** The change of the input signal $\boldsymbol{g}_{si}$ *(A)* is displayed after every step of the preprocessing pipeline. Lowpass filtering leads to $\boldsymbol{g}_{filt}$ *(B)*, followed by $\boldsymbol{g}_{smooth}$ *(C)* after the application of a smoothing filter. The signal was then squared resulting in $\boldsymbol{g}_{sq}$ *(D)*, and normalized to $\boldsymbol{g}_{norm}$ *(E)*. The downsampling procedure in the last step does not considerably change the outcome and was thus performed to increase efficiency. The resulting output signal $\boldsymbol{g}_{resampled}$ *(F)* was used for template matching.

Template Creation

When investigating the preprocessed $\boldsymbol{g}_{si}$ axis, every *gait test series* is represented as a consecutive series of turning peaks. Thus, the basic component for a template was the preprocessed $\boldsymbol{g}_{si}$ measurement of a transversal 180° turning.
Fischer et al. aimed at detecting separate 4x10MWTs. Consequently, they suggested creating a template by averaging different randomly picked preprocessed $\boldsymbol{g}_{si}$ signals from manually labeled 4x10MWTs performed in the clinic [Fis20]. Their resulting template consisted of three subsequent turning peaks, since there was no intermediate rotation as stated in assumption (i) (Figure 4.6).
However, in this thesis the template matching aimed at the detection of *gait test series*, and as stated in assumption (iii) it is presumed that every compliant *gait test series* included twelve turnings. Therefore, a *gait test series* template was created by lining up twelve succeeding turnings.
For this purpose, template creation by averaging resampled manually labeled 4x10MWTs or *gait test series* was not feasible for semi-standardized 4x10MWTs in a free-living environment. Due to their unsupervised character, distances between subsequent peaks differed both between and within subjects due to different speed levels, resting period lengths, and varying spatial conditions. Despite resampling the 4x10MWTs to their average length, double peaks occurred (Figure 4.6).
To avoid further manual inspection or quality estimation, in this work, a semi-artificial template creation method was used. That was possible, because the preprocessed $\boldsymbol{g}_{si}$ signal of a *gait test series* describes recurrent turnings, and thus is a periodically repeating pattern. To recreate this, three subjects, and for each subject, one manually labeled preferred speed 4x10MWT, were chosen randomly. The $\boldsymbol{g}_{si}$ signals were extracted and preprocessed separately. The length of the resulting template was determined by the mean length of the extracted sequences. Each sequence was then resampled to this length using *linear interpolation*. From the resulting sequence, an even turning peak with a distinct edge was chosen and extracted after visual inspection. This peak provided the basic component for the template. It was then repeated twelve times subsequently. This procedure was implemented for the foot as well as the hip mounted sensor signal to prevent possible distortion of the peak shape resulting from different sensor placements (Figure 4.6).

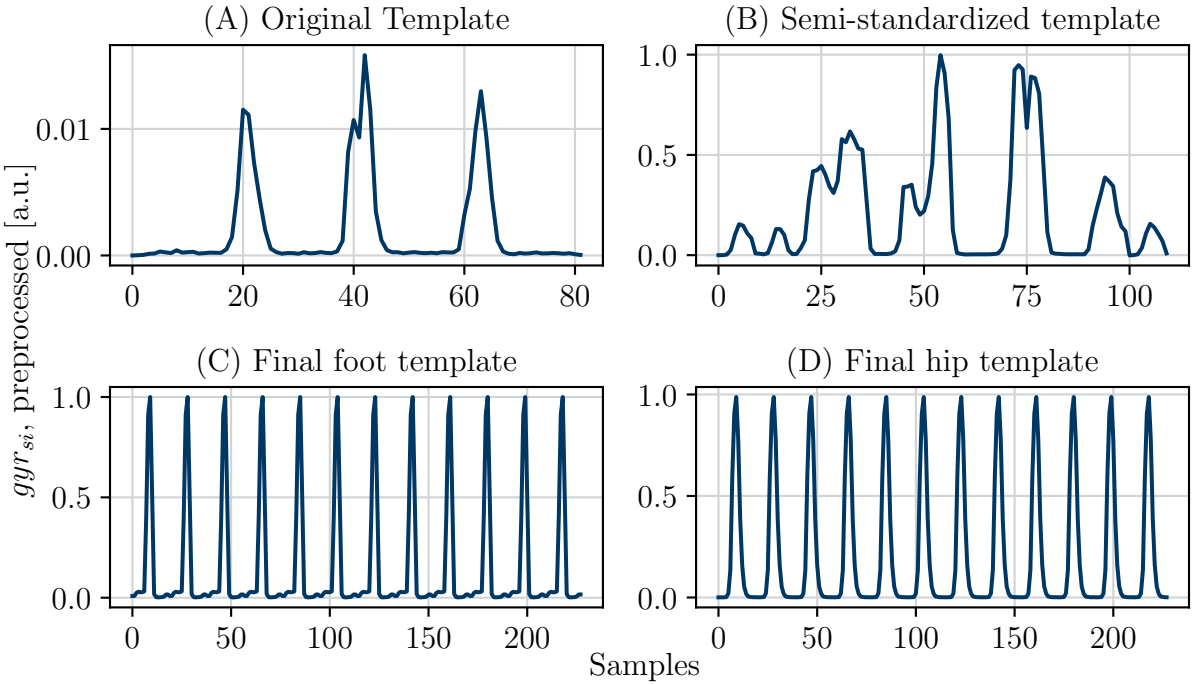

Figure 4.6: **Template creation.** The template used by Fischer et al. [Fis20] constructed from single 4x10MWT is shown in *(A)*. *(B)* outlines the result when averaging five unsupervised 4x10MWTs. The templates finally used for this work are displayed in *(C)* and *(D)* for foot and hip mounted sensors, respectively.

Subsequence Dynamic Time Warping

The fundamental idea of dynamic time warping (DTW) is mapping a template function to a given sequence and calculating the cost for this procedure [Mye81]. To be able to make a reliable prediction about the similarity of sequences, the lowest possible costs of this warping operation must be investigated.
An extension of this concept that allows the application to longer time series is sDTW. This includes a template being warped to a longer time series to determine the subsequences that show the greatest similarity. Based on that, the beginning and end timestamp of the corresponding match can be determined. Even though DTW has its origin in speech recognition [Mü07], it is also successfully applied in gait analysis, e.g. for stride segmentation [Bar15]. In the following, the algorithm is described more in detail.

Distance Matrix First of all, a distance matrix D was calculated for the combination of template t and the input signal, in the following referred to as s. Each matrix entry $D(m, n)$ was determined by calculating the Euclidean distance between the function values of $t[m]$ and $s[n]$:

$$D(m,n) = \sqrt{(t[m] - s[n])^2} \tag{4.6}$$

Consequently, the closer two values $t[m]$ and $s[n]$ are in Euclidean space, the smaller is the corresponding entry in the distance matrix. If t has the length M and s the length N, the resulting value ranges for m and n are given by

$$\begin{aligned} m &\in \{0, 1, \ldots, M-1\} \\ n &\in \{0, 1, \ldots, N-1\} \end{aligned} \tag{4.7}$$

which caused D to be a $M \times N$ matrix. This results in the runtime complexity of sDTW to be $\mathcal{O}(M \cdot N)$[Mü07].

Accumulated Cost Matrix Given the distance and thus the cost for warping each pair of points of template t and input signal s to one another, the next step was to determine the overall cost for warping t completely to s in a specific, preferably cheap way. Therefore, every sample of s was considered as a possible starting point for matching with t. In order not to have to test all possible combinations for the most favorable one, which would have caused an exponential runtime complexity, a dynamic programming approach was used [Bar15, Mü07].

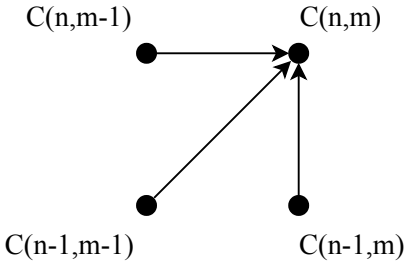

Figure 4.7: **Accumulated Cost Matrix Calculation.** A new entry is calculated dynamically by comparing its three predecessors and adding the minimum cost value. [Fis19]

First, the costs for warping the first template sample $t[0]$ to every sample $s[n]$ were stored to the bottom row of C. Since the Euclidean distances were used as a cost measure, the values were simply transferred from D:

$$C(0, n) = D(0, n) \tag{4.8}$$

In a similar way, the first column of C was also initialized. For warping every sample $t[m]$ for $m = 0, 1, \ldots, M - 1$ to $s[0]$, the corresponding distance values of D needed to be summed up.

$$C(m, 0) = \sum_{i=0}^{m} D(i, 0) \tag{4.9}$$

Given these cost values, all other costs were calculated by searching for the cheapest match within the previous, directly surrounding entries (Figure 4.7). Finally, the distance between the last appended sample pair (m, n) needed to be added to this cheapest value to receive the entry $C(m, n)$. This calculation specification can be denoted as follows:

$$C(m, n) = min\{C(m - 1, n - 1), C(m - 1, n), C(m, n - 1)\} + D(m, n) \tag{4.10}$$

In summary, the minimum costs for warping the first m samples of t to s ending at the nth sample of s were stored in the accumulated cost matrix entry $C(m, n)$. When filled completely, the upper row entries of the accumulated cost matrix $C(M - 1, n)$ were interpreted as the lowest costs for warping the complete template t to s ending at the respective sample $s[n]$. Consequently, the upper row can be displayed as a discrete function, which is in the following referred to as cost function Δ. Local minima of Δ indicated the best matches. To determine the beginning and end of a possible *gait test series*, the warping path must be reconstructed first.

Warping Path Each entry of the cost function contains the total costs for one warping path w. More precisely, the warping path describes, which pairs of points (m, n) are mapped to each other. This means that each sample m of template t was uniquely assigned to a sample n of the input signal s. w was expressed as a series of index pairs referring to entries of D (Equation 4.11). Consequently, w was interpreted as the path that exactly yielded the warping cost described in $\Delta[n]$.

$$w = (w_0, \ldots, w_{L-1}) \text{ with } w_l = (m, n) \in C(m, n);\ l \in \{0, \ldots L-1\} \tag{4.11}$$

This warping path depends on which of the adjacent entries (Figure 4.7) was selected when creating the accumulated cost matrix and could therefore be reconstructed backwards for each element of the cost function. There were some limitations to be observed [Mü07]:

- w always stretches from the top to the bottom line of the accumulated cost matrix, i.e. it describes at which point of the warping sequence the beginning and end of the template are matched, respectively.
- w progresses monotonously, i.e. a template sample is always mapped to the same or the succeeding gait sequence sample as its predecessor, never to a previous one.
- Subsequent entries of w are maximally one index apart for every dimension, where one index is always changing.

Following these constraints, the calculation of w was performed. Starting at the end of the matched sequence, the cheapest of the previous entries of D was read. The result was appended to the warping path w. This was repeated, always considering the last entry added to w, until the beginning of the matched sequence was reached.

Input Signal

There were two distinct functions suitable to be utilized as input signal s for the sDTW algorithm. One one hand, a single-dimensional subsequence dynamic time warping (ssDTW) approach can be used, which means that the distance matrix was calculated based on the g_{si} axis separately for every sensor position and corresponding template. As a result, the whole *gait test series* detection step was performed three times for every subsequence per recording day, once for every single sensor dataset.

On the other hand, taken into account the fact that all sensor data used was synchronized, multi-dimensional subsequence dynamic time warping (msDTW) can be applied. Therefore, as input signal for the sDTW algorithm a matrix consisting of a vector for every sample was used (Equation 4.12).

$$s = \begin{pmatrix} \boldsymbol{g}_{si,left} \\ \boldsymbol{g}_{si,right} \\ \boldsymbol{g}_{si,hip} \end{pmatrix} = \begin{pmatrix} \boldsymbol{g}_{si,left}[0] & \dots & \boldsymbol{g}_{si,left}[N-1] \\ \boldsymbol{g}_{si,right}[0] & \ddots & \vdots \\ \boldsymbol{g}_{si,hip}[0] & \dots & \boldsymbol{g}_{si,hip}[N-1] \end{pmatrix} = \begin{pmatrix} \boldsymbol{g}_{all}[0] \\ \vdots \\ \boldsymbol{g}_{all}[N-1] \end{pmatrix}^T \tag{4.12}$$

A 3-d template was created accordingly. Similar to the 1-d case, the cost matrix was simply calculated by determining the Euclidean distances between the vectors for each sample. This has the advantage that the algorithm only needed to be applied once for every gait sequence of each recording day.

Match Selection

To determine, which elements of the cost function Δ were taken into account for further investigations, a threshold Θ was applied to Δ. All local minima beneath Θ were calculated using the SciPy function `find_peaks` [Vir20]. Θ was found to be a critical parameter and required optimization (section 5.2.2). As a result, a list of all indices of Δ that were considered as possible matches was received.
For these gait test candidates, the warping path was reconstructed. Its beginning and end index, i_{start} and i_{end}, respectively, were calculated and stored as they represent start and stop of a possible 4x10MWT (Figure 4.8).

4.2.3 Postprocessing

Due to turning and other movements of the patients directly before and after the *gait test series*, or inaccuracies when performing the 4x10MWTs, several local minima occurred in the cost function. They were often located tightly together, and therefore caused several overlapping matches. Fischer et al. assumed that the absolute minimum of the cost function is always the optimal match. However, this caused the resulting interval limits to be imprecise [Fis20]. Furthermore, false positives might have been found caused by activities of daily life including irregular turnings (Figure B.2).
To find precise borders, and simultaneously include non-compliant tests and exclude coincidentally similar daily-live movements, postprocessing was performed, not only to finalize whether a *gait test series* was in the gait sequence, but also to determine the corresponding border labels as accurately as possible. Therefore, the detected gait test candidates, i.e. the matches identified as local minima of the cost function below Θ, were checked against some constraints. First of all, for each match the $\boldsymbol{g}_{si}$ interval between beginning and end index was extracted (Equation 4.13).

$$\boldsymbol{g}_{candidate} = (\boldsymbol{g}_{si}[i_{start}], \boldsymbol{g}_{si}[i_{start}+1], \ldots, \boldsymbol{g}_{si}[i_{end}]) \tag{4.13}$$

Second, to be confirmed as a *gait test series*, two crucial requirements were defined:

1. $\boldsymbol{g}_{candidate}$ has an adequate length.
2. $\boldsymbol{g}_{candidate}$ has a sufficient amount of turnings.

If these conditions were met, an interval was classified as a *gait test series* (Figure 4.5). Requirement 1 was confirmed by simply applying thresholds for the minimum and maximum length of a *gait test series*. They were estimated by averaging the lengths of all manually labeled 4x10MWTs (section 6.1.1). For the minimum length of one 4x10MWT, the mean duration of the fast and thus shortest 4x10MWTs was used subtracting twice its standard deviation (Equation 4.14). For the maximum length, this procedure was repeated analogously using the slow 4x10MWTs and adding the doubled standard deviation (Equation 4.15).

$$\kappa_{min} = \mu_{fast} - 2 \cdot \sigma_{fast} \tag{4.14}$$

$$\kappa_{max} = \mu_{slow} + 2 \cdot \sigma_{slow} \tag{4.15}$$

When considering a compliant *gait test series*, the detected candidates $\boldsymbol{g}_{candidate}$ of the sDTW contained three 4x10MWTs each, therefore their length was required to range between $3 \cdot \kappa_{min}$ and $3 \cdot \kappa_{max}$. To confirm the number of turnings within $\boldsymbol{g}_{candidate}$, Fischer et al. used a measurement based on the derivative of the cost function [Fis20]. In this thesis, however, two other attempts are compared, which are independent of the sDTW result. They are introduced in the following.

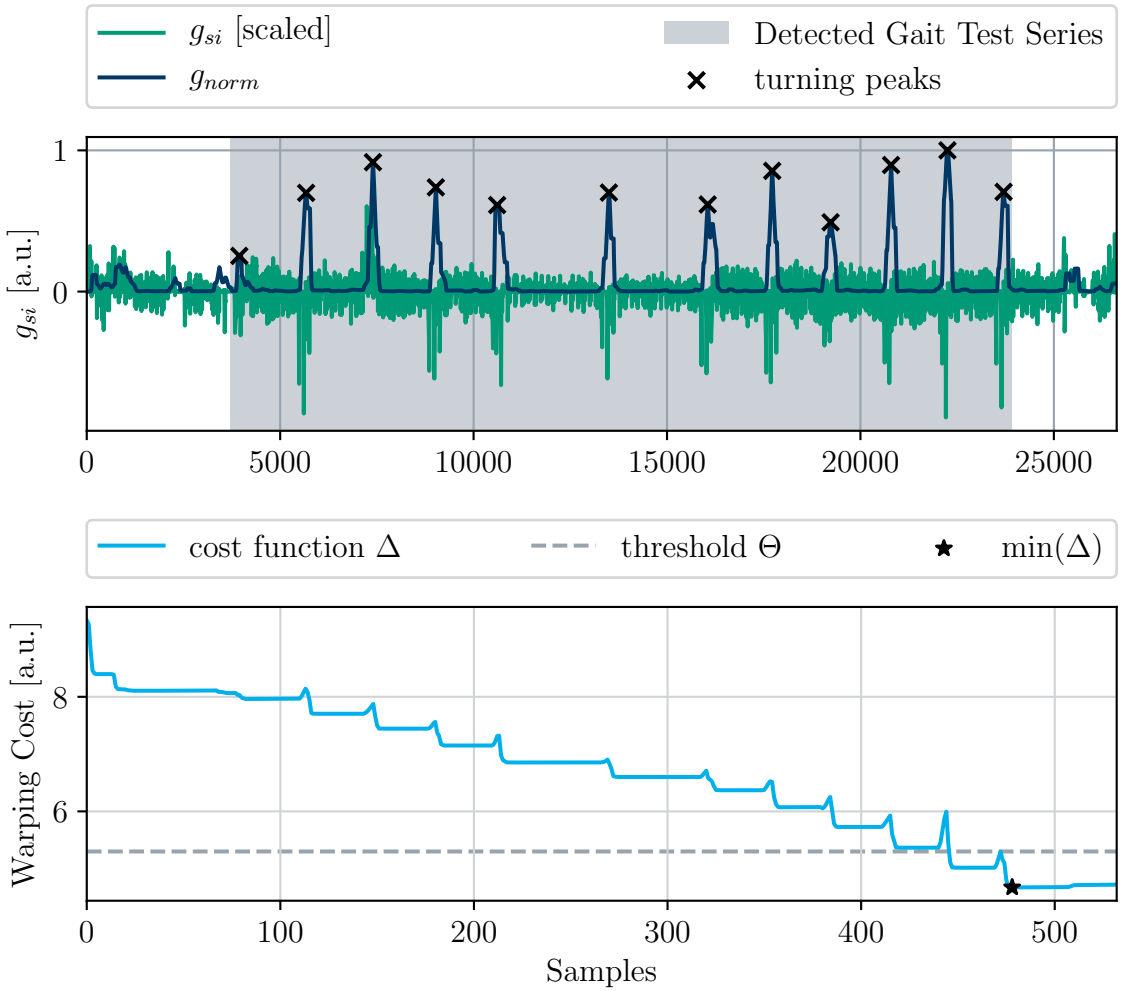

Figure 4.8: **Gait Test Series Detection.** Here, the result after applying the complete *gait test series* detection is displayed. Using msDTW and a template consisting of twelve turnings resulted in the cost function Δ shown in the lower figure. Afterwards, the local minima of the cost function below Θ were determined, and the interval borders resulting from the corresponding warping path calculated. The most favorable one, which fulfilled all constraints, was chosen. The corresponding interval limits were transferred back to the original sampling rate of the signal and are presented in the upper figure.

Peak Detection

Preprocessing was performed to highlight turning peaks (section 4.2.1). Therefore, the most intuitive approach to determine the number of rotations in the *gait test series* candidate sequence was to count the local maxima of this signal. For this purpose, the `find_peaks` function by SciPy [Vir20] was used. Additionally, the function parameters *peak prominence* and *distance* were defined. The first was set to sum of the $\boldsymbol{g}_{si}$ signal's mean and standard deviation to only detect peaks above noise level. The latter was chosen to be the minimum approximate distance between two turning peaks. To estimate this value, the mean and standard deviation of all manually labeled 4x10MWTs in preferred speed were used similar to the length threshold parameter assessment (section 4.2.3). As four turning peaks are expected per 4x10MWT, the result was divided by four (Equation 4.16).

$$\kappa_{max} = \frac{1}{4} \cdot (\mu_{pref} - 2 \cdot \sigma_{pref}) \tag{4.16}$$

The number of turning peaks was then determined separately for every sensor's signal for both ssDTW and msDTW. A sequence $\boldsymbol{s}_{candidate}$ was rejected, when its estimated number of turning peaks was not part of the fixed range $[\eta_{min}, \eta_{max}]$. These interval limits were found to be decisive for the classification results and thus their values were optimized (section 5.2.2).

Turning Detection

The second attempt to calculate the number of turning peaks in a signal subsequence is based on the physical meaning of gyroscope signals as they represent an angular velocity. When integrated over time, the covered distance, i.e. the turning angle, is retrieved. This was used to detect considerable turnings.

In detail, an algorithm adopted from El-Gohary et al. [EG14] was used. This approach only works when applied the hip sensor signal. Especially elderly, motor-impaired patients perform turns slowly taking numerous small steps [Sta08]. Therefore, no complete $180°$-turnings can be reconstructed by foot sensor gyroscope signal integration. As a consequence, the turning detection according to El-Gohary et al. is not applicable for signals recorded by lower limb mounted sensors [EG14]. Thus, it is suitable for msDTW, but not for the separated ssDTW assessment.

The calculations were done using the calibrated signal of the lower back mounted sensor $\boldsymbol{g}_{si,hip}$. First, it was lowpass filtered to $\boldsymbol{g}_{filt}$ with a *Butterworth filter* (section 4.2.1) to remove high frequency noise. The cutoff frequency was $f_c = 1.5\,\mathrm{Hz}$.

Afterwards, local maxima above a threshold of $15°/s$ were searched for in this preprocessed signal. They were interpreted as indicators for turns. For each detected maximum, the closest adjacent points on the left- and right-hand side of the peak intersecting with a horizontal line at $5°/s$ altitude were determined. These crossing points were considered as beginning and end of a turn, respectively. To include potentially missing sections, all turning intervals with a shorter intra-turn distance than $10s$ were concatenated if the consecutive turns were performed in the same direction. Whether two turns were carried out in the same direction could be determined by considering the sign of the integral $\boldsymbol{g}_{filt}$. If the rotation was performed in mathematically positive direction around the superior to inferior axis, the integral value is positive and vice versa. [EG14] All detected turns were undertaken duration thresholds excluding all turns shorter than $0.5s$ or longer than $10s$. Finally, the angle of all leftover turning sequences was determined by integrating over $\boldsymbol{g}_{filt}$, which, as the signal is discrete, corresponds to summing up all values within the respective section:

$$\alpha = \sum_{i=i_{start}}^{i_{end}} \boldsymbol{g}_{filt}[i] \tag{4.17}$$

The original implementation aimed at differentiating between distinct turning angles [EG14]. Because target for this application was a binary decision, whether a turning is considerable or not, only a broad threshold for the turning angle α was used to classify a turn as valid. All angles smaller than $\alpha_{min} = 80°$ and larger than $\alpha_{min} = 240°$ were excluded.
This turning detection approach was applied to the msDTW pipeline, using the same criteria as described in the peak detection approach (section 4.2.3). That means, every sequence with an overall turning number outside the range $[\eta_{min}, \eta_{max}]$ was rejected.

Overlapping Matches

Even after filtering the matches according to the two constraints explained above, intervals with overlapping borders occurred. In this case, another postprocessing step was applied to eliminate overlapping matches.
As the template covered twelve turnings, the aim of finding a complete, compliant *gait test series* was achieved with a single match. Therefore, the cost function values of all overlapping intervals were compared and only the cheapest match was kept (Figure 4.9).

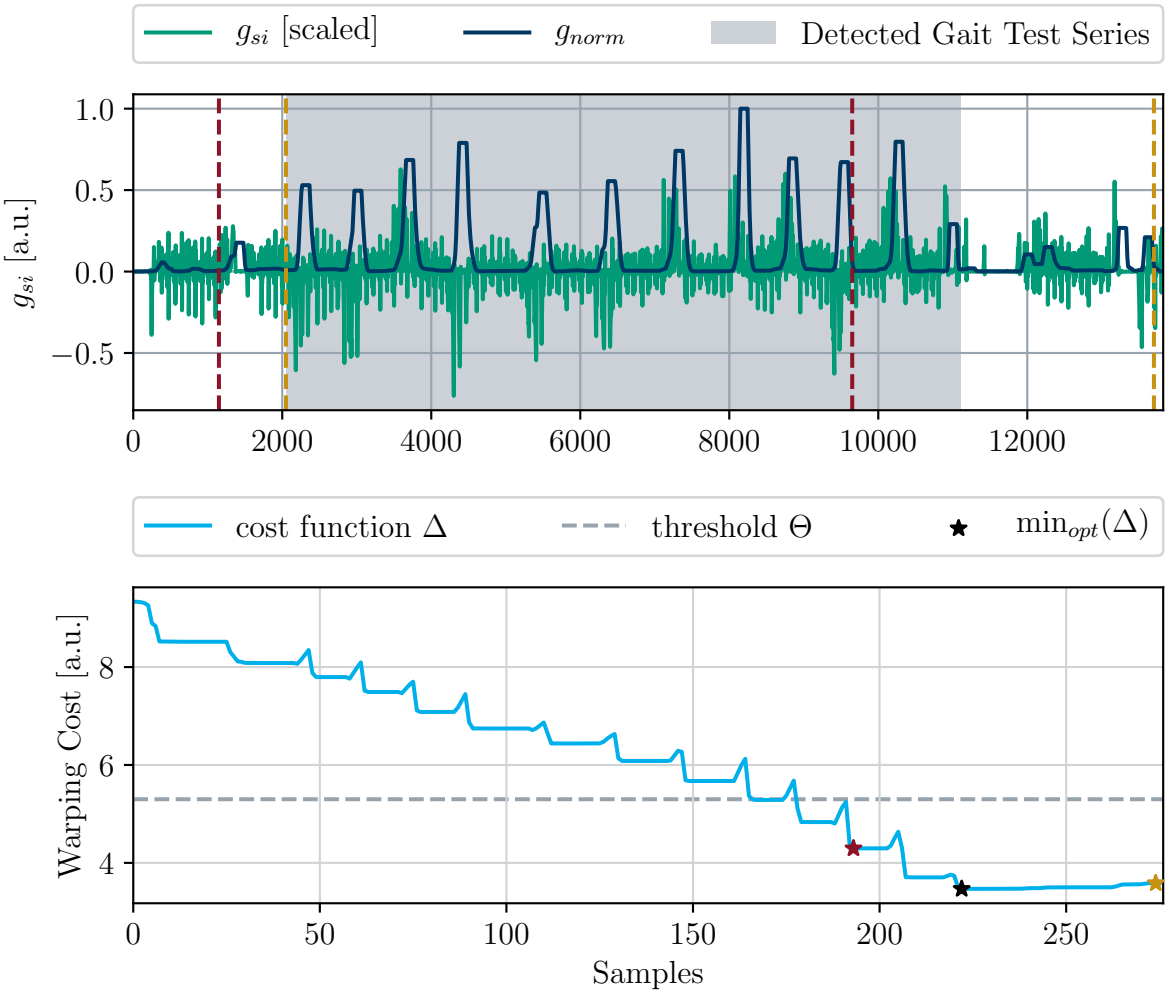

Figure 4.9: **Overlapping Matches.** In this example, three matches were detected fulfilling all constraints. The match with the lowest corresponding warping cost $\min_{opt}(\Delta)$ was chosen as final *gait test series* interval, which is highlighted grey. The excluded matches resulting from the other local cost function minima and their corresponding borders are marked in red and yellow, respectively.

4.3 Gait Test Series Decomposition

To enable separate evaluation and interpretation of the single 4x10MWTs included in a detected *gait test series*, it was required to find their borders and assign the correct speed levels to each 4x10MWT. This was implemented stride-based, i.e. strides were segmented from the *gait test series*. Their stride velocity was then estimated and described as a function over time. For this function, an edge detection was performed to detect steep changes in the velocity. The points in time of this changes were interpreted as 4x10MWT borders. Afterwards, for each detected 4x10MWT, the relative speed level was assigned resulting in the final 4x10MWT labels including start and end time as well as velocity (Figure 4.10).

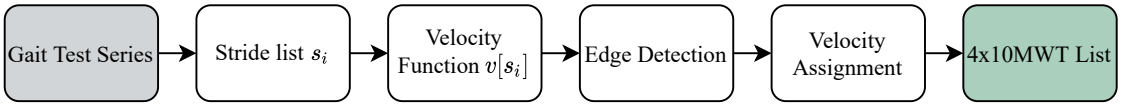

Figure 4.10: **Flowchart for the proposed Gait Test Series Decomposition Pipeline.**

4.3.1 Stride Velocity Calculation

To measure the velocity of each stride in the detected *gait test series*, first of all, a stride segmentation was performed. For this purpose, an algorithm according to Barth et al. using msDTW was applied to the foot sensor signal [Bar15]. As input signal, all axes of the gyroscope signal were used. Since the detected strides differ depending on the investigated foot, only the signal of one foot sensor was used for the following calculations. For the refinement of this stride list, a consistency check was performed based on characteristic temporal gait events such as IC, MS, and TC (section 2.1) using signal processing as introduced by Rampp et al. [Ram15]. Any stride that did not contain the events in the mentioned order was excluded from further evaluation. As a result, for all confirmed strides s_i, the corresponding timestamps were returned. Two subsequent mid stance events were defined as beginning and end of a stride. For each stride, the stride length was calculated by double integrating the gravity corrected, drift compensated accelerometer signal. The stride length was defined as the euclidean distance between beginning and end point of a stride in the floor plane [Ram15]. Afterwards, the velocity was calculated by dividing the length of each stride by stride time, resulting in a function v of velocities $v[i]$ for every stride s_i and a total of S strides:

$$v = (v[0], v[1], \ldots, v[S-1]);\ i \in \{0, 1, \ldots, S-1\} \tag{4.18}$$

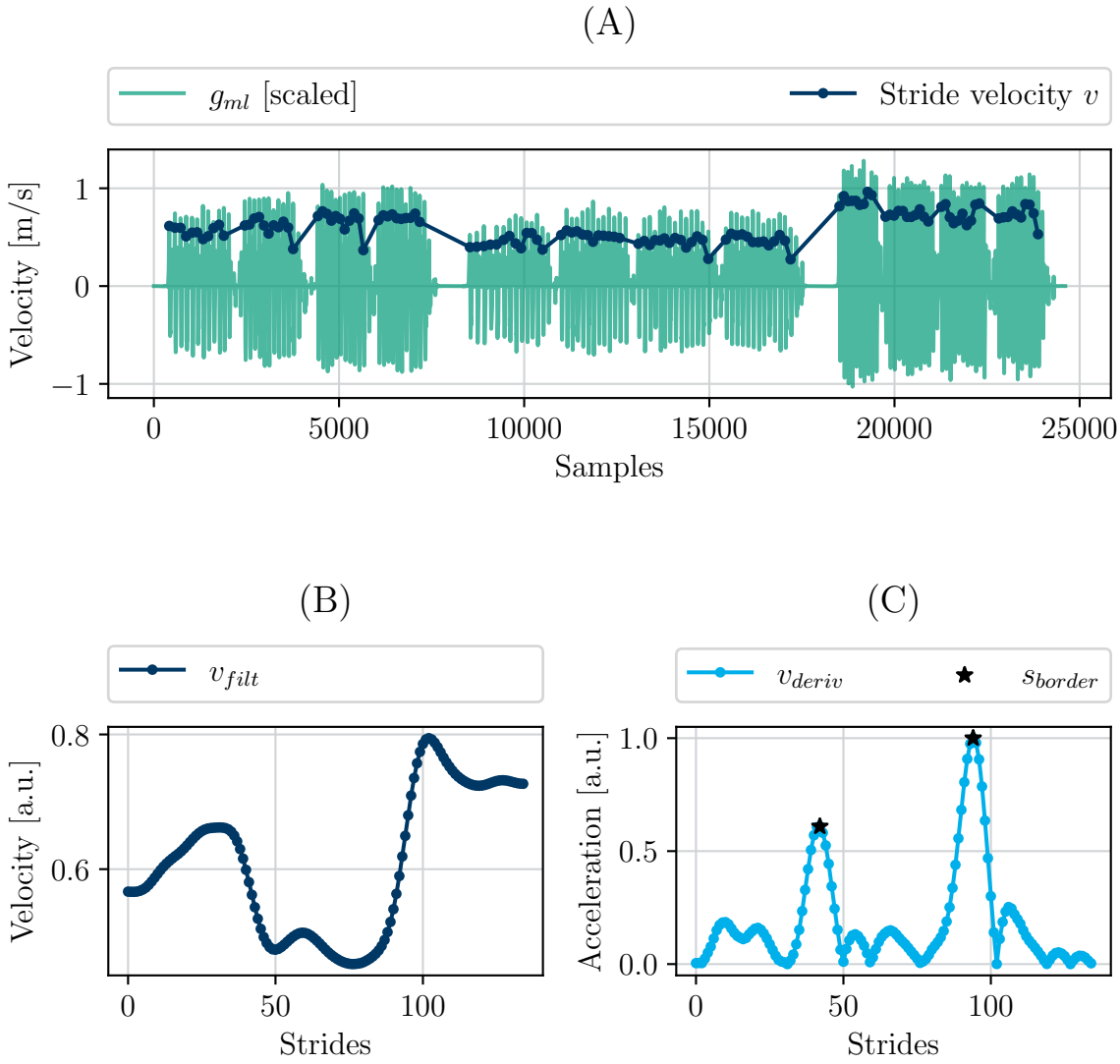

Figure 4.11: **Edge Detection for Gait Test Border Calculation.** In *(A)*, the stride velocity progression over one *gait test series* is shown. After smoothing, the velocity function looks as displayed in *(B)*. Derivation, absolute value calculation and normalization yield a function with two clear peaks depicted in *(C)*. The corresponding strides were interpreted as the last strides of the respective 4x10MWT.

4.3.2 Gait Test Border Detection

To detect steep changes in the velocity function v and thus transitions between 4x10MWTs of different speeds, edge detection was performed for v. This was accomplished using a signal processing pipeline consisting of smoothing, edge enhancement and peak detection (Figure 4.11). The details are outlined in the following.

Smoothing

During one 4x10MWT, a subject's gait velocity is usually slightly varying caused by deceleration during turns and acceleration on the straight $10\,\mathrm{m}$ walkway. When considering the velocity function v, these variations can be interpreted as noise. Before the application of a derivative-based edge detection filter, noise reduction is required in order to only keep sharp changes and prevent noise amplification. Hence, the velocity function v was lowpass filtered using a Gaussian window as filter kernel (Equation 4.19), which is a common smoothing filter for edge detection [Bas02, Can87]. The window was created using the SciPy function `gaussian` [Vir20].

$$h_{gauss}[n] = e^{-\frac{1}{2}(\frac{n}{\sigma})^2} \tag{4.19}$$

Crucial parameters are the standard deviation σ, which defines the width of the underlying Gaussian bell, and the filter width w in samples. The choice of w and σ is directly related. The larger w is compared to σ, the smaller are the weights of the kernel's boundary values. If w was chosen smaller in relation to σ, the boundary filter weights increased (Figure 4.12). A width of $\pm 3\sigma$ includes all values of the Gaussian bell considerably different from zero according to the so-called *three-sigma-rule* [Gra06]. Therefore, the relation between filter width and standard deviation was chosen as $w = 2 \cdot 3\sigma = 6\sigma$. The filter width w was optimized in following experiments (section 5.3). The velocity function v was then filtered by convolution with the resulting Gaussian window (Equation 4.19).

$$v_{filt} = v * h_{gauss} \tag{4.20}$$

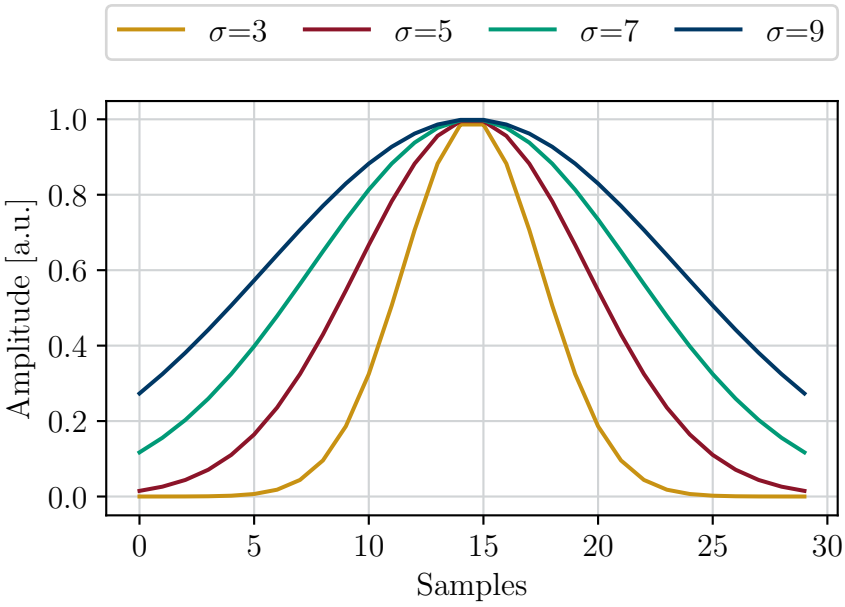

Figure 4.12: **Different Gaussian Windows.** For all depicted Gaussian windows, the filter width was set to $w = 30$, whereas the standard deviation σ was varied. The larger σ was chosen, the higher the window's border values were. For the red window, the complete range of amplitude values is just covered.

Derivation

An edge in a signal occurs when adjacent function values differ strongly, i.e. when the signal's slope is a local extremum. Consequently, edges in the smoothed signal v_{filt} were searched to find points where the velocity over w strides changed considerably. This was implemented by calculating the first derivative of v_{filt} as it represents the signal's slope. From a physical point of view, this is equivalent to computing the acceleration of v_{filt}.
For discrete signals, the derivative for each point $v_{filt}[i]$ can be approximated by calculating the slope of the secant through the preceding and following point (Equation 4.21). This is equivalent to convolution (Equation 4.22) with an one-dimensional edge detection filter kernel (Equation 4.23). As the type of the extremum was irrelevant, the absolute value of v_{deriv} was used for the following considerations. This enabled the possibility to perform a simply peak detection for finding the two largest extrema. Furthermore, the signal was *min-max normalized* as described previously (section 4.2.1). For the sake of simplicity, this postprocessed version of the derived velocity function is still referred to as v_{deriv} in the following.

$$v_{deriv}[i] = \frac{v[i+1] - v[i-1]}{2} \tag{4.21}$$

$$v_{deriv} = v_{filt} * h_{deriv} \tag{4.22}$$

$$h_{deriv} = \frac{1}{2} \cdot \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{4.23}$$

Peak Detection

The maximum values of v_{deriv} were detected using the SciPy function `find_peaks` [Vir20]. Two constraints needed to be considered. First, as stated in assumption (i), tests with three different velocities were performed, i.e. two velocity transitions with particularly outstanding peaks were expected within the *gait test series*. Therefore, the two values with the maximal absolute value were taken. In order to avoid that boundary values of the speed function were mistakenly recognized as maxima, they were checked for a second condition. Assuming that a 4x10MWT must contain at least ten recognized steps, maximum points that were detected within the first or last 15 strides of the derived velocity function v_{deriv} were excluded.

4.3.3 Gait Test Extraction

The peak detection returned two points $v[border, 1]$ and $v[border, 2]$ of the velocity function. The associated strides $s_{border,1}$ and $s_{border,2}$ were interpreted as the last stride of a 4x10MWT, respectively. That means, the first 4x10MWT ranges from the beginning of the first stride s_0 up to the end of the first border stride $s_{border,1}$. The subsequent stride of the first border stride was then assigned to the second 4x10MWT, covering all strides until and including the second border stride $s_{border,2}$. Accordingly, the area between the beginning of the stride after $s_{border,2}$ and the end of the last stride was assigned to the third 4x10MWT (Equation 4.24). Using the respective signal samples of the beginning and end point of the strides, a label list was created completely describing the points in time where all 4x10MWTs were performed.

$$\begin{aligned} \text{test}_1 &= [s_0, s_1, \ldots, s_{border,1}] \\ \text{test}_2 &= [s_{border,1+1}, \ldots, s_{border,2}] \\ \text{test}_3 &= [s_{border,2+1}, \ldots, s_{S-1}] \end{aligned} \tag{4.24}$$

Finally, to assign the three speed levels, i.e. slow, preferred, and fast walking, to the right sequences, the velocities of all strides were averaged over each detected 4x10MWT. The results were compared and assigned to the corresponding speed levels. Consequently, the 4x10MWT with the lowest average stride velocity was marked as the slow test, the one with the highest average stride velocity as the fast test, and so forth (Figure 4.13).

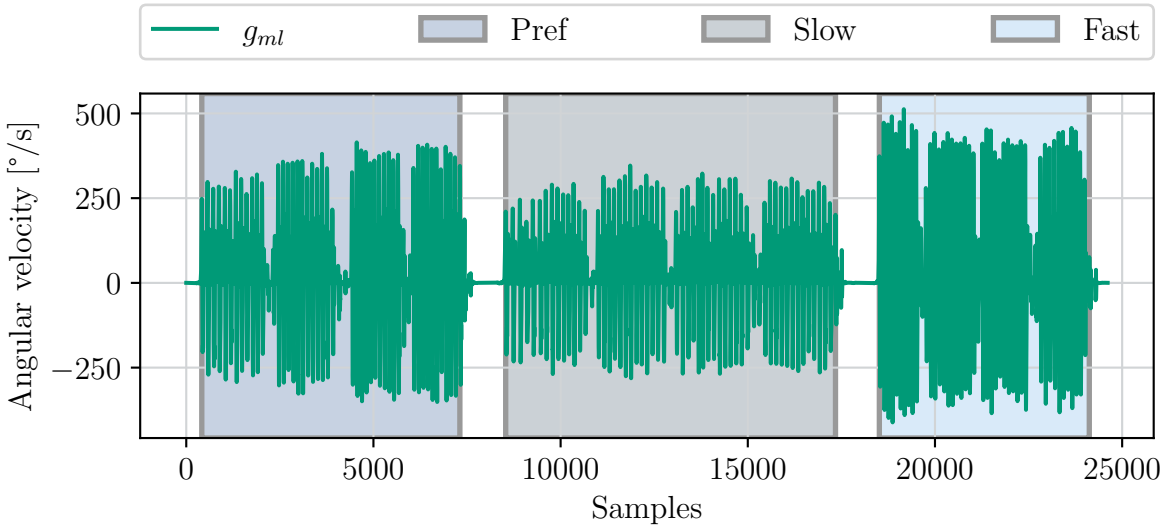

Figure 4.13: **Gait Test Series Decomposition.** As a result of the decomposition algorithm, that received one detected *gait test series* as input, the borders of the single 4x10MWT were set. For each 4x10MWT, the corresponding velocities were assigned.

Chapter 5

Experiments

In this chapter, the performance assessment criteria for evaluating the algorithms for *gait test series* detection and *gait test series* decomposition are defined (section 5.1). Additionally, experiments performed to optimize the crucial parameters of the gait test detection pipeline are explained. To receive a better overview of strengths and weaknesses of the complete pipeline, the two major parts *gait test series* detection (section 5.2) and *gait test series* decomposition (section 5.3) were evaluated separately.

5.1 Performance Assessment

For estimating performance, different statistical measures were utilized. They are always based on the number of detected true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). For the use case of this thesis, true negatives, i.e. free-living gait sequences, which did not contain a gait test, were not meaningful and not relevant for the result. Thus, *sensitivity*, *precision*, and *F1-score* were chosen as performance criteria, since they do not consider the number of true negatives.
Sensitivity, also referred to as true positive rate or recall, defines the rate of true positives in relation to the real number of positives (P) resulting from the ground truth (Equation 5.1) [Tha18]. An increasing *sensitivity* is accompanied by a higher number of true positives. However, it is not influenced by the number of false positives. Consequently, the more true positives are found, the higher is the *sensitivity*, regardless of whether the number of false positives is also increasing.

$$\text{sensitivity} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.1}$$

To estimate the relation of all true positives (TP) with regards to all positives predicted by the algorithm, i.e. all detected positives (DP), *precision*, also referred to as positive predictive value, is calculated (Equation 5.2) [Tha18]. Hence, with an increased number of false positives, *precision* decreases. However, *precision* does not consider the number of false negatives, i.e. the number of overlooked positives.

$$\text{precision} = \frac{\text{TP}}{\text{DP}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.2}$$

To combine *sensitivity* and *precision* and therefore enable simultaneous assessment of true positives, false negatives, and false positives of an algorithms' classification results, *F1-score* is used. It is calculated as the harmonic mean of *sensitivity* and *precision* (Equation 5.3) [Tha18]. *F1-score* maximizes when an algorithm achieves a high amount of true positives, while at the same time obtaining a small amount of false positives and also a small amount of false negatives.

$$\text{F1-score} = 2 \cdot \frac{\text{precision} \cdot \text{sensitivity}}{\text{precision} + \text{sensitivity}} \tag{5.3}$$

5.2 Gait Test Series Detection Evaluation

5.2.1 Evaluation Methods

The *gait test series* detection is a binary classification, which can be investigated on different levels of granularity. All evaluation methods were applied session-wise, i.e. the numbers of true positives, false positives, and true negatives were calculated separately for every session and summed up afterwards for all investigated sessions to estimate the final *sensitivity*, *precision*, and *F1-score* values.

On the coarsest level, it can be asserted that a gait sequence does either contain a *gait test series* or not. To evaluate the algorithm based on this criteria, an area-based evaluation was performed. When at least one sample of a detected match was included in a manually labeled *gait test series* of the ground truth, i.e. the area of a *gait test series* was found correctly, the detected label was classified as a true positive. Every detected label not fulfilling this condition was interpreted as false positive. The number of false negatives was received by counting the amount of ground truth labels that were not covered by any of a session's detected labels.

Furthermore, the accuracy of the predicted interval borders of a *gait test series* can be considered. If the borders of a detected label were set within a tolerance range around the ground truth labels, a detected label was counted as true positive. If the borders were outside of the tolerance range, a detected label was declared as false positive. The number of false negatives was estimated

indirectly by subtracting the amount of true positives from the sum of ground truth *gait test series*. This test-based evaluation was adapted from Fischer et al. [Fis20]. Equally to their evaluation, the tolerance range was set to $15\,\mathrm{s}$.
The most precise level for accuracy evaluation is when considering every stride of a detected *gait test series*. This stride-based evaluation is based on a stride list generated from a DTW-based stride segmentation algorithm proposed by Barth et al. [Bar15]. When a stride of a detected *gait test series* was also included in a manually labeled *gait test series*, it was interpreted as true positive. All predicted strides not included in the ground truth stride list were counted as false positives. The number of false negatives was calculated indirectly by subtracting the amount of true positives from the sum of all strides covered by the ground truth list.

5.2.2 Parameter Optimization

DTW Method	Parameter	Values
ssDTW	Ω	$2.1, 2.2, 2.3, 2.4, 2.5$
	$[\eta_{min}, \eta_{max}]$	None, $[0, 24]$, $[2, 22]$, $[4, 20]$, $[6, 18]$, $[8, 16]$, $[10, 14]$, $[11, 13]$
msDTW	Ω	$[5.2, 5.3, 5.4, 5.5, 5.6]$
	$[\eta_{min}, \eta_{max}]$	None, $[0, 24]$, $[2, 22]$, $[4, 20]$, $[6, 18]$, $[8, 16]$, $[10, 14]$
	turning verification	peak detection, turning detection

Table 5.1: **Value Ranges of Crucial Parameters for Grid Search.**

To optimize the crucial parameters, a 4-fold cross validation with grid search was performed for all algorithms. The available datasets were split randomly and patient-wise into four subsets including data of three patients each. The same split was used for optimization of all algorithms. All parameter combinations were applied to a training set consisting of three out of the four subsets. A training set thus included nine patients. The parameter combination achieving the best performance for this training dataset was tested on the remaining subset, i.e. data of the remaining three patients. For every fold, another subset served as test dataset. Based on the test results of all folds, the final performance parameters for each algorithm were calculated. As explained above, there are three different options for performance estimation. With regards to the desired further use of the detected *gait test series* for spatio-temporal gait parameter calculation of every

stride, the stride-based evaluation is most relevant. Therefore, the *F1-score* of the stride-based evaluation was chosen as optimization criterion. Test- and area-based performance were calculated additionally for evaluation, but had no impact on the parameter optimization.
A separate 4-fold cross validation was performed for the ssDTW- and msDTW-based *gait test series* detection, respectively. Crucial parameters to be optimized, and reasonable value ranges for the grid search were determined in a conservative manner based on extensive experiments. For the ssDTW-based approach, cost threshold Ω (section 4.2.2) and turning threshold $[\eta_{min}, \eta_{max}]$ (section 4.2.3) were chosen as crucial parameters. For the turning threshold, symmetric interval borders around the expected amount of detected turnings were used as parameter options for the grid search (Table 5.1). The same parameters were investigated for the msDTW-based algorithm. Additionally, it was distinguished between the applied turning verification methods (section 4.2.3), either according to El-Gohary et al. [EG14], or by using peak detection (Table 5.1).

5.3 Gait Test Series Decomposition Evaluation

The *gait test series* decomposition is a multi-class classification task, where each stride is either assigned to one of the speed level classes slow, preferred, fast, or to a rejection class, when the algorithm is not able to separate a given sequence due to unfulfilled constraints. The *gait test series* decomposition algorithm was evaluated once based on the ground truth *gait test series* labels for parameter optimization (section 5.3.1), and once based on the automatically detected labels to assess the feasibility of the complete detection pipeline (section 5.3.2). For both cases, the recording subsets from right as well as left foot sensor served as input.

5.3.1 Parameter Optimization

To optimize the results of the *gait test series* decomposition completely independent of the influence of previous pipeline steps, the manually labeled compliant *gait test series* were used as algorithm input. The evaluation was performed stride-based using the stride list calculated during the decomposition algorithm. To every stride included in this stride list, a speed level was assigned. Once, this assignment was made based on the manual gait velocity labels, and once based on the output of the decomposition, resulting in a ground truth label list and a predicted label list, respectively. Based on those label lists, a confusion matrix was calculated displaying the classification results for all of the three speed level classes slow, preferred and fast. The rejected strides were counted separately.

As crucial parameter, the window width w of the Gaussian window (section 4.3.2) was optimized using a 4-fold cross validation with grid search as described above (section 5.2). This optimization was based on the compliant dataset. The mean *F1-score* of all speed level classes was used as optimization criterion. Considering that the window size refers to the amount of strides covered when smoothing the velocity function, $w \in \{18, 24, 30, 36, 42\}$ was chosen as a reasonable range.

5.3.2 Pipeline Validation

In a second evaluation, it was examined whether the *gait test series* decomposition is not only applicable to the ground truth labels, but also to the labels generated automatically by the *gait test series* detection. Consequently, the *gait test series* labels generated for the respective test sets of the 4-fold cross validation with the best average performance (section 5.2.2) were used as input for the *gait test series* decomposition. As filter parameters, the optimal window width and standard deviation resulting from the 4-fold cross validation using the manually labeled *gait test series* were utilized.

To assess the results independently from the *gait test series* detection results, only detected strides included in the ground truth were considered in this evaluation, whereas the classification results of false positive and false negative strides were ignored. However, strides of sequences rejected by the *gait test series* decomposition algorithm were still collected in a separate rejection class.

Chapter 6

Results

In the following chapter, the resulting threshold parameters and performance values of the *gait test series* detection (section 6.1) and the *gait test series* decomposition (section 6.2) are presented.

6.1 Gait Test Series Detection

6.1.1 Statistical Threshold Adjustment

Thresholds were calculated for minimum and maximum duration of a *gait test series* based on the mean duration of single 4x10MWTs and their standard deviation, as described earlier (section 4.2.3). Therefore, mean and standard deviation of all labeled 4x10MWTs were calculated separated by speed level (Table 6.1).

Velocity Level	**Mean [s]**	**SD [s]**
Slow	65.94	27.85
Preferred	53.48	19.24
Fast	43.52	17.08

Table 6.1: **Average Duration of 4x10MWTs.**

The resulting thresholds were $\kappa_{min} \approx 43.52\,\mathrm{s} - 2 \cdot 17.08\,\mathrm{s} = 9.36\,\mathrm{s}$ and $\kappa_{max} \approx 65.94\,\mathrm{s} + 2 \cdot 27.85\,\mathrm{s} = 93.79\,\mathrm{s}$. This range applies to all available manually labeled data except for one outlier (Figure 6.1).
Furthermore, the mean duration results caused the parameter *distance* for the peak detection algorithm (section 4.2.3) being set to $(53.48\,\mathrm{s} - 2 \cdot 19.24\,\mathrm{s}) \cdot \frac{1}{4} \approx 3.25\,\mathrm{s}$.

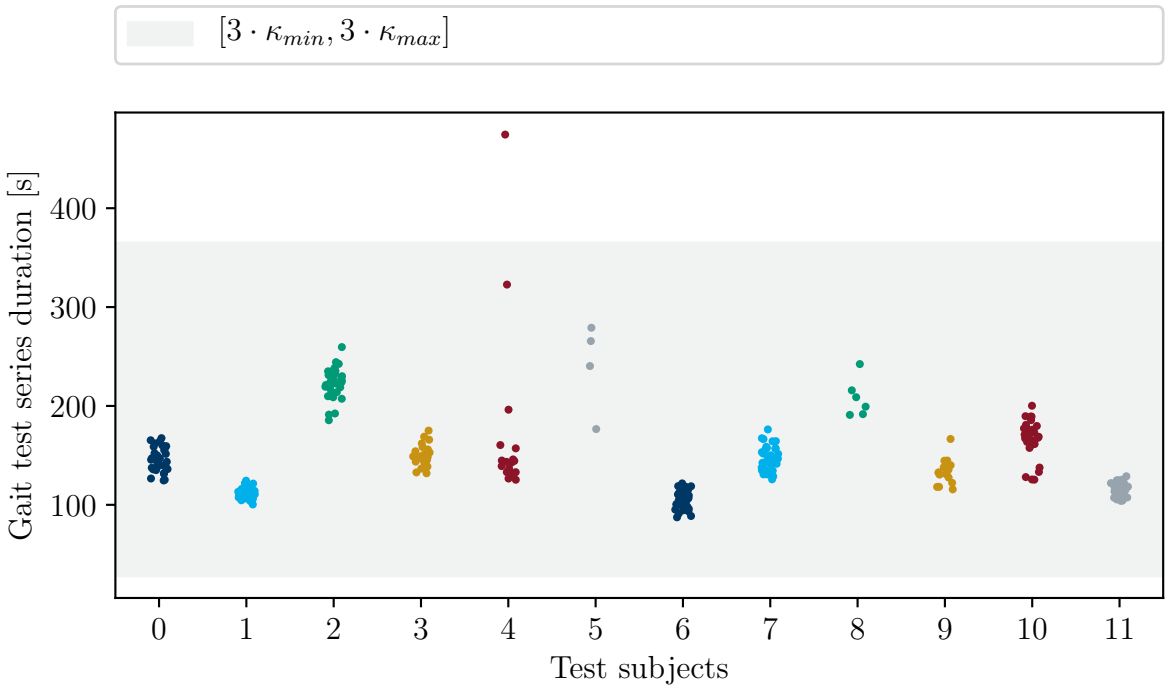

Figure 6.1: **Length Range for Gait Test Candidates.** Each data point represents the duration of a manually labeled compliant *gait test series*. The grey shaded area is the part covered by $[3 \cdot \kappa_{min}, 3 \cdot \kappa_{max}]$.

6.1.2 Postprocessing Parameters

Single-dimensional Dynamic Time Warping

A 4-fold cross validation was performed to optimize the cost threshold Θ and the turning verification range $[\eta_{min}, \eta_{max}]$. The optimal parameters were defined based on the average performance values of the training datasets of this cross validation (Figure 6.2). On average, The best stride-based *F1-score* was achieved for $\Theta = 2.4$, $\eta_{min} = 11$, and $\eta_{max} = 13$. The *sensitivity* was maximized for larger cost thresholds and a broader or no turning peak range. The *precision* increased for a lower cost threshold and a more narrow turning peak range.

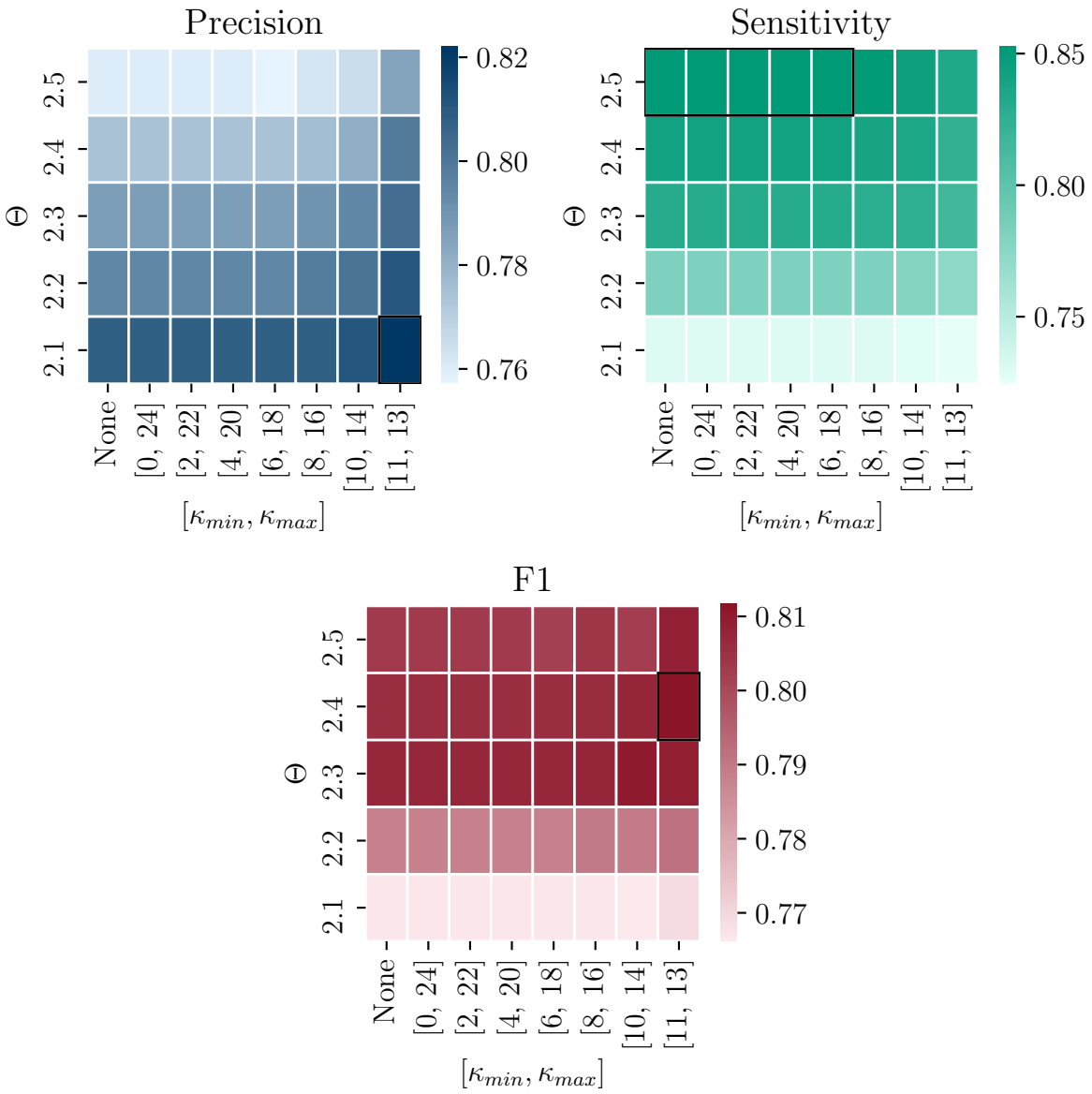

Figure 6.2: **ssDTW training performance.** Mean values of *precision*, *sensitivity*, and *F1-score* during training for different postprocessing parameter combinations averaged over training datasets of a 4-fold cross validation. The maximum value is reached in the black framed areas.

For each fold, the best parameter combination was estimated separately and then applied to the respective test data. Resulting optimal parameter combinations ranged from $\Theta \in \{2.3, 2.4\}$ and $[\eta_{min}, \eta_{max}] \in \{[10, 14], [11, 13]\}$. The mean stride-based *F1-score* of this test data was $80.5\% \pm 13.8\%$ for all data (Table 6.2). The area-based *F1-score* was 92.8%, and the test-based *F1-score* yielded to 51.6%. For the compliant data, performance increased for all evaluation methods. The stride-based *F1-score* resulted in 87.6%, and the area-based yielded to 97.6%.

Evaluation	All 4x10MWT [%]	Compliant 4x10MWT [%]
Stride-based	80.5 (13.8)	87.6 (3.3)
Test-based	51.6 (12.9)	61.7 (15.0)
Area-based	92.8 (3.8)	97.6 (2.0)

Table 6.2: **ssDTW test performance.** *F1-scores* for a 4-fold cross validation of the ssDTW-based algorithm as mean [%] (SD [%]) for the complete as well as the compliant dataset for all evaluation methods.

Multi-dimensional Dynamic Time Warping

Similar to the ssDTW optimization procedure, a 4-fold cross validation was performed for the msDTW-based algorithm. In addition to the optimization of cost threshold Θ and turning verification range $[\eta_{min}, \eta_{max}]$, two different procedures for estimating the amount of turnings, namely peak detection and turning detection, were compared. The latter turning verification method was only applicable for msDTW, since it requires the lower back sensor signal. Therefore, the two distinct turning verification methods were only compared for msDTW. The optimal parameters achieving the largest stride-based *F1-score* when using turning detection according to El-Gohary [EG14] were $\Theta = 5.3$, $\eta_{min} = 8$, and $\eta_{max} = 16$ (Figure 6.3). Using peak counting, the optimal values were achieved for $\Theta = 5.3$. Regarding the turning verification range, there was no improvement observed when a turning verification threshold was used. However, for $\eta_{min} \leq 8$ and $\eta_{max} \geq 16$, the stride-based *F1-score* did not deteriorate (Figure 6.4). Similar to the observations made for ssDTW, the *precision* was maximized for lower Θ and a more narrow turning peak range for both turning verification methods, whereas the *sensitivity* reached its maximum for a higher cost threshold and a broader peak range than the optimal values (Figure 6.3, Figure 6.4).

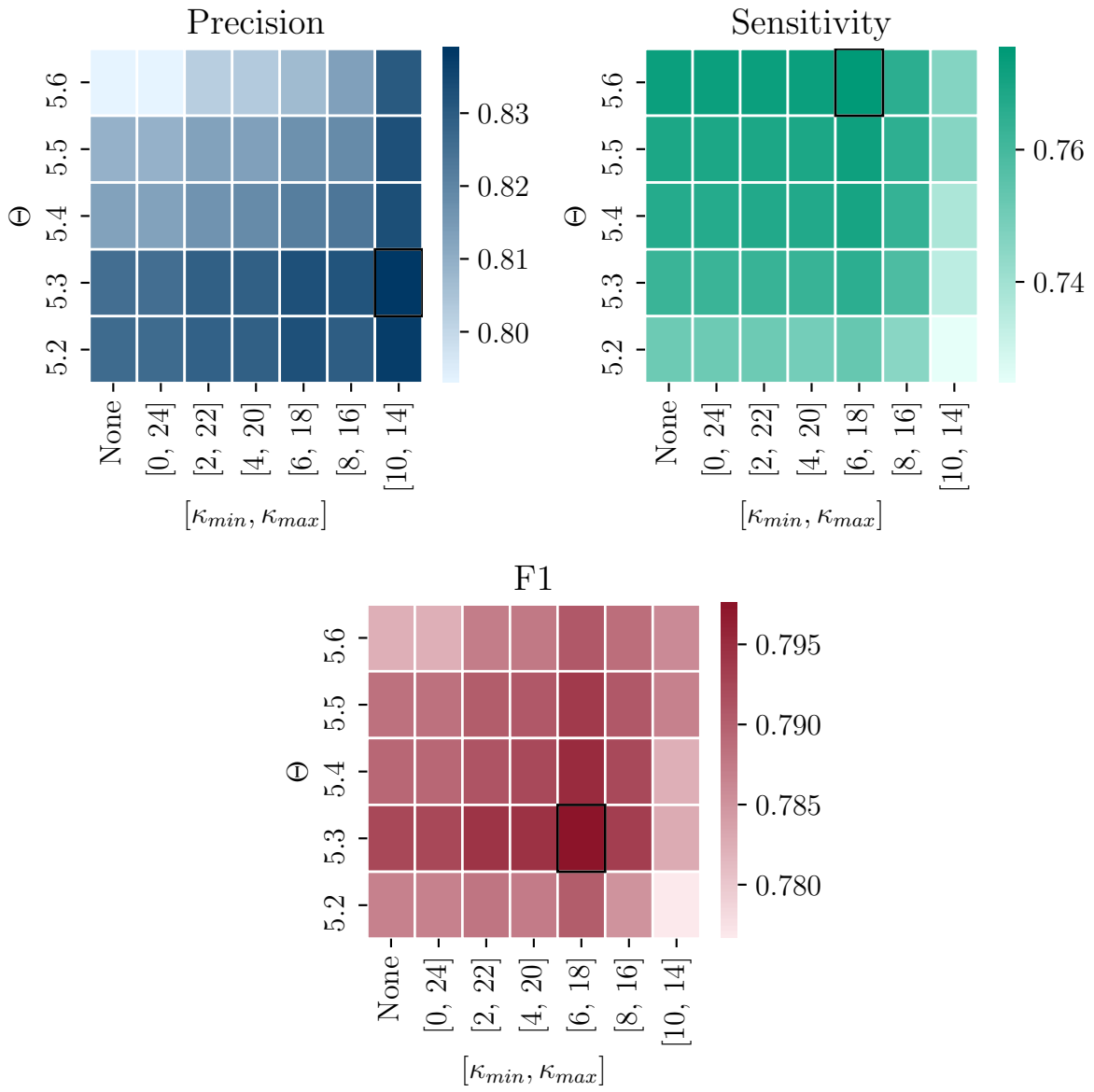

Figure 6.3: **msDTW Training Performance with Turning Detection.** Mean values of *precision*, *sensitivity*, and *F1-score* during training for different postprocessing parameter combinations averaged over training datasets of a 4-fold cross validation. Turning detection was used as turning verification method. The maximum value is reached in the black framed areas.

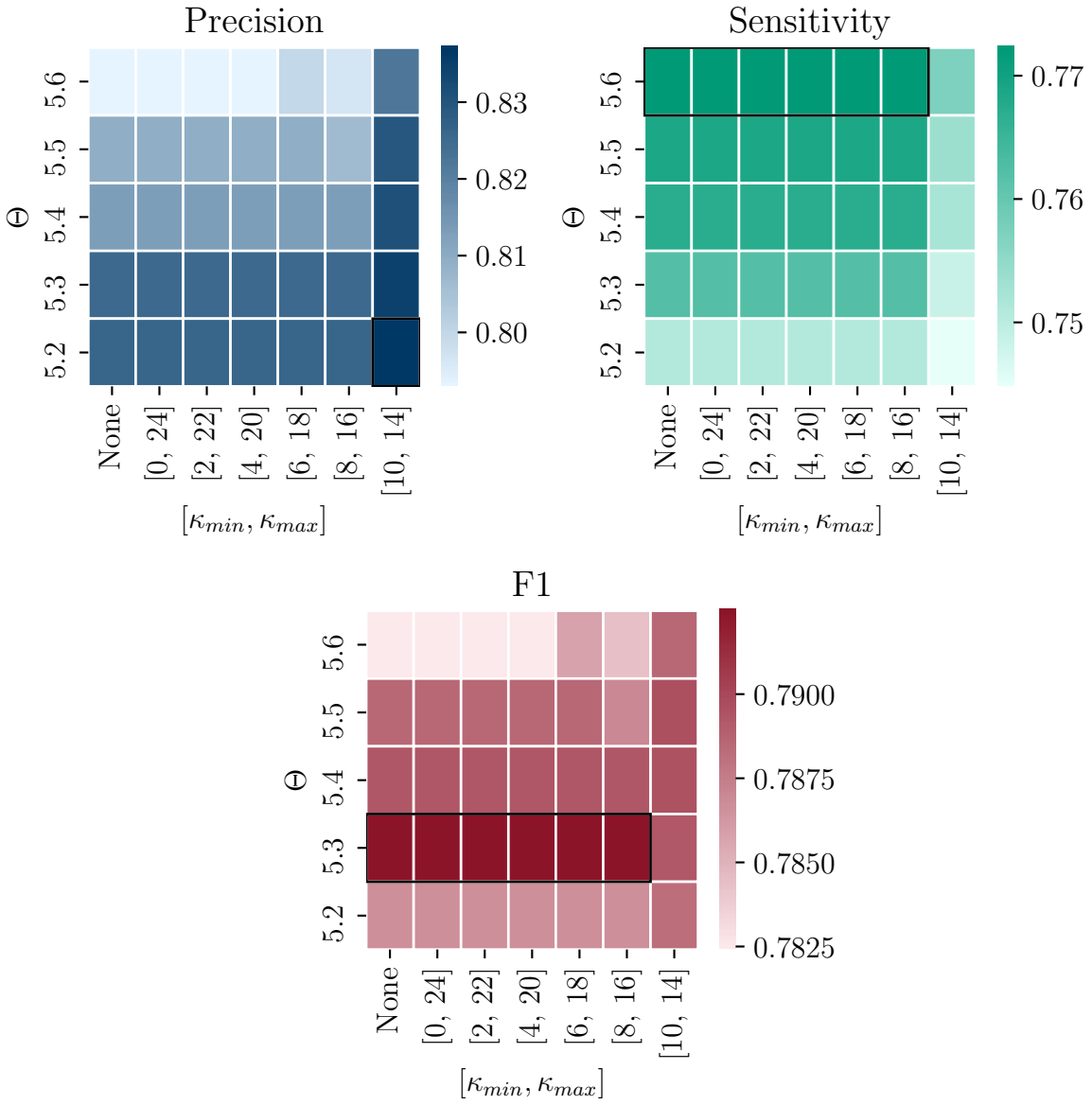

Figure 6.4: **msDTW Training Performance with Peak Detection.** Mean values of *precision*, *sensitivity*, and *F1-score* during training for different postprocessing parameter combinations averaged over training datasets of a 4-fold cross validation. Peak detection was used as turning verification method. The maximum value is reached in the black framed areas.

Again, the best parameter combinations for each of the turning verification methods were estimated and validated on the respective test set of the unrefined data. The mean stride-based *F1-score* for was $78.1\% \pm 15.4\%$ (Table 6.3). For each fold, the best parameter combination was achieved using peak detection as turning verification method. The optimal cost threshold for each fold varied between $\Theta = 5.3$ and $\Theta = 5.4$, and the turning peak range was in the range $[\eta_{min}, \eta_{max}] \in \{[6, 18], [8, 16], [10, 14]\}$. The area-based *F1-score* was estimated to 91.0%, and the test-based *F1-score* resulted in 67.2%. For the compliant dataset, the stride-based *F1-score* was 77.4%.
Additionally, for each fold, the best parameters achieved by the respective other turning verification method, i.e. turning detection, were tested separately resulting in a stride-based *F1-score* of 66.5% for all and 76.9% for compliant data (Table 6.3).

Evaluation	Peak Detection		Turning Detection	
	All 4x10MWT [%]	**Compliant 4x10MWT [%]**	**All 4x10MWT [%]**	**Compliant 4x10MWT [%]**
Stride-based	78.1 (15.4)	85.2 (4.5)	77.3 (14.5)	84.5 (4.9)
Test-based	67.2 (13.5)	77.4 (11.6)	66.5 (13.8)	76.9 (12.2)
Area-based	91.3 (5.2)	95.4 (1.9)	90.9 (4.0)	94.7 (2.6)

Table 6.3: **msDTW Test Performance.** Stride-based *F1-scores* for the 4-fold cross validation of the ssDTW-based algorithm as mean [%] (SD [%]) for the complete as well as compliant dataset for all evaluation methods. The results of both turning verification methods are displayed separately.

6.2 Gait Test Series Decomposition

The performance based on the foot sensor strides of the *gait test series* decomposition was optimized using a 4-fold cross validation. For all training sets, the window width $w = 30$ and hence the standard deviation $\sigma = w/6 = 5$ were determined to achieve the best stride-based *F1-score*. When evaluating the *gait test series* decomposition based on the ground truth *gait test series*, summed up over all test sets, a total of 51502 4x10MWT strides was investigated (Figure 6.5). 218 strides (0.4%) were rejected since the algorithm was not able to detect suitable peaks for the corresponding 4x10MWTs. The mean performance scores averaged over all classes and test folds are 91.7% for *precision*, 91.4% for *sensitivity*, and 91.5% for *F1-score* (Table 6.4).

When considering the *gait test series* labels set by the *gait test series* detection method with the best performance, i.e. the ssDTW approach, as input, 46147 valid strides were included, where 382 (0.8%) of them were rejected by the algorithm. The resulting performance values were a *precision* of 88.1%, a *sensitivity* of 87.9%, and an *F1-score* of 88.0% (Figure 6.6, Table 6.5).

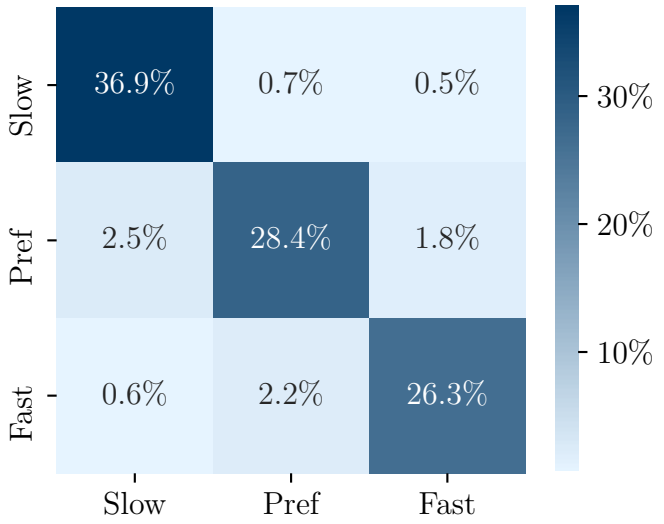

Figure 6.5: **Confusion Matrix for Gait Test Series Decomposition Test Performance based on Ground Truth Labels.** Sum of classification results of all not rejected strides of 4-fold cross validation test datasets.

Velocity Level	Precision [%]	Sensitivity [%]	F1-Score [%]
Slow	92.4 (3.4)	96.7 (0.4)	94.5 (1.8)
Preferred	90.9 (4.4)	86.8 (4.2)	88.8 (4.1)
Fast	91.8 (5.6)	90.6 (6.8)	91.2 (6.1)
Total	91.7 (0.6)	91.4 (4.1)	91.5 (2.3)

Table 6.4: **Gait Test Series Decomposition Test Performance based on Ground Truth Labels.** *Sensitivity*, *precision*, and *F1-scores* for a 4-fold cross validation of the *gait test series* decomposition algorithm as mean [%] (SD [%]) separated by velocity level classes and in total.

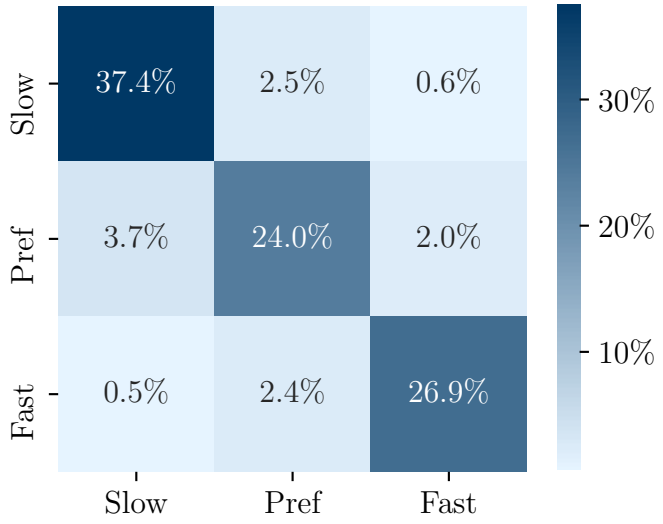

Figure 6.6: **Confusion Matrix for Gait Test Series Decomposition Test Performance based on Automatically Detected Labels.** Classification results of all not rejected strides included in the ground truth.

Velocity Level	Precision [%]	Sensitivity [%]	F1-Score [%]
Slow	89.9	92.4	91.2
Preferred	83.2	80.8	82.0
Fast	91.2	90.4	90.8
Total	88.1	87.9	88.0

Table 6.5: **Gait Test Series Decomposition Test Performance based on Automatically Detected Labels.** *Sensitivity*, *precision*, and *F1-scores* of the *gait test series* decomposition algorithm as mean [%] separated by velocity level classes and in total.

Chapter 7

Discussion

In this thesis, an improved pipeline for detecting 4x10MWTs in free-living gait recordings was developed and evaluated. A major problem when investigating standardized gait tests performed at home without supervision were slight changes of the execution compared to the predefined protocol. For instance, resting periods between individual 4x10MWTs were omitted in many cases, and the order and number of repetitions of succeeding 4x10MWTs varied between and within patients. Furthermore, the test specifications could not be met under the same strict conditions as in the clinic, because the execution of the 4x10MWTs was strongly depending on patients' housing conditions.

All these factors make automated detection of standardized tests using pattern recognition challenging, and prevent existing algorithms that have been validated on short, supervised recordings from working as expected. Hence, addressing the two main research questions of this thesis, a pipeline consisting of two parts was developed and evaluated.

The first aim was to detect *gait test series* in full-day recordings realized in the first part of the presented pipeline. It has been shown that the developed algorithm was suitable to find *gait test series* in free-living recordings. However, the accuracy of the detected labels could still be improved (section 7.1).

The second goal was the decomposition of single 4x10MWTs and the assignment of the corresponding gait speed levels, which was implemented by the second pipeline part. The performed experiments have shown that the proposed algorithm for this task produced reliable results (section 7.2). Strengths and weaknesses of both pipeline parts are discussed more in detail in the following.

7.1 Gait Test Series Detection

The main objective of the first part of the pipeline was to detect series of 4x10MWTs within free-living recordings. In the following, differences to the implementation of Fischer et al. [Fis20] are emphasized and discussed (section 7.1.1). Furthermore, the evaluation results are analyzed (section 7.1.2). Finally, persisting problems and limitations of the algorithm are outlined (section 7.1.3).

7.1.1 Comparison with Reference Algorithm

The *gait test series* detection was adapted from Fischer et al. [Fis20], who proposed a DTW-based algorithm for 4x10MWT detection applicable to clinical recordings including different supervised gait tests. The main advantage of DTW is that it is an intuitive approach to compare time series independent from their scaling [Mye81]. This enables the processing of distinct sessions despite various walking path lengths and gait velocities. However, the algorithm of Fischer et al. did not succeed on semi-standardized 4x10MWTs in free-living environments, and therefore required an adaption. As discussed in the following, several reasons were analyzed and the required improvements implemented.

For both algorithms, a gait sequence detection was applied first to extract relevant subsequences of a complete dataset, which has proven to be robust for clinical applications [Ull20]. A problem already reported was that if resting periods between single 4x10MWTs were too small, all 4x10MWTs were concatenated to one sequence [Fis20]. As the importance of breaks between the single tests was not clearly emphasized in the *FallRiskPD* study protocol, resting periods conducted by the patients were consistently performed too short or not at all. Additionally, if a patient's gait was shuffling or the 4x10MWT was suddenly interrupted for a short period of time, e.g. due to FOG, a test was often split into several small gait sequences. Since the gait behavior tends to worsen at home compared to the clinic [Bro16], this problem was further aggravated for the free-living data used in this work. Therefore, since all gait sequences were investigated separately, it was decided to merge short successive gait sequences to provide unintended splitting of 4x10MWTs. Consequently, due to negligible resting periods, a *gait test series* consisting of three 4x10MWTs instead of one was expected in a gait sequence.

The preprocessing pipeline for the gyroscope signal prior to DTW developed and validated by Fischer et al. [Fis20] was adopted mostly unchanged. However, to enable processing of large data amounts generated by free-living recordings in a reasonable amount of time, downsampling of the data was added. Besides the increase of processing speed this additional step did not influence the performance of the algorithm.

Even though the separate detection of individual 4x10MWTs with a three- or four-turn template appeared more intuitive in the first place, it turned out to be error-prone with regards to detected false positives and was therefore discarded. Hence, the template for the DTW step was chosen accordingly to consist of three consecutive 4x10MWTs. Fischer et al. created several templates automatically and estimated the best one using a validation dataset. They recommended to average five 4x10MWTs using randomly chosen patients [Fis20]. Again, despite of resampling for standardization, for templates created from manually labeled unsupervised 4x10MWTs, unwanted side effects such as double peaks occurred. Thus, to avoid creating several templates, followed by either manual inspection or additional evaluation on test data to estimate their quality, in this thesis, a semi-artificially created template was used. It was constructed by extracting a clear turning peak from averaged 4x10MWTs and repeating this peak periodically. Because the template was composed manually, there was no evaluation bias caused by a template created from known data. Relevant data of turnings was equally recorded by several sensors at both feet as well as the lower back. For the sake of simplicity, and because their data was not synchronized, Fischer et al. limited their considerations to the left foot sensor [Fis20]. In contrast, in this thesis, all available sensors were investigated with regard to the question whether a combination of synchronized data recorded with several sensors can provide information gain.

A postprocessing part of the algorithm was adapted focusing on determining *gait test series* interval borders as precisely as possible. Fischer et al. stated that for their pipeline, the borders estimated for the 4x10MWTs based on the cheapest warping path of the DTW turned out to be imprecise [Fis20]. They assumed the gait sequence borders to be identical with the 4x10MWT borders, i.e. that one gait sequence only contained a 4x10MWT and no additional gait signal. This assumption succeeded for clinical recordings, since patients were clearly instructed to rest directly before and after a 4x10MWT. Accordingly, their postprocessing pipeline was focused on the complete gait sequence including turning peak checking and stride counting [Fis20]. For the application to free-living recordings, however, this assumption is not feasible, since resting periods were not only often missing between subsequent 4x10MWTs, but also directly before and after a *gait test series*. Possible reasons are that patients were walking around to prepare their test setup, including an obstacle-free straight walkway, a chair for the TUG, and the smartphone for operating the annotation app. Afterwards, they had the instruction to perform a TUG test, which was consequently often included in the same gait sequence. To counteract this problem, a postprocessing pipeline was developed focusing on warping path borders predicted by DTW. To detect the best possible result, all *gait test series* borders predicted by local minima of the cost function underneath a certain cost threshold were undertaken turning and length checking leading to improved results in many examples (Figure 7.1).

Fischer et al. reported a test-based *F1-score* of 86.7% [Fis20]. The best test-based result for the presented algorithm was 67.2%, which is considerably lower. However, this difference can be explained considering the fundamental differences of the underlying datasets used for the evaluation. Within the free-living data used for this thesis, as mentioned above, gait movement was often performed directly before and after the *gait test series*. As this movement most happened within the scope of preparing the test setup or performing a TUG, it included turnings with a high probability. As the proposed algorithm focuses only on turnings, it can not distinguish between previous or succeeding movement including turns, and *gait test series* (Figure 7.2). The decision, which turns to match, is consequently made based on the DTW cost only, not considering further criteria, e.g. consistent walking between all turnings.

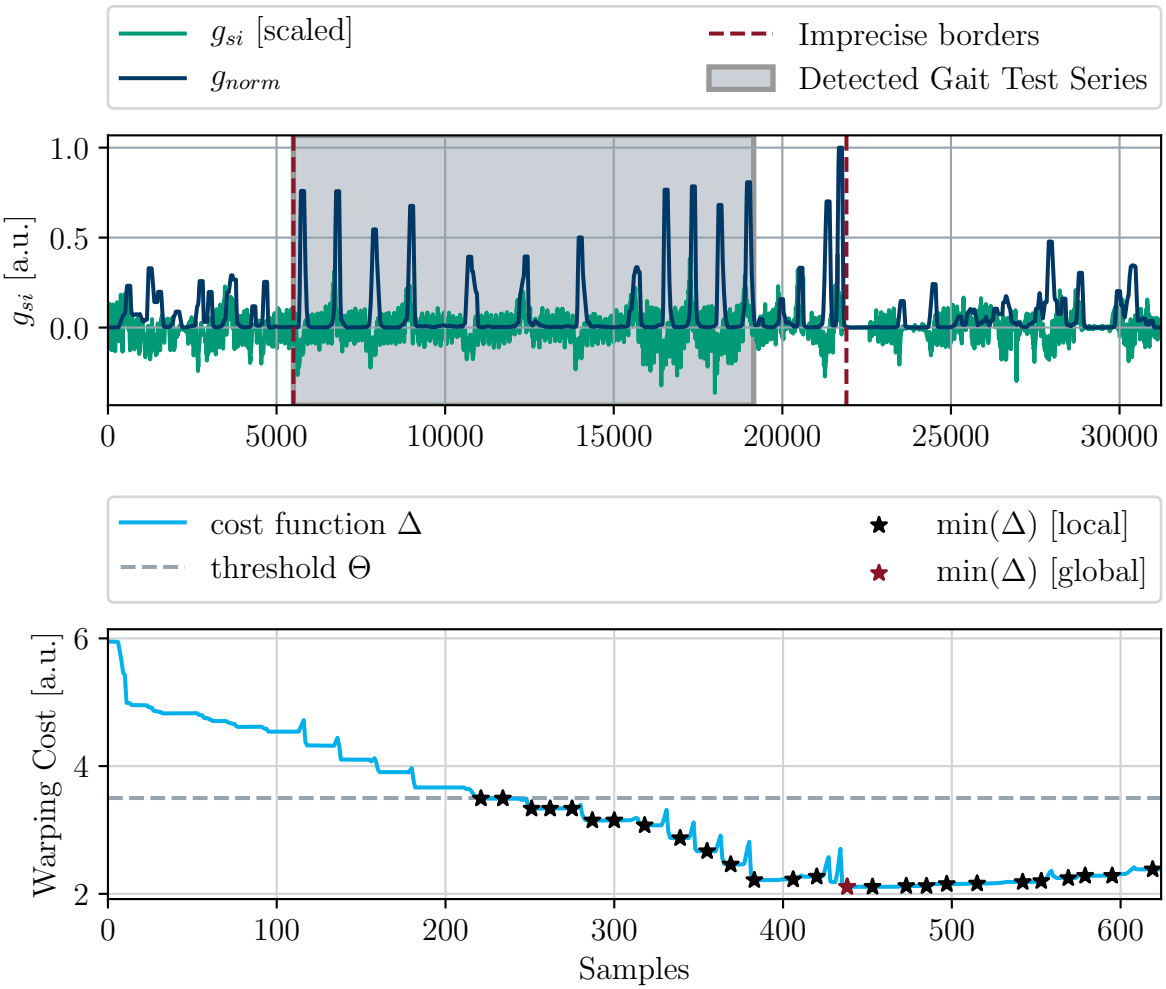

Figure 7.1: **Improved Labeling.** The global minimum of the cost function and the corresponding interval borders marked in red did not always belong to the most accurate label. Usage of further postprocessing lead to the grey-shaded label and thus increased accuracy.

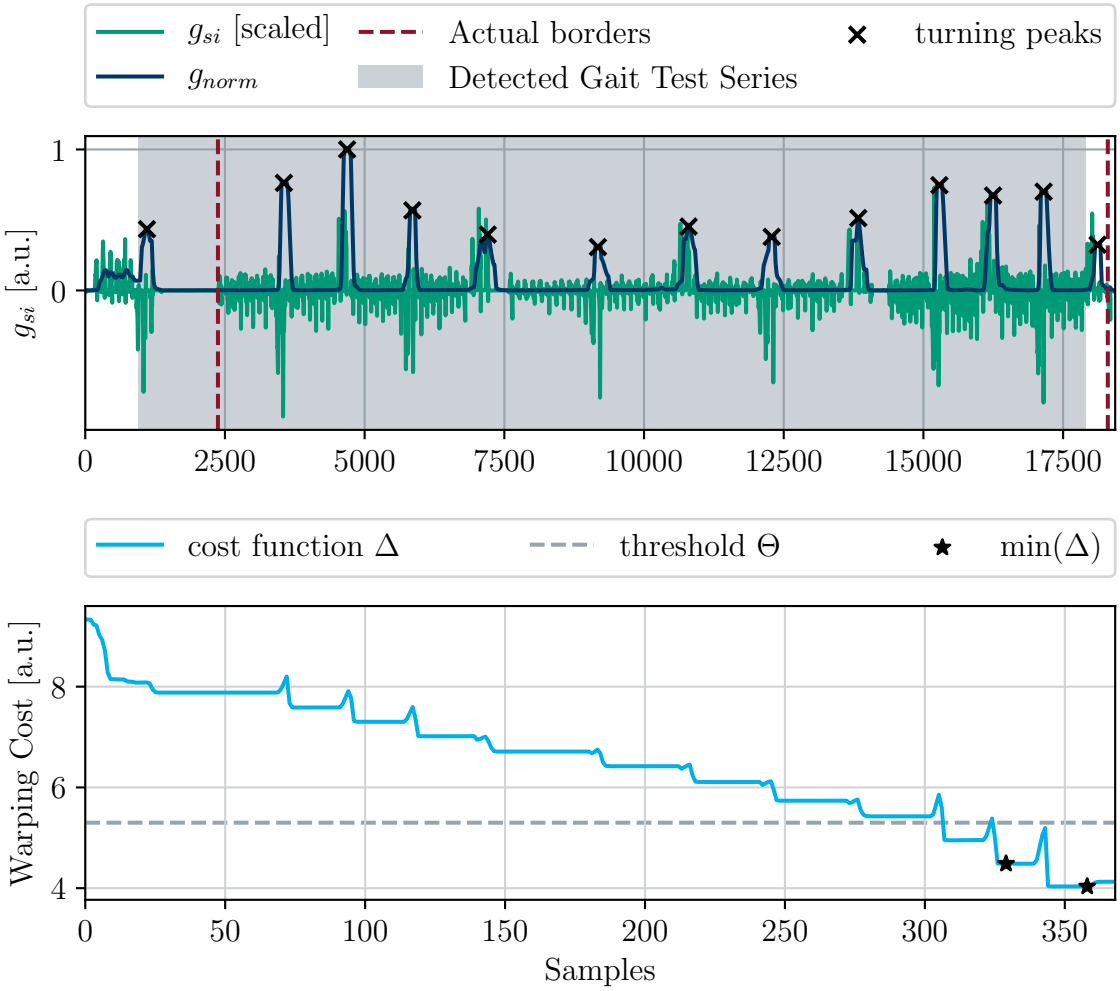

Figure 7.2: **Imprecise Warping Path Borders.** In this example, the patient was turning around directly before the start of the *gait test series*. Still, the match including this previous turn fulfilled all criteria, achieved a local minimum point in the cost function, and was therefore falsely chosen as the best match.

7.1.2 Performance

Two different DTW approaches were compared in the evaluation. One one hand, there was ssDTW processing all sensor positions separately, and on the other hand, msDTW combining all sensor recordings to a 3-d signal performing only one DTW step for each gait sequence.
The optimal postprocessing hyperparameters were grid-searched during the evaluation for both approaches. For ssDTW, the results were relatively consistent for every test set. For msDTW, the results for the cost threshold Θ were also consistent. The turning verification range, however, only had a minor impact on the msDTW-based training results of the 4-fold cross validation. This caused the resulting turning threshold parameters to vary between every fold while causing only minor performance differences. Especially for the peak detection based turning verification, the impact of a turning verification threshold was negligible. For the turning detection based turning verification, a broad turning threshold turned out to be the best choice. None of the turning verification methods performed considerably better than the other. Therefore, it is concluded that peak detection is the more convenient method, as it is widely applicable for ssDTW as well as msDTW without requiring a hip mounted sensor. Furthermore, the calculation is more intuitive and requires less computation time.
When considering the stride-based *F1-scores* of the 4-fold cross validation defined as main performance criterion, the performance of both approaches is comparable. For the test set, the difference in *F1-score* of ssDTW and msDTW is 2.4% and hence considerably low, where ssDTW performed slightly better than msDTW with an average *F1-score* of 80.5%.
For the test-based evaluation, performance drops for both DTW approaches. However, msDTW appears to be more robust in terms of detecting precise borders for *gait test series* showing a 15.6% higher test-based *F1-score* (67.2%) than ssDTW. In comparison, the stride-based evaluation is considered as a more valuable evaluation method than the test-based evaluation. Reasons are in the first place, that the *gait test series* decomposition as well as the gait parameter calculation for analysis of the 4x10MWTs from a clinical point of view are stride-based. The test-based evaluation on the other side refers to the borders of manually set labels. These labels, however, where set in a manner that the longest possible resting periods before and after each test were included, which have no relevance for the physiological evaluation.
Another evaluation method considered was the area-based evaluation. With regard to the aim of not yet achieving fully automatic detection of 4x10MWTs, but simply creating a smart labeling tool detecting areas of interest within full day recordings, this score can be investigated. With a maximum *F1-score* of 92.8% for ssDTW and even 97.6% when considering the compliant test data only, it can be concluded that the presented algorithm provides a reliable basis for such a labeling tool. Again, the msDTW is only slightly outperformed with an average *F1-score* 1.5%

beneath ssDTW. When integrated to a graphical user interface, manual refinement of labels might still be required, however, patients are relieved and technical errors prevented as smartphone annotation setting or diary keeping becomes obsolete.
Another aspect that was observed is the large standard deviation of the performance scores caused by large performance differences between folds. This phenomenon occurs more pronounced for the evaluation of the complete dataset instead of the compliant dataset. As each test set consisted of data of three patients, the impact of each patient's *gait test series* performance was relatively large, resulting in worse results if the patient was consistently executing gait tests different than specified in the study protocol (Figure 7.3).
In general, it was found that ssDTW and msDTW perform approximately equal. Only for the most intuitive, but physiologically irrelevant test-based evaluation approach, msDTW considerably outperformed ssDTW. However, ssDTW has the main advantage that it is applicable to a dataset even if recordings of single sensors are not available, which is a frequent issue especially with regards to long-term monitoring. Possible reasons are uncharged or broken sensors or issues during data transfer. As soon as there is a recording of one sensor available, the ssDTW approach can be used without the need for further hyperparameter adjustment. Although, it needs to be emphasized that in the evaluation performed here, all sensors of all datasets were used for the ssDTW performance estimation. Especially in terms of *sensitivity*, the results may worsen when not all sensor positions are used.

7.1.3 Limitations and Weaknesses

One major problem of the proposed algorithm is that although it is supposed to achieve precise stride-based results, it is not working based on stride information, but on turning information. This causes results to be inaccurate especially in terms of *precision*, when further turnings are performed directly before or after a *gait test series*. Further criteria might be required to prevent false decisions in this case. One tested approach was to add a further gyroscope axis as input of the DTW to additionally match the steady stride pattern accompanying the *gait test series*. However, this was found not to be feasible, since on one hand, the amount of strides required for one 4x10MWT fluctuates strongly inter- and intra-patient-wise. On the other hand, the stride information could not be downsampled to the same extent as the turning peak signal and would therefore cause a strong runtime increase of the DTW step.

Another persisting limitation of a template with a static amount of turning peaks is that it is not applicable to non-compliant tests that were performed with a reduced number of turnings. Non-compliant tests with an increased number of turnings, however, were detected, but the defined borders only covered a subsequence of the test series and would thus might require manual refinement.

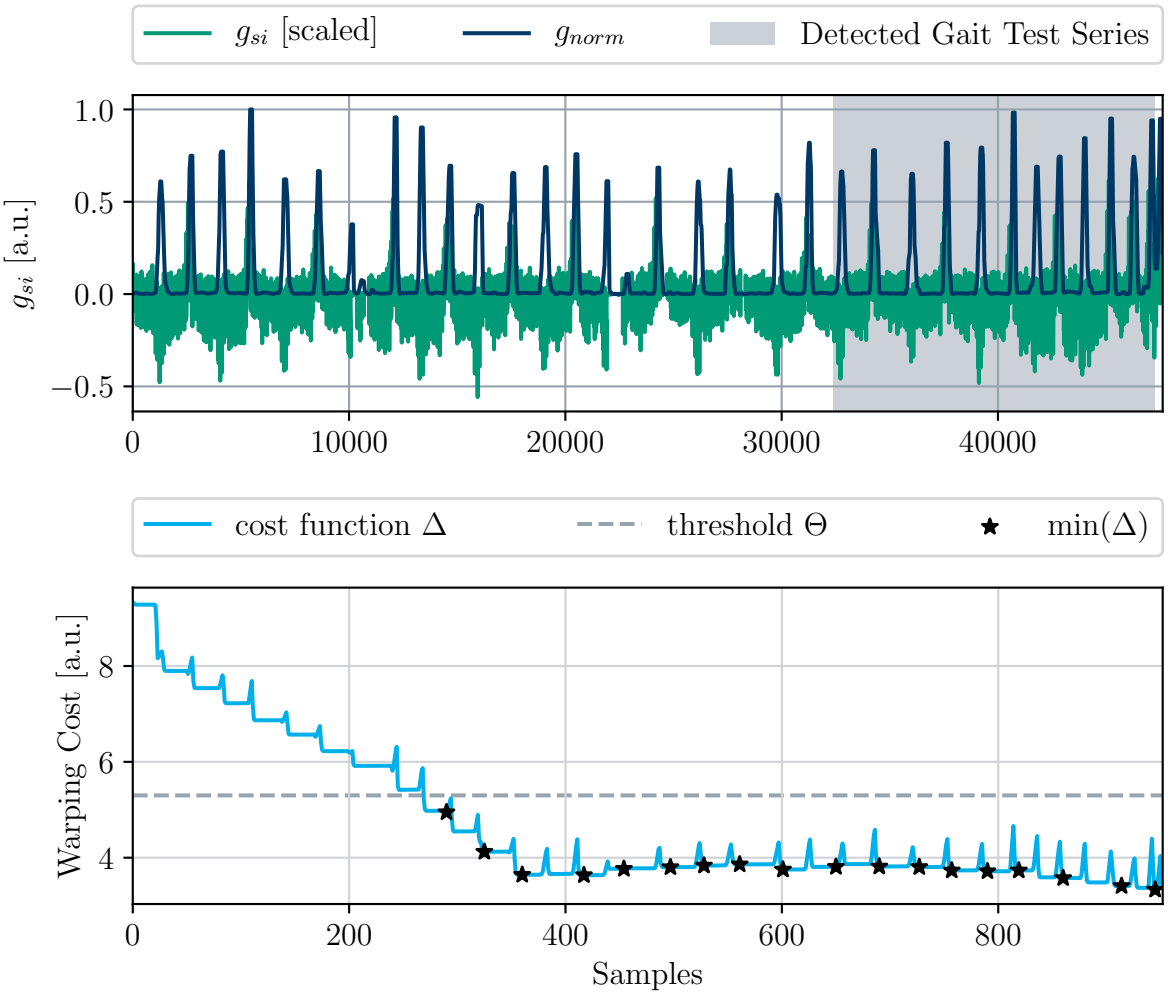

Figure 7.3: **Non-compliant Gait Test Series.** In the depicted example, too many repetitions of 4x10MWTs were performed in a *gait test series*. Thus, only a part of it was recognized by the proposed algorithm.

7.2 Gait Test Series Decomposition

In the second part of the proposed pipeline, the beginning and end of each 4x10MWT were defined and corresponding speed levels were assigned. As input sequences for parameter optimization, first, the ground truth *gait test series* labels were used. Afterwards, the *gait test series* decomposition was tested using the automatically generated labels resulting from the *gait test series* detection algorithm. In the following, the result of both methods is compared and discussed (section 7.2.1). Afterwards, the still existing weaknesses of the decomposition algorithm are described (section 7.2.2).

7.2.1 Input-dependent Comparison of the Results

The proposed algorithm exploits the fact that the gait speed level varied between every 4x10MWT of a *gait test series*. That means, steep changes of the gait velocity over time indicated speed translations and thus 4x10MWT borders. For calculating the gait velocity function, strides were segmented and the velocity for each stride was calculated using already proven algorithms [Bar15, Ram15]. Since the stride velocity fluctuates within each 4x10MWT due to acceleration, deceleration, and turning strides as well as shuffling gait, the velocity function was smoothed to eliminate the impact of small variations. The steepest changes of this smoothed velocity function were then determined based on the maximal values of the derivative of the velocity function. The corresponding strides were defined as borders for the respective 4x10MWTs. The speed levels of all 4x10MWTs were assigned using the velocity function again by only comparing the mean velocities of all 4x10MWTs relative to each other. Consequently, there were no fixed velocity thresholds required allowing the algorithm to work despite the fact, that the self-selected gait speeds can considerably differ between patients.

The pipeline showed a reliable performance with an *F1-score* of 91.7% averaged for all test data when using the ground truth *gait test series* as input. The result based on the automatically detected *gait test series* was only slightly lower with an *F1-score* of 88.0%.

However, the overall amount of investigated strides was lower for the automatically detected sequence based evaluation, since wrongly classified strides were excluded in order to eliminate the impact of the classification result of the previous pipeline step.

For both input variants, when observing the result for the separate classes, it is noticeable that *sensitivity* and *precision* of the preferred speed class are lower than those of the remaining classes. This is due to the fact that considering the velocity, preferred speed is situated between the slow and fast speed classes and therefore, misclassification occurs in both directions. In contrast, strides

from fast to slow and vice versa are rarely assigned to the respective other class. They are caused by imprecisely detected edges resulting in wrongly assigned border steps.
A small percentage of *gait test series*, 0.4% based on the ground truth sequences and 0.8% based on the detected sequences, was excluded from the evaluation, because no maxima fulfilling all constraints were detected in the smoothed acceleration function. This was mainly caused by too few detected strides (Figure 7.4) and might require and improvement of the stride segmentation algorithm. The higher value for the automatically detected *gait test series* as input additionally was a consequence of imprecise borders cutting off relevant strides.

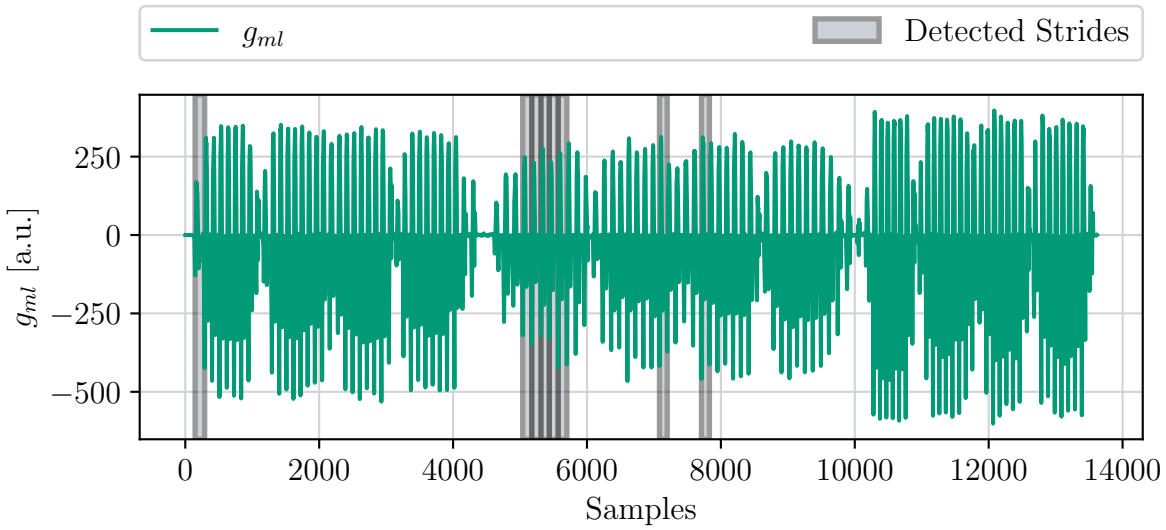

Figure 7.4: **Stride List Refinement Issue.** For this example, the majority of strides was mistakenly excluded during the event-based stride list refinement. Therefore, no meaningful analysis of the velocity function could be performed and the displayed *gait test series* was rejected by the decomposition algorithm.

7.2.2 Limitations and Weaknesses

The evaluation was performed for the compliant *gait test series* data only since for the proposed algorithm, it was assumed that exactly three subsequent 4x10MWTs were included within the observed *gait test series*. The algorithm then detects the two points with the maximal acceleration within one signal and therefore only two speed level translation borders. That means, *gait test series* non-compliant for the reason that too many or too few 4x10MWTs were included

could not be analyzed. However, *gait test series* that are non-compliant because they included a different amount of performed 10 meters walkways, e.g. 2x10MWTs or 6x10MWTs, should be decomposable using the proposed algorithm.
Furthermore, the proposed algorithm has difficulties, if the speed changes abruptly within a 4x10MWT, or if the gait speed does not considerably change within a *gait test series* at all. However, both phenomenons may occur in PD patients with advanced stage of disease caused by FOG [Jan08], or difficulties in gait speed adaption [Ste08].
The classification results based on the automatically detected *gait test series* are promising and indicate that the overall detection pipeline consisting of *gait test series* detection and *gait test series* decomposition is functioning well. However, the classification results of the *gait test series* decomposition are not to be interpreted as overall performance of the detection pipeline. This is due to the fact that falsely classified strides from the *gait test series* detection were excluded from the decomposition evaluation.

Chapter 8

Conclusion and Outlook

The goal of this thesis was to develop a method for automatic detection of semi-standardized 4x10MWTs in free-living inertial sensor recordings.
At first, *gait test series*, i.e. subsequently performed 4x10MWTs of different gait velocities, were detected and extracted. The algorithm for this pipeline step achieved a mean *F1-score* of 80.5% for stride-based detection of *gait test series* when using the recordings of all three inertial sensors, performing ssDTW, and combining the resulting matches of all sensors.
Second, extracted *gait test series* were decomposed into single 4x10MWTs and the corresponding gait speed levels were assigned. Based on the ground truth *gait test series* labels, 90.7% of all included strides were assigned to the right speed level and therefore to the correct 4x10MWT. Results did not considerably worsen when the validation was performed for the automatically detected labels.
When considering the complete pipeline, it needs to be noticed that the majority of 4x10MWTs is recognized by the proposed algorithm. However, information is lost on one hand by completely missed or falsely detected *gait test series* as well as false negative and false positive detected strides due to imprecise *gait test series* borders. On the other hand, results can be falsified by strides incorrectly assigned to a speed level caused by the second pipeline part. Thus, for assessing whether the results are sufficient to enable the usage of the proposed pipeline as base for a tool for automatic evaluation of 4x10MWTs in free-living gait recordings, further investigations are required. The next step will be to calculate spatio-temporal gait parameters for all patients based on the ground truth labels and the automatically detected labels, respectively. A comparison of both results will provide information on whether and how much clinically relevant information is lost when using the proposed detection algorithm.

With regard to the study design, there are two decisive aspects considerably influencing the outcome of the detection pipeline. On one hand, sufficient resting periods performed before and after every 4x10MWT have a major impact on the detection accuracy. On the other hand, the compliance of a *gait test series*, i. e. that it consists of exactly three 4x10MWTs, turned out to be important. These two aspects need to be communicated more clearly to participants of future studies. When patients have become accustomed to this protocol in long-term monitoring after initial uncertainties, the algorithm may be able to provide more robust results.
However, it needs to be emphasized that all proposed methods only work for series of 4x10MWTs, as they form a characteristic pattern of turnings, which is easily recognizable even in full-day recordings. The search for single 4x10MWTs or shorter tests such as the TUG based on a transversal turning pattern is more difficult and might not be feasible for free-living recordings, mainly because such turnings are also frequently performed during daily live activities of individuals besides from standardized gait tests. With regards to that, it needs to be investigated in future research to what extent the integration of standardized 4x10MWTs in mobile gait analysis provides an information gain or other advantages for the evaluation and interpretation of free-living recordings. Furthermore, different approaches for the detection of standardized gait tests in longer time series, such as window-based supervised machine learning approaches or the Hidden Markov Model, are worth a closer examination.
Nevertheless, due to the good area-based *F1-score* of 92.8%, the *gait test series* detection algorithm may allow to eliminate the necessity of smartphone annotations. It could be applied as base of a manual labeling tool instead of the smartphone annotations, to provide areas of interest in the scope of a GUI for expert-labeling. The user would then only be required to confirm the estimated results or adjust the proposed 4x10MWT borders.

Appendix A

Patents

A.1 System for clinical assessment of movement disorders

Publication Number	US8876739B2
Date of Publication	Apr. 12, 2017
Inventors	Arash Salarian, Fay Horak, Kamiar Aminian
Assignee	Oregon Health Science University
Abstract	According to one embodiment of the present invention, the system for clinical assessment of movement disorders (iTUG) is comprised of a) a protocol to assess gait, balance, and mobility; b) a plurality of wearable sensors including accelerometers, gyroscopes, magnetometers, optical sensors, and goniometers to record kinematics data obtained from a patient during said protocol; c) means for wirelessly transmitting said kinematics data to a storage and data processing server; and d) a plurality of statistical and biomedical signal processing methods to analyze said kinematic data and derive a plurality of metrics (outcomes) to objectively quantify movement disorders. A specially important outcome for the assessment of movement disorders is described, namely, the quantification of the onset and offset parameters during turning. A method for quantification of said onset and offset turning parameters involves 1) collecting data to measure the angle of the trunk during turning, 2) modeling said angle using a mathematical model, and 3) using numerical optimization and estimation methods for fitting the model in the data to determine said onset and offset turning parameters.

A.2 System and method for 3D gait assessment

Publication Number	US9307932B2
Date of Publication	May 16, 2013
Inventors	Benoît Mariani, Kamiar Aminian
Assignee	Ecole Polytechnique Federale de Lausanne (EPFL)
Abstract	The invention relates to a system and a method for assessment of walking and miming gait in human. The method is preferably based on the fusion of a portable device featuring inertial sensors and several new dedicated signal processing algorithms: the detection of specific temporal events and parameters, 5 optimized fusion and de-drifted integration of inertial signals, automatic and online virtual alignment of sensors module, 3D foot kinematics estimation, a kinematic model for automatic online heel and toe position estimation, and finally the extraction of relevant and clinically meaning-full outcome parameters. Advantageously including at least one wireless inertial module attached to foot, the system provides common spatio-temporal parameters (gait cycle time, stride length, and stride velocity), with the 10 advantage of being able to work in unconstrained condition such as during turning or running. It furthermore may provide original parameters for each gait cycle, both temporal (load, foot-flat and push duration) and spatial (foot clearance and turning angle), and their inter-cycles variability. The system and method according to the invention allows the assessment of various aspects of gait which have shown recently to be of premium importance in research and clinical field, including foot clearance, 15 turns, gait initiation and termination, running, or gait variability. The system may be light weight, easy to wear and use, and suitable for any application requiring objective and quantitative evaluation of gait without heavy laboratory settings.

A.3 Wireless sensor based quantitative falls risk assessment

Publication Number	US8805641B2
Date of Publication	Aug. 12, 2014
Inventors	Barry R. Greene
Assignee	Care Innovations LLC
Abstract	Methods and systems may provide for a plurality of kinematic sensors to be coupled to a corresponding plurality of shanks of an individual, a processor, and a memory to store a set of instructions. If executed by the processor, the instructions can cause the system to calculate a timed up and go (TUG) time segment based on angular velocity data from the plurality of kinematic sensors. The instructions may also cause the system to calculate a derived parameter based on the angular velocity data, and generate a falls risk assessment based on at least one of the TUG time segment and the derived parameter.

A.4 Automated near-fall detector

Publication Number	US10548512B2
Date of Publication	Feb. 04, 2020
Inventors	Jeffrey M. Hausdorff, Nir Giladi
Assignee	Medical Research Infrastructure and Health Services Fund of Tel Aviv Medical Center
Abstract	A method of gait data collection, the method comprising collecting movement data, determining from the data a movement parameter that includes a third order derivative of position, comparing the movement parameter with a threshold value, and counting at least a near fall if the movement parameter exceeds the threshold value.

Appendix B

Additional Figures and Tables

Patient	Valid Sessions	All 4x10MWT	All Gait Test Series	Compliant Gait Test Series
0	15	152	43	34
1	12	115	36	33
2	13	125	40	33
3	9	73	23	22
4	10	54	18	18
5	14	209	38	4
6	13	117	38	32
7	13	120	39	39
8	15	128	43	6
9	12	73	27	17
10	13	108	35	26
11	14	130	38	34
Total:	153	1404	418	298

Table B.1: **Amount of Gait Test Series per Patient.**

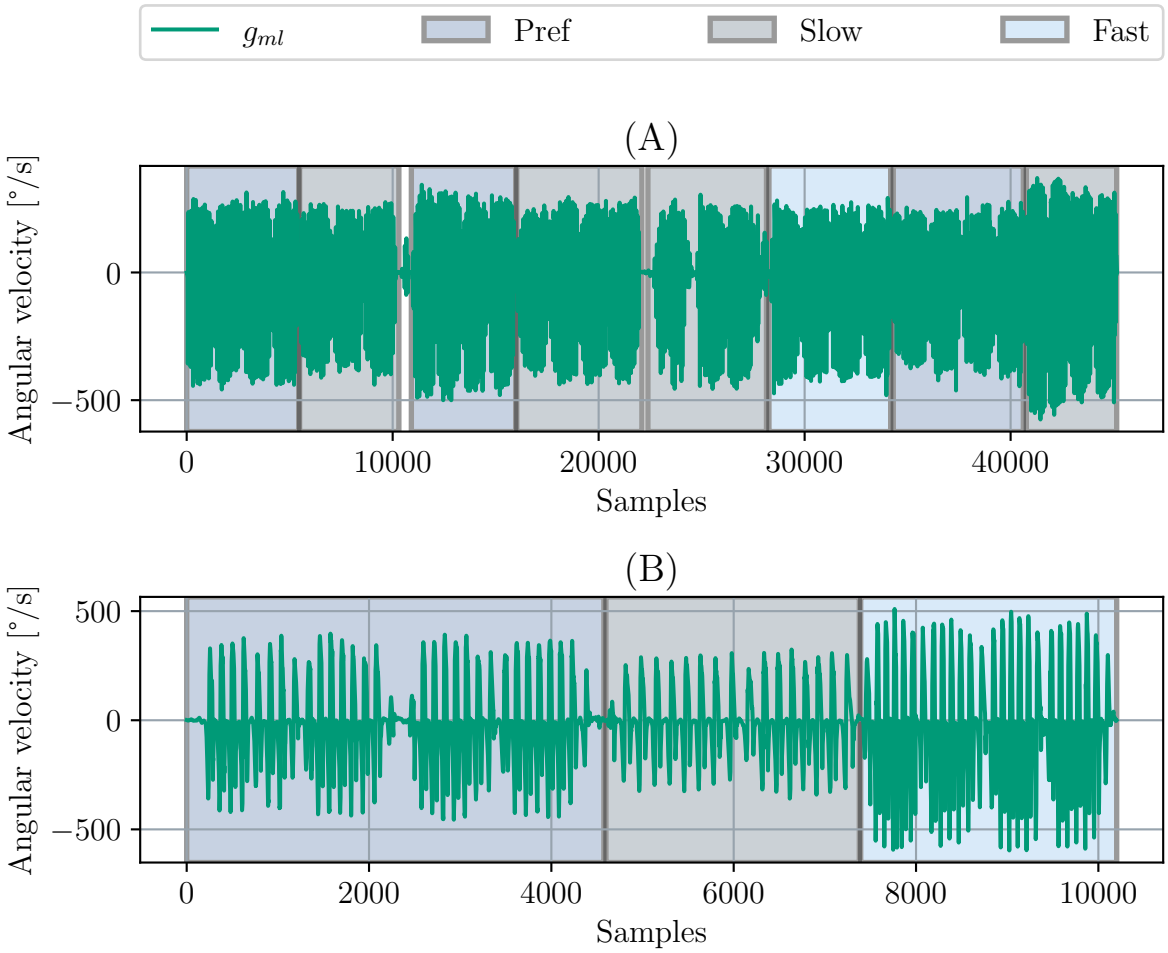

Figure B.1: **Examples for Non-compliant Gait Test Series.** The gait test series depicted in *(A)* includes eight 4x10MWTs of different speed levels instead of three. In *(B)*, the slow 4x10MWT was performed deviating from the study protocol by repeating the straight walkway two instead of four times.

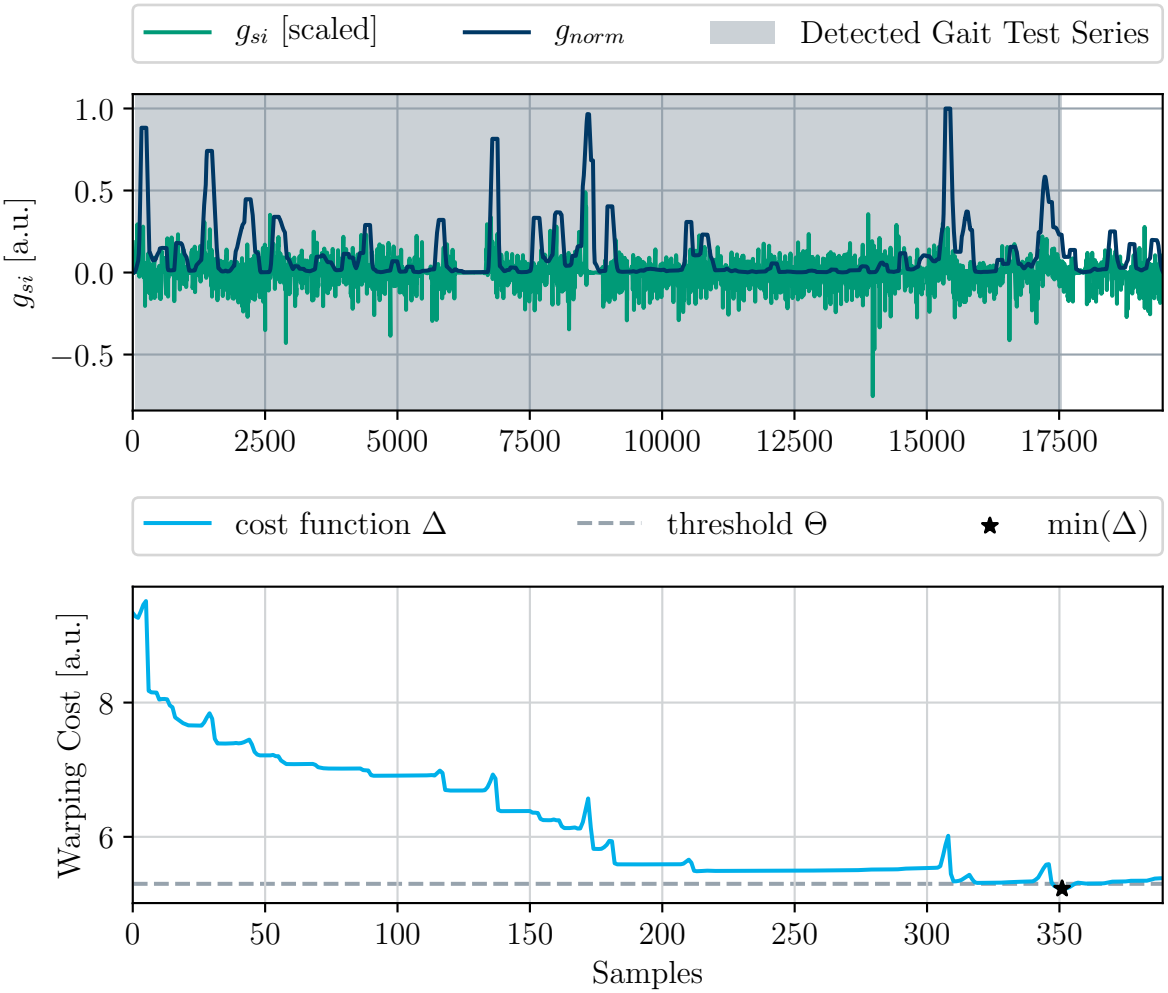

Figure B.2: **Detected Result without Turning Verification.** Because the amount of turning peaks in the preprocessed signal was not verified, this sequence of daily-life activity was mistakenly detected as a gait test.

Acronyms

10MWT 10 meter walking test

4x10MWT 4x10 meter walking test

DTW dynamic time warping

FOG freezing of gait

H&Y Hoehn & Yahr

IC initial contact

ICC intra-class correlation coefficient

IMU inertial measurement units

iTUG instrumented Timed Up and Go Test

MEMS micro-electro-mechanical systems

MS mid stance

msDTW multi-dimensional subsequence dynamic time warping

PD Parkinson's disease

sDTW subsequence dynamic time warping

ssDTW single-dimensional subsequence dynamic time warping

TC terminal contact

TUG Timed Up and Go Test

UPDRS Unified Parkinson's Disease Rating Scale

List of Figures

List of Tables

Bibliography

[Alb09] A. Albarbar, A. Badri, J. K. Sinha, and A. Starr. Performance evaluation of mems accelerometers. *Measurement*, 42(5):790 – 795, 2009.

[Bam08] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, et al. Gait analysis using a shoe-integrated wireless sensor system. *IEEE Transactions on Information Technology in Biomedicine*, 12(4):413–423, 2008.

[Bar11] J. Barth, J. Klucken, P. Kugler, et al. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 868–871, 2011.

[Bar15] J. Barth, C. Oberndorfer, C. Pasluosta, et al. Stride segmentation during free walk movements using multi-dimensional subsequence dynamic time warping on inertial sensor data. *Sensors*, 15(3):6419–6440, 2015.

[Bas02] M. Basu. Gaussian-based edge-detection methods-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(3):252–260, 2002.

[Ben15] E. T. Benser. Trends in inertial sensors and applications. In *2015 IEEE International Symposium on Inertial Sensors and Systems (ISISS) Proceedings*, pages 1–4, 2015.

[Ber18] M. Bertoli, A. Cereatti, D. Trojaniello, et al. Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: Multicenter validation among parkinson, mildly cognitively impaired and healthy older adults. *BioMedical Engineering OnLine*, 17, 2018.

[Blo16] B. R. Bloem, J. Marinus, Q. Almeida, et al. Measurement instruments to assess posture, gait, and balance in Parkinson's disease: Critique and recommendations. *Movement Disorders*, 31(9):1342–1355, 2016.

[Bro16] M. A. Brodie, M. J. Coppens, S. R. Lord, et al. Wearable pendant device monitoring using new wavelet-based methods shows daily life and laboratory gaits are different. *Medical and Biological Engineering and Computing*, 54(4):663–674, 2016.

[Bro17] M. A. Brodie, M. J. Coppens, A. Ejupi, et al. Comparison between clinical gait and daily-life gait assessments of fall risk in older people. *Geriatrics and Gerontology International*, 17(11):2274–2282, 2017.

[Bro19] L. Brognara, P. Palumbo, B. Grimm, and L. Palmerini. Assessing Gait in Parkinson's Disease Using Wearable Motion Sensors: A Systematic Review. *Diseases*, 7(1):18, 2019.

[Buc19] C. Buckley, L. Alcock, R. Rehman, et al. The role of movement analysis in diagnosing and monitoring neurodegenerative conditions: Insights from gait and postural control. *Brain Sciences*, 9:34, 02 2019.

[But30] S. Butterworth. On the theory of filter amplifiers. *Experimental Wireless and the Wireless Engineer*, 7:536–541, 1930.

[Can87] J. Canny. A computational approach to edge detection. In M. A. Fischler and O. Firschein, editors, *Readings in Computer Vision*, pages 184 – 203. Morgan Kaufmann, San Francisco (CA), 1987.

[Car13] B. Carse, B. Meadows, R. Bowers, and P. Rowe. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3d motion analysis system. *Physiotherapy*, 99(4):347 – 351, 2013.

[Del16] S. Del Din, A. Godfrey, C. Mazzà, et al. Free-living monitoring of Parkinson's disease: Lessons from the field. *Movement Disorders*, 31(9):1293–1313, 2016.

[Del17] S. Del Din, B. Galna, A. Godfrey, et al. Analysis of Free-Living Gait in Older Adults With and Without Parkinson's Disease and With and Without a History of Falls: Identifying Generic and Disease-Specific Characteristics. *The Journals of Gerontology: Series A*, 74(4):500–506, 12 2017.

[Dew14] D. Campbell Dewey, S. Miocinovic, I. Bernstein, et al. Automated gait and balance parameters diagnose and correlate with severity in parkinson disease. *Journal of the Neurological Sciences*, 345(1):131 – 138, 2014.

[Dib17] V. Dibilio, A. Nicoletti, G. Mostile, et al. Dopaminergic and non-dopaminergic gait components assessed by instrumented timed up and go test in parkinson's disease. *Journal of Neural Transmission*, 124:1539–1546, 2017.

[Dor07] E. R. Dorsey, R. Constantinescu, J. P. Thompson, et al. Projected number of people with parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5):384–386, 2007.

[EG14] M. El-Gohary, S. Pearson, J. McNames, et al. Continuous monitoring of turning in patients with movement disability. *Sensors (Switzerland)*, 14(1):356–369, 2014.

[Fer95] F. Ferraris, U. Grimaldi, and M. Parvis. Procedure for effortless in-field calibration of three-axis rate gyros and accelerometers. *Sensors and Materials*, 7:311–311, 1995.

[Fis19] S. Fischer. Macro analysis of free-living gait in parkinson's disease. Bachelor's thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019.

[Fis20] S. Fischer, M. Ullrich, A. Küderle, et al. Automatic clinical gait test detection from inertial sensor data. In *2020 42th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020.

[For11] K. B. Foreman, O. Addison, H. S. Kim, and L. E. Dibble. Testing balance and fall risk in persons with parkinson disease, an argument for ecologically valid testing. *Parkinsonism & Related Disorders*, 17(3):166 – 171, 2011.

[Gal13] B. Galna, S. Lord, and L. Rochester. Is gait variability reliable in older adults and parkinson's disease? towards an optimal testing protocol. *Gait & Posture*, 37(4):580 – 585, 2013.

[Gal19] I. Galperin, I. Hillel, S. Del Din, et al. Associations between daily-living physical activity and laboratory-based assessments of motor severity in patients with falls and Parkinson's disease. *Parkinsonism and Related Disorders*, 62(January):85–90, 2019.

[Gha19] N. H. Ghassemi, J. Hannink, N. Roth, et al. Turning analysis during standardized test using on-shoe wearable sensors in parkinson's disease. *Sensors (Switzerland)*, 19(14):16–18, 2019.

[Gra06] E.W. Grafarend. *Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models*. Walter de Gruyter, 2006.

[Gre10] B. R. Greene, A. O'Donovan, R. Romero-Ortuno, et al. Quantitative falls risk assessment using the timed up and go test. *IEEE Transactions on Biomedical Engineering*, 57(12):2918–2926, 2010.

[Har80] J. E. Hardebo and C. Owman. Barrier mechanisms for neurotransmitter monoamines and their precursors at the blood-brain interface. *Annals of Neurology*, 8(1):1–11, 1980.

[Har13] S. Harada, D. Sato, H. Takagi, and C. Asakawa. Characteristics of elderly user behavior on mobile multi-touch devices. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8120 LNCS(PART 4):323–341, 2013.

[Hau03] J. Hausdorff, J. Schaafsma, Y. Balash, et al. Impaired regulation of stride variability in parkinson's disease subjects with freezing gait. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 149:187–94, 03 2003.

[Hil19] I. Hillel, E. Gazit, A. Nieuwboer, et al. Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. *European Review of Aging and Physical Activity*, 16(1):1–12, 2019.

[Hor06] O. Hornykiewicz. The discovery of dopamine deficiency in the parkinsonian brain. In P. Riederer, H. Reichmann, M. B. H. Youdim, and M. Gerlach, editors, *Parkinson's Disease and Related Disorders*, pages 9–15. Springer Vienna, 2006.

[Hun10] C.-W. Hung, Y.-C. Chen, W.-L. Hsieh, et al. Ageing and neurodegenerative diseases. *Ageing Research Reviews*, 9:S36 – S46, 2010. Ageing Research in Chinese Societies.

[Ilu13] T. Iluz, E. Gazit, T. Herman, et al. Automated detection of missteps during community ambulation in patients with parkinson's disease: a new approach for quantifying fall risk in the community setting. *Journal of NeuroEngineering and Rehabilitation*, 11:48 – 48, 2013.

[Jan08] J. Jankovic. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(4):368–376, 2008.

[Kan15] C. M. Kanzler, J. Barth, A. Rampp, et al. Inertial sensor based and shoe size independent gait analysis including heel and toe clearance estimation. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5424–5427, 2015.

[Klu13] J. Klucken, J. Barth, P. Kugler, et al. Unbiased and mobile gait analysis detects motor impairment in parkinson's disease. *PLOS ONE*, 8:1–9, 02 2013.

[Liu09] T. Liu, Y. Inoue, and K. Shibata. Development of a wearable sensor system for quantitative gait analysis. *Measurement*, 42(7):978 – 988, 2009.

[Mü07] M. Müller. *Information Retrieval for Music and Motion*, chapter 4, pages 69–84. Springer, 2007.

[Mac05] H. G. MacDougall and S. T. Moore. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology*, 99(3):1164–1173, 2005.

[Mac07] M. Macht, Y. Kaussner, J. C. Möller, et al. Predictors of freezing in parkinson's disease: A survey of 6,620 patients. *Movement Disorders*, 22(7):953–956, 2007.

[Man15] M. Mancini, M. El-Gohary, S. Pearson, et al. Continuous monitoring of turning in Parkinson's disease: Rehabilitation potential. *NeuroRehabilitation*, 37(1):3–10, aug 2015.

[Mar10] B. Mariani, C. Hoskovec, S. Rochat, et al. 3d gait assessment in young and elderly subjects using foot-worn inertial sensors. *Journal of Biomechanics*, 43(15):2999–3006, 2010.

[McD01] A. L. McDonough, M. Batavia, F. C. Chen, et al. The validity and reliability of the gaitrite system's measurements: A preliminary evaluation. *Archives of Physical Medicine and Rehabilitation*, 82(3):419 – 425, 2001.

[Mil13] S. Miller, M. Diehl, J. Filip, and E. Long. Short-distance walking speed tests in people with parkinson disease: Reliability, responsiveness, and validity. *Gait & posture*, 39, 10 2013.

[Mye81] C. S. Myers and L. R. Rabiner. A comparative study of several dynamic time-warping algorithms for connected-word recognition. *Bell System Technical Journal*, 60(7):1389–1409, 1981.

[Naf14] M. Naftali and L. Findlater. Accessibility in context: Understanding the truly mobile experience of smartphone users with motor impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, page 209–216. Association for Computing Machinery, 2014.

[Ngu19] A. Nguyen, N. Roth, N. H. Ghassemi, et al. Development and clinical validation of inertial sensor-based gait-clustering methods in Parkinson's disease. *Journal of NeuroEngineering and Rehabilitation*, 16(1):1–14, 2019.

[Osh06] R. Oshana. 4 - overview of digital signal processing algorithms. In R. Oshana, editor, *DSP Software Development Techniques for Embedded and Real-Time Systems*, Embedded Technology, pages 59 – 121. Newnes, Burlington, 2006.

[Par02] J. Parkinson. An essay on the shaking palsy. 1817. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 14(2):223–236, 2002.

[Pir16] W. Pirker and R. Katzenschlager. Gait disorders in adults and the elderly: A clinical guide. *Wiener klinische Wochenschrift*, 129, 10 2016.

[Pul18] C. L. Pulliam, D. A. Heldman, E. B. Brokaw, et al. Continuous assessment of levodopa response in parkinson's disease using wearable motion sensors. *IEEE Transactions on Biomedical Engineering*, 65(1):159–164, 2018.

[Ram15] A. Rampp, J. Barth, S. Schülein, et al. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE Transactions on Biomedical Engineering*, 62(4):1089–1097, 2015.

[Rot18] N. Roth, C. F. Martindale, B. M. Eskofier, et al. Synchronized sensor insoles for clinical gait analysis in home-monitoring applications. *Current Directions in Biomedical Engineering*, 4(1):433 – 437, 2018.

[Sal09] A. Salarian, C. Zampieri, F. B. Horak, et al. Analyzing 180° turns using an inertial system reveals early signs of progression of Parkinson's disease. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, pages 224–227, 2009.

[Sal10] A. Salarian, F. B. Horak, C. Zampieri, et al. itug, a sensitive and reliable measure of mobility. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(3):303–310, 2010.

[Sam04] A. Samii, J. G. Nutt, and B. R. Ransom. Parkinson's disease. *The Lancet*, 363(9423):1783 – 1793, 2004.

[Saz14] E. Sazonov. *Wearable Sensors: Fundamentals, Implementation and Applications*. Elsevier Science, 2014.

[Sch17] J. C. M. Schlachetzki, J. Barth, F. Marxreiter, et al. Wearable sensors objectively measure gait parameters in parkinson's disease. *PLOS ONE*, 12(10):1–18, 10 2017.

[Sko17] M. Skorvanek, P. Martinez-Martin, N. Kovacs, et al. Differences in mds-updrs scores based on hoehn and yahr stage and disease duration. *Movement Disorders Clinical Practice*, 4(4):536–544, 2017.

[Smi16] E. Smith, L. Walsh, J. Doyle, et al. The reliability of the quantitative timed up and go test (QTUG) measured over five consecutive days under single and dual-task conditions in community dwelling older adults. *Gait and Posture*, 43:239–244, 2016.

[Sta08] E. Stack and A. Ashburn. Dysfunctional turning in parkinson's disease. *Disability and rehabilitation*, 30:1222–9, 06 2008.

[Ste08] T. Steffen and M. Seney. Ambulation Tests, the 36-Item Short- Form Health Survey, and the Unified Parkinson Disease Rating Scale in People With Parkinsonism. *Physical Therapy*, 88(1-2):733–746, 2008.

[Tan96] C. M. Tanner and S. M. Goldman. Epidemiology of parkinson's disease. *Neurologic Clinics*, 14(2):317 – 335, 1996.

[Tha18] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2018.

[The09] S. Theodoridis and K. Koutroumbas. Chapter 8 - template matching. In S. Theodoridis and K. Koutroumbas, editors, *Pattern Recognition (Fourth Edition)*, pages 481 – 519. Academic Press, Boston, 2009.

[Tun17] C. Tunca, N. Pehlivan, N. Ak, et al. Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders. *Sensors*, 17(4), 2017.

[Ull20] M. Ullrich, A. Küderle, J. Hannink, et al. Detection of gait from continuous inertial sensor data using harmonic frequencies. *IEEE Journal of Biomedical and Health Informatics*, 24(7):1869–1878, 2020.

[Ur 19] R. Z. Ur Rehman, S. Del Din, J. Q. Shi, et al. Comparison of walking protocols and gait assessment systems for machine learning-based classification of parkinson's disease. *Sensors (Switzerland)*, 19(24), 2019.

[Vir20] P. Virtanen, R. Gommers, T. E. Oliphant, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020.

[Wal06] B. L. Walter and J. L. Vitek. Parkinson's disease. In R. T. Johnson, J. W. Griffin, and J. C. McArthur, editors, *Current Therapy in Neurologic Disease (Seventh Edition)*, pages 281 – 288. Mosby, Philadelphia, 2006.

[Wei14] A. Weiss, T. Herman, N. Giladi, and J. M. Hausdorff. Objective assessment of fall risk in parkinson's disease using a body-fixed sensor worn for 3 days. *PLOS ONE*, 9(5):1–10, 05 2014.

[Whi07] M. W. Whittle. Chapter 2 - normal gait. In *Gait Analysis (Fourth Edition)*, pages 47 – 100. Butterworth-Heinemann, Edinburgh, 2007.

[Woo07] O. J. Woodman. An introduction to inertial navigation, 2007.

[Zam11] C. Zampieri, A. Salarian, P. Carlson-Kuhta, et al. Assessing mobility at home in people with early parkinson's disease using an instrumented timed up and go test. *Parkinsonism and Related Disorders*, 17(4):277 – 280, 2011.