



TALLINN UNIVERSITY OF TECHNOLOGY

SCHOOL OF ENGINEERING

Department of Mechatronics

**AUTOMATED GAIT EVENT DETECTION WITH THE
HELP OF WEARABLE SENSORS
AUTOMAATNE KÕNNIJUHTUMI TUVASTAMINE
KANTAVATE SENSORITE ABIL**

Student: Andrii Boryshkevych

Student code: 201619MAHM

Supervisor: Jeffrey Andrew Tuhtan

Associate Professor

Co-supervisor: Cecilia Monoli

PhD Candidate

Co-supervisor: Mart Tämre

Professor emeritus

Tallinn 2022

(On the reverse side of title page)

AUTHOR'S DECLARATION

Hereby I declare, that I have written this thesis independently.

No academic degree has been applied for based on this material. All works, major viewpoints and data of the other authors used in this thesis have been referenced.

"18" May, 2022.

Author: Andrii Boryshkevych

/signature /

Thesis is in accordance with terms and requirements

"18" May, 2022.

Supervisor: Jeffrey Andrew Tuhtan

/signature/

Accepted for defence

"....."20... .

Chairman of theses defence commission:

/name and signature/

Non-exclusive licence for reproduction and publication of a graduation thesis¹

I, Andrii Boryshkeych (date of birth: 14.06.1998) hereby

grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis Automated Gait Event Identification with the Help of Wearable Sensors,

supervised by Associate Professor_Jeffrey Andrew Tuhtan,

to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

¹ *The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.*

**TalTech Department of Electrical Power Engineering and
Mechatronics**

THESIS TASK

Student: Andrii Boryshkevych 201619MAHM

Study programme, MAHM02/18 – Mechatronics

Supervisor(s): Associate Professor Jeffrey Andrew Tuhtan PHONE, PhD Candidate

Cecilia Monoli PHONE, Professor emeritus Mart Tamre PHONE

Thesis topic:

Automated Gait Event Identification with the Help of Wearable Sensors

Automaatne kõnnijuhumi tuvastamine kantavate sensorite abil

Thesis main objectives:

1. Validate IMU sensor developed at TalTech for human gait analysis
2. Develop a sensor mounting system for the lower limb of the human body that will assure firm fixation and convenient utilization
3. Develop an automated gait event identification data processing algorithm and implement it on the study group to validate the developed data processing method

Thesis tasks and time schedule:

No	Task description	Deadline
1.	Literature review	December 2021
2.	TalTech sensor validation for gait analysis	January 2021
3.	Data acquisition, automated gait event detection data processing algorithm development, and testing	April 2022
4.	Writing Thesis and Defense	May - June 2022

Language: English **Deadline for submission of thesis:** "18" May 2022

Student: Andrii Boryshkevych "18" May 2022

/signature/

Supervisor: Jeffrey Andrew Tuhtan "18" May 2022

/signature/

Head of study programme: Anton Rassõlkin "18" May 2022

/signature/

Terms of thesis closed defence and/or restricted access conditions to be formulated on the reverse side

CONTENTS

1. INTRODUCTION.....	6
2. LITERATURE REVIEW	10
2.1 Sensor placement.....	11
2.2 Utilized sensors and data analysis methods	15
2.3 Literature review conclusions	20
2.4 Aims of the Thesis	20
3. MATERIALS AND METHODS.....	21
3.1 Materials and Tools.....	21
TinyTag.....	21
Sensor casing design and mounting system.....	22
Buffer materials for mechanical damper	24
3.2 Methodology	27
Pilot study 1	30
Pilot study 2	32
Experimental study.....	33
3.3 Numerical filter	34
3.4 Automated gait events identification algorithm.....	35
Matrix profiling.....	35
Peak identification	36
Novel gait analysis data processing	36
4. RESULTS AND DISCUSSION.....	40
4.1 TinyTag validation for gait analysis	40
4.2 Mechanical damper.....	45
4.3 Numerical filter and its parameters.....	48
Window length for Savitzky-Golay filter.....	48
Numerical filter performance evaluation	53
4.4 Efficiency of the developed algorithm.....	55
5. CONCLUSIONS	59
6. KOKKUVÕTE	61
7. LIST OF REFERENCES	63
APPENDIX 1	68
APPENDIX 2	70

1. INTRODUCTION

Every human has a distinctive gait that contains unique information about a person's health [1], making it an ideal topic for a clinical study to assess fundamental kinematics of human motion [2], identify developing pathologies [3], or aid rehabilitation process [4]. Apart from clinical applications, human gait analysis finds application in various fields such as sports, computer games, surveillance, human recognition, modeling, and some other fields [5], [6].

The act of walking is defined by the constant repetition of gait cycles marked by specific occurrences, called gait events. The gait events are *heel strike*, the moment when the foot hits the ground, and *toe-off*, which describes the moment of foot leaving the ground. A complete cycle is identified by the period of time and specific motion patterns occurring between two homolateral heel strikes. A gait cycle can be divided into two main phases: the *stance phase*, a period of time when the foot is in contact with the ground, and the *swing phase*, the period of time where the foot is flying. Stance and swing are the two main phases that can also further spread in smaller sub-phases. The sub-phases nomenclature varies in the literature, but they can be described as follows (see Figure 1.1): initial contact, loading response, mid stance, terminal stance, and pre-swing, initial swing, mid-swing, and terminal swing. The gait events are following one another and together conclude human gait.

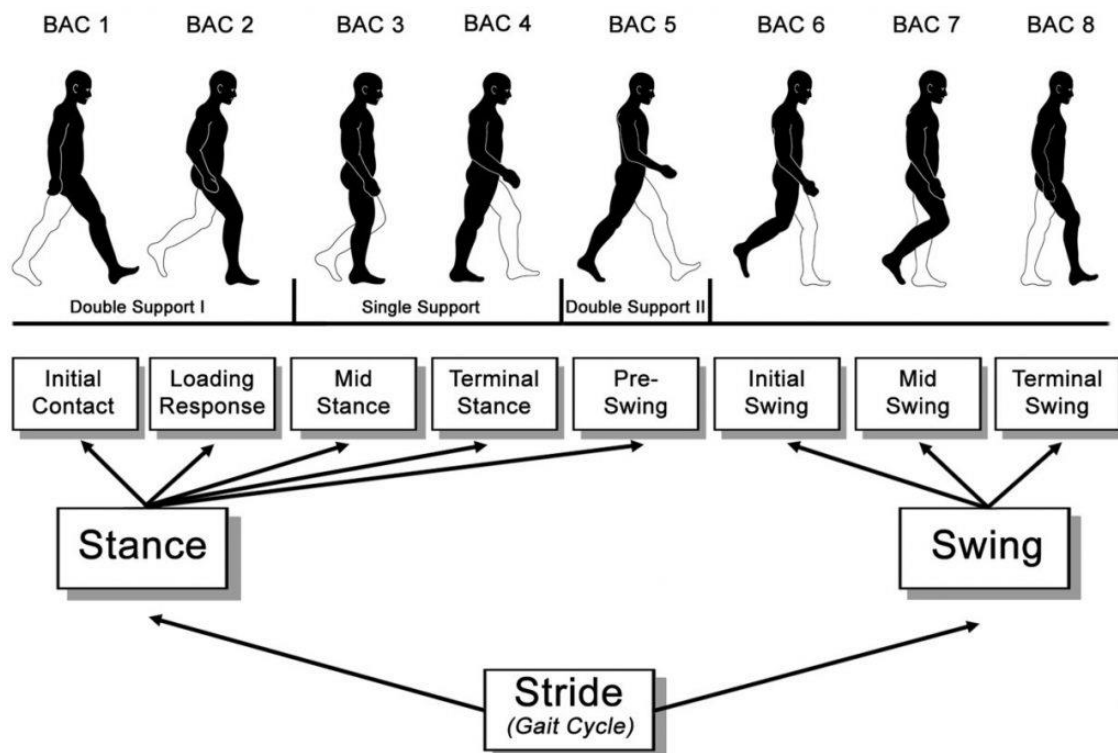


Figure 1.1 Phases and Sub-Phases of the gait cycle [7]

Gait analysis focuses on the study of the broad range of the parameters that characterize human gait, such as, for example, gait cycle time, cadence, walking velocity, step length, stride duration, stride length, swing, and stance percentage of the gait cycle, swing and stance time [8]. Gait parameters, in turn, are calculated based on the occurrence and duration of gait events [8]. The occurrence of changes in the normal gait may point to the development of various diseases, such as, for instance, Parkinson's disease, where a decrease in cadence [9] (number of steps per minute) occurs.

There are various methods to evaluate human gait but they all can be categorized into two main groups, namely clinician's direct assessment and data-driven methods [10]. The first method is based on the visual observation of the doctor and his evaluation of the gait. Even when it is attempted to categorize and classify gait with questionnaires and scales, these methods rely completely on the clinicians' expertise [1] and therefore they are highly subjective. Conversely, the other category of methods relies on quantitative data obtained from the sensors and devices capturing human gait. These methods enable recording and quantifying human locomotion and then through data processing techniques produce objective data for gait analysis and evaluation.

The data-driven gait analysis, in turn, can be distinguished into three main groups: image processing, floor sensors, and inertial wearable sensors [10]. An overall summary of these methods along with their strengths and weaknesses is provided in Table 1.1. The most accurate system derives from image processing and is called the optoelectronics motion capture method. This system provides undoubtedly the best quality data for gait analysis of clinical interest. Light-reflecting markers are placed on specific anatomical points of the body and, recorded by multiple infrared cameras, allow reconstruction of human motion in its complexity. However, this methodology has some limitations ([11], [12]), being restricted to the laboratory environment, as it requires a large amount of expensive equipment and personnel to carry out the experiment [13]. It also introduces certain limitations to the parameters of the experiments, primarily the length of the walking path, due to the limited field of view of the cameras. It has also been pointed out in [12], [14] that in-laboratory monitored walking gait and outdoor day to day walking gait may differ, as in lab walking is often scripted according to the research purposes and walking gait might be unrealistic due to psychological factor. The predecessor of optoelectronics gait analysis is a camera-based method, where gait analysis is done through direct image (video recordings) processing and this approach shares the same restrictions. Similar limitations apply to the floor sensors methods [15], where the study subject walks along the platform equipped with pressure sensors. Moreover, rather than the measuring kinematic

characteristics of motion, force plates enable the investigation of the dynamic components of the gait.

Table 1.1 Gait analysis methods classification

	Clinician's visual observation	Data-driven methods		
		Image processing	Floor Sensors	Wearables
Strengths	Doesn't require expensive equipment, nor personnel to carry out the gait evaluation	Allows precise 3D capturing of human motion and enables comprehensive gait analysis and quantification	Enables gait analysis based on the Ground Reaction Forces, Produces reliable data for gait analysis	Cost-effective, Allows gait investigation in the real-life conditions, Sensors are tiny and lightweight
Weaknesses	Subjective, Doesn't allow to quantify gait parameters, Evaluation based on clinician's skills and expertise	A large cost of the equipment, Requires a considerable amount of personnel for an experiment, Experiment is restrained to the lab environment	A large cost of the equipment, Requires a considerable amount of personnel for an experiment, Experiment is restrained to the lab environment	Lack of standardization regarding sensor placement, data reporting, and processing techniques

The third group of objective methods which implements wearable sensors enables overcoming these limitations. This caused a growing amount of research, in the past two decades, utilizing inertial sensors [16]. Initially, data collection was performed using accelerometers only. Subsequent advances in microelectronics now allow for the widespread use of Inertial Measurement Units (IMUs) for wearable sensors. An IMU combines an accelerometer, gyroscope, and magnetometer, allowing for a more comprehensive kinematic assessment of the gait. When data are combined properly, the three sensors can be used to complement each other, improving both the efficacy and quality of the gathered data. The state-of-the-art wearable sensors ([17], [18]) now make ubiquitous use of IMUs for human gait analysis. Once mounted on the test subject, these devices can record linear acceleration, angular velocity, and magnetic field data which is later processed for gait analysis. The major advantage of this method is its flexibility in the test environment, as no special laboratory equipment nor large personnel is required to carry out data gathering. Also, it provides an

opportunity for the data collection for an extended period of time in the real-world environment, when the proper battery life time is assured.

Despite such rapid growth in the amount of research implementing IMUs for gait analysis, there are still problems to be solved. One of those problems is the lack of standardization for data gathering and evaluation methods [11]. Another gap that is infrequently addressed in the literature is optimal sensor placement. The placement of sensors varies broadly, mainly according to the investigation purpose and the algorithm used afterward. Most commonly met are foot, shank, and thigh, however, there remains little justification or critical evaluation of sensor placement and sensitivity to the attachment method for studies of the walking gait [13].

2. LITERATURE REVIEW

The purpose of the literature review was to examine previous studies and identify the optimal inertial sensor placement, for the robust and accurate gait data collection. Along with that, algorithms that are applied for data processing, together with their advantages and disadvantages were considered.

Table 2.1 provides the summary of all the keywords combinations that were used during the literature review along with a quantity of the given results. Only publications of the past 10 years were taken into consideration. 784 research papers in total were found. After removing duplicates and unrelated studies based on the title and abstract scanning, 34 studies in total were considered for the full-text investigation.

Table 2.1 Keywords combinations used for literature review

Keywords combinations	Search results
Wearable IMU gait analysis	180
IMU gait event extraction	27
IMU gait event detection	63
IMU gait analysis placement	83
IMU mounting gait analysis	72
Wearable gait analysis sensor location	161
IMU optimal placement gait analysis	53
Wearable optimal placement gait analysis	145

Out of 34 research papers, 18 conducted experiments exploiting wearable sensors for gait analysis or investigated the efficiency of various algorithms. Seven of those studies ([14], [19]–[24]) utilized sensors located on the lower back, chest, or wrist. They were removed from the final analysis, due to findings in [4], [13], [21], [25], where authors concluded that for a more comprehensive gait analysis lower limbs, and feet specifically, seem to be the best choice. Indeed, the closer the sensor to the foot, the wider range of motion it will be able to record, and therefore produce better data for further analysis. This was best highlighted in [26] where the research group was investigating the sensor's accuracy for gait analysis in the following four locations: L5

spinal cord, thighs, shanks, and foot. Results of this study showed that sensors when placed on the L5 resulted in the worst accuracy. Hence, based on these conclusions in the present MSc Thesis the focus will be placed on the investigation of sensor performance placed on the lower limb.

2.1 Sensor placement

Table 2.2 summarizes findings regarding sensor placements and fixation methods. In terms of the sensor location, three main areas can be highlighted, one for each body segment: thigh, shank, and foot. Sensors were placed on healthy subjects as well as on subjects with pathological gait caused by various conditions (e.g. Parkinson's disease, Cerebral Palsy, stroke survivors). For both healthy and abnormal gait shanks and foot located sensors were giving the best results for gait events recognition, reaching over 90% alignment with gold standard systems used as a reference.

Three studies ([26]–[28]) were investigating how the sensor's location influences the gait events identification, and thus gait analysis. The study group in [26] consisted of 22 patients with Neurological diseases, where only the stroke group had a considerable population (11 patients) and the other 9 were representatives of four other diseases. As the result of this research, the authors highlighted that for neurological gait, shanks might be the optimal sensor placement. In [27] authors were investigating the influence of the sensor locations on gait recognition for children with Cerebral Palsy, whereas in [28] researchers investigated specific foot locations, not concerning other parts of the human body.

Studies also vary significantly in the number of utilized sensors. However, as highlighted in [25], at least two sensors, one for each lower limb, should be utilized. This is because a single sensor doesn't provide enough data for comprehensive gait analysis, especially in the case of pathological, asymmetrical gait.

For sensor mounting, straps, sometimes combined with elastic belts, are preferred over the tape, presumably due to better fixation. However, no cross-comparison between these two methods was found. Interestingly, in none of the reviewed studies the soft-tissue artifact, as the relative motion between skin and sensor, was taken into account. This may cause discrepancy which directly influences the accuracy and robustness of the data collection [3]. Additionally, the review paper [29] highlights that relative linear and angular displacement is lower when sensors are mounted on the shanks and ankles, which is coherent with the results of the aforementioned studies.

As to the identified gait events, it can be seen from Table 2.2 that mainly researchers are targeting Heel Strike (HS) and Toe-Off (TO), and only one study [27] approached to quantify 6 gait events. This is due to the fact that HS and TO stand-alone enable estimation of the broad range of the spatial and temporal parameters for an extensive gait analysis.

To conclude, there is still a lack of agreement on the optimal sensor location for accurate gait events detection [15], [30]. Moreover, evidence is present that for pathological gait general optimal sensor location might not be established, due to high variability in the gait abnormality for the specific disease. Therefore, as pointed out in [31], the sensor location should be based on the purpose of the application and, for pathological gait analysis, the target group of the study. Regardless of the study group, straps are preferred to fix the sensor on the subject, and HS and TO are the gait events that enable broad, quantitative gait analysis.

Table 2.2 Details regarding sensor placement identified during literature review classified by the sensor location, reasoning of the location, mounting system and target gait events of the research, heel strike (HS), flat foot (FF), mid-stance (MSt), heel off (HO), toe-off (TO) and mid-swing (MSw)

Nr.	Study ref	Year	Purpose of the study	Sensor Location	Placement Justification	Mounting System	Identified gait events
1	[32]	2017	Validation of the algorithm for the spatial gait parameters estimation based on Initial Contact detection	Shanks	NONE	Straps	HS
2	[33]	2017	Heel strike and Toe-Off events recording using accelerometers	Waist, wrist, and both ankles	Based on the previous studies utilizing tested algorithms	Elastic band and Velcro straps	HS and TO
3	[34]	2017	IC and TC contact identification to estimate stance and swing time; assess 4 different data processing algorithms	The instep of the foot and mid-shank	NONE	Suitable elastic belts	HS and TO
4	[35]	2017	Development and validation of the new method for HS and TO gait events identification	Shanks and Thighs	To repeat previous studies	Elastic open-patella knee brace with sewn-in straps	HS and TO
5	[31]	2018	HS and TO gait events identification for temporal parameters calculations	Trunk (at L5 level), shanks (about five centimeters above lateral malleolus), and feet (on the dorsal surface of each shoe)	To repeat previous studies	NOT MENTIONED	HS and TO
6	[26]	2018	Investigate how sensor placement influences the accuracy of gait event identification	7 sensors in total; L5, thighs, shanks, and foot	Investigation of the sensor placement on its performance	Medical-grade tape	HS and TO

Table 2.2 continued

Nr.	Study ref	Year	Purpose of the study	Sensor Location	Placement Justification	Mounting System	Identified gait events
7	[28]	2018	Analyze how sensor location on the foot influences the data collecting process	Metatarsal, Proximal phalange, side metatarsal, Talus, and Achilles tendon	Investigation of the sensor placement on its performance	Self-designed case and attached to the foot with Velcro elastic belt	HS and TO
8	[36]	2018	Real-Time gait event identification to feed algorithm for robot control	The backside of the foot	Purpose of the application	Custom designed mounting	HS and TO
9	[27]	2018	Evaluate how sensor placement influences the accuracy of the gait events detection in children with CP	Thighs and shanks, feet	Investigation of the sensor placement on its performance s	Hypoallergenic adhesive film	HS and TO
10	[25]	2019	Evaluate what data, acceleration, angular velocity, or both, provides the most accurate gait events identification	Side of the foot	NONE	Straps	HS, FF, MSt, HO, TO, and MSw,
11	[37]	2020	Validation of an IMU gait analysis algorithm for gait monitoring in daily life situations	Top of the foot	NONE	NOT MENTIONED	HS and TO

2.2 Utilized sensors and data analysis methods

A summary of utilized sensors and data analysis methods is presented in Table 2.3. It can be noticed that predominantly commercially available IMUs are utilized, with only one research group [35] using the self-made sensor. Only one study [33] reported using the accelerometer stand-alone, whereas all others were utilizing IMUs. High inconsistency was noticed in terms of reported sensor parameters, where even sensitivity and dimensions are not always specified. Similar inconsistency is present in the sampling frequency preferred for experiments. The most commonly met is 50 Hz (three times), however, eight other studies used a frequency that exceeds it at least twice, with one study [25] using a frequency of 400 Hz.

As to raw data filtering, the investigated studies were varying significantly. Only two studies, [25] and [35], declared to use raw data for the analysis. Among applied data filters, the most commonly utilized were Kalman, Savitzky-Golay, and Butterworth filters. For analyzed data representation Bland-Altman plot and Boxplot have been used frequently.

Both angular velocity and acceleration data were used for gait analysis, noticing no consistency. Despite magnetometer being present in the vast majority of utilized devices, it wasn't used at all. Among 28 data analysis processes that were discussed in the studies, gait events were identified based on acceleration 15 times and 12 times based on the angular velocity. One study [25] went in-depth to analyze what data, acceleration or angular velocity, is more suitable for gait identification. Researchers used different configurations of the signal (e.g. both acceleration and angular velocity from both feet, acceleration from both feet, angular velocity from both feet, single foot acceleration, etc.) to feed a Hidden Markov Model based algorithm. It was found that angular velocity alone or a combination of both acceleration and angular velocity gives the best accuracy for HS and TO gait events identification. This is coherent with a systematic review made in 2016 [4], where authors concluded that angular velocity of the foot shows the best performance among other inertial quantities.

Predominantly acceleration or angular velocity peaks identification method is used for HS and TO gait events recognition. Alternatively, another common method was an analysis of the wavelet transform of the sensor signal. In terms of automatic gait event extraction primarily self-developed algorithms were used. It is worth mentioning that all studies except for one [26] used either an optoelectronics system, force plates, or a combination of the two as the ground truth for the comparison. These

systems are considered to be the gold standard in the field of gait analysis and therefore serve as a reference point for validating wearable devices.

Table 2.3 Details regarding data analysis methods and utilized sensor identified during literature review classified by type of sensor, sampling frequency, utilized data filter, and data processing method, accelerometer (ACC), gyroscope (GYRO), magnetometer (MAG), Neural Network (NN), Hidden Markov Model (HMM)

Nr	Study ref	Year	Utilized sensor			Sampling frequency	Commercial/ Custom made	Sensor Raw Data Filtering	Data Processing
			ACC	GYRO	MAG				
1	[32]	2017	3-axes; ±16 g	3-axes; ±2000 °/s	-	256 Hz	NOT MENTIONED	PRESENT	Finding the first derivative of the acceleration (Jerk), which peaks identify HSs
2	[33]	2017	3-axes; ±8 g	-	-	128 Hz	Commercial	NOT MENTIONED	Composite acceleration
									Acceleration peaks identification using thresholds
									HS and TO detection based on the stride duration approximation using data from the accelerometer
									Acceleration wavelet transform
								Symbol based method that uses piecewise linear segmentation followed by clustering to symbolize the 2-axis acceleration signal	

Table 2.3 continued

Nr	Study ref	Year	Utilized sensor			Sampling frequency	Commercial/ Custom made	Sensor Raw Data Filtering	Data Processing
			ACC	GYRO	MAG				
									Domain knowledge based on the gait cycle and gait event frequency detection using data from the accelerometer
3	[34]	2017	3-axes; $\pm 160 \text{ m/s}^2$	3-axes; $\pm 1200 \text{ }^\circ/\text{s}$	-	50 Hz	Commercial	PRESENT	Acceleration relative minimum peaks
									Acceleration relative minimum peaks for TO and mid-swing; HS as a linear interpolation between the sample above the zero rate and the sample below the zero rate just after the mid-swing events
									Acceleration continuous wavelet form
									Scalar continuous Hidden Model Markov with data from the gyroscope
4	[35]	2017	3-axes; $\pm 160 \text{ m/s}^2$	3-axes; $\pm 2000 \text{ }^\circ/\text{s}$	-	50 Hz	Custom	RAW DATA ANALYSIS	Angular velocity noise zero-crossing method which is based on the inverted pendulum model of the gait
5	[31]	2018	3-axes; $\pm 8 \text{ g}$	3-axes; $\pm 1000 \text{ }^\circ/\text{s}$	-	285 Hz	Commercial	NOT MENTIONED	Acceleration peak identification and zero-crossing
									Acceleration/ Angular velocity continuous wavelet transform
									Angular Velocity peaks identification

Table 2.3 continued

Nr	Study ref	Year	Utilized sensor			Sampling frequency	Commercial/ Custom made	Sensor Raw Data Filtering	Data Processing
			ACC	GYRO	MAG				
6	[26]	2018	3-axes; ±16 g	3-axes; ±2000 °/s		148 Hz	Commercial	NOT MENTIONED	Angular Velocity peaks identification
7	[28]	2018	3-axes; ±8 g	3-axes; ±500 °/s	-	50 Hz	Commercial	PRESENT	Acceleration peak identification for mid-swing
8	[36]	2018	3-axes	3-axes	-	150 Hz	Commercial	PRESENT	Heuristics and zero-crossing method based on the angular velocity
									Gait sequence and peak angular acceleration
									Acceleration maximum and minimum peaks
									Acceleration maximum peak and zero-crossing
									Angular velocity wavelet decomposition
									Single-axis angular velocity thresholds taken from the baseline signal
Manual identification by visual inspection of gyroscope signal									
9	[27]	2018	3-axes	3-axes	-	100 Hz	Commercial	NOT MENTIONED	Angular Velocity peaks identification
									Angular velocity peaks identification and norm Acceleration of the feet
10	[25]	2019	3-axes	3-axes	-	400 Hz	Commercial	RAW DATA ANALYSIS	Acceleration and Angular velocity rule-based sliding window method (peaks identification, flat-zone detection, zero-crossing) that was fed to NN\HMM hybrid model
11	[37]	2020	3-axes	3-axes	-	128 Hz	Commercial	PRESENT	Angular velocity negative peaks

2.3 Literature review conclusions

Through the selected keywords a total of 34 research papers have been analyzed and summarised in this literature review. The main outcome of the conducted survey is the general lack of standardization regarding protocols, devices, and analysis methods applied. Furthermore, the following conclusions can be made:

- Shanks and foot are the most suitable locations for comprehensive gait analysis, however, exact placement should be considered based on the purpose of the application and the population investigated.
- Straps are preferred as a mounting method for better fixation.
- Both angular velocity and acceleration are suitable for identifying gait events.
- Peaks identification method produces agreeable results in comparison to the gold standard reference systems.

More uncertainty remains in terms of the necessity of data filtering, optimal sampling frequency, and algorithms for automatic gait events extraction.

2.4 Aims of the Thesis

The main aim of this MSc thesis is to develop and validate an automatic gait events detection system utilizing TalTech's miniature wearable IMU. The developed system is aimed to be utilized for gait analysis in clinical applications. To achieve that, the following sub-goals have been outlined:

- Develop a sensor mounting system for the lower limb of the human body that will assure firm fixation and convenient utilization.
- Validate the sensor for gait analysis versus a camera-based approach.
- Develop a data filtering system to limit noise and soft-tissue artifacts.
- Develop and implement an automatic gait events identification algorithm.
- Implement the algorithm on the study group to validate the developed system.

3. MATERIALS AND METHODS

This chapter contains a complete overview of the data-acquisition set-up, materials, and tools that were involved, as well as a description of the methodology, data processing and exploited statistical methods. The sections of this chapter were organized according to the actual timeline of the thesis project. Since outcomes of this thesis might be used in conferences and future works, for the entire content of the manuscript a *dot* instead of the *comma* was used as a decimal value indicator.

3.1 Materials and Tools

This section outlines the materials and tools used in this master thesis. Among them are the TinyTag motion sensor, the custom casing that was designed and 3D printed, along with materials used for filter development.

TinyTag

TinyTag is a small motion and pressure logger for human and animal gait analysis developed by TalTech's Centre for Environmental Sensing and Intelligence, Department of Computer Systems. This device is designed to be used in hospitals, rehabilitation clinics, and outdoors. In its essence, it is an IMU device that also has pressure and temperature sensors. Due to lack of the communication systems, it is equipped with a removable SD card for data storage and a lithium-polymer battery for power (see Figure 3.1). Battery size and capacity can be adjusted according to the objectives of the performed task. The sensor also has a real-time clock that provides the timestamp of every measurement once the sensor is activated. Sensor activation is done through the magnetic switch. Charging and data transfer are carried out through 4 pins in the head of the sensor.

Dimensions of the sensor and utilized battery are captured in Figure 3.2. Together, their weight results in 5.2 grams. Once switched on, it records acceleration, angular velocity, and magnetic field in 3 axes, as well as pressure and temperature. However, for gait analysis, only accelerometer data were utilized. Acceleration was recorded at the frequency of 100 Hz, with an accelerometer set to the range of ± 8 g with the sensitivity of 1/4096 g. The sensor stores the data as a *.txt* file.

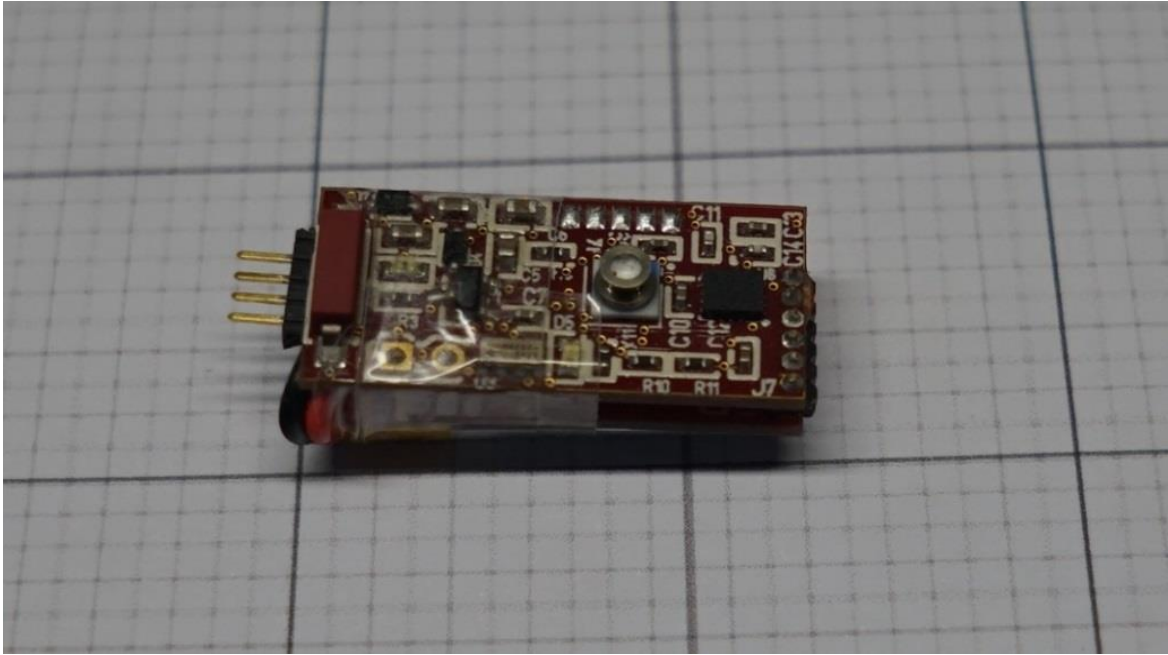


Figure 3.1 Image of the final embodiment of the device, after assembly and testing

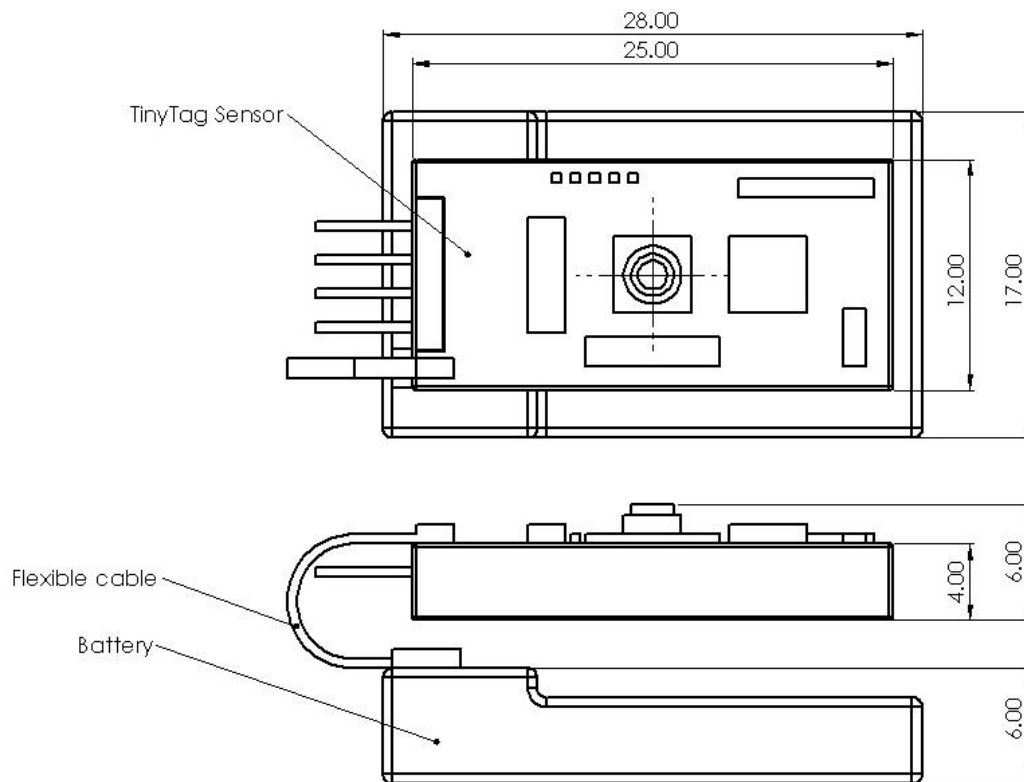


Figure 3.2 TinyTag and Lithium battery dimensions, all dimensions are given in millimeters

Sensor casing design and mounting system

As it was concluded in the literature review, a mounting system with straps is preferred over adhesive double-sided tape or any other tape-alike material. Straps

allow more firm attachment of the sensor and enable quick relocation of the sensor, within one experiment, from one mounting location to the other. Therefore, the mounting system consisted of a custom casing, designed and later 3D printed, and a strap.

For design, Solidworks 2021 (Dassault Systèmes, France) was used. The complete design is shown in Figure 3.3. The main dimensions of the casing are depicted in Figure 3.4. During the experiments, as will be better described later, the sensor had to be switched on at the beginning of every trial and switched off at the end. For that, the casing was purposefully left uncovered since the sensor is turned on with the magnetic switch.

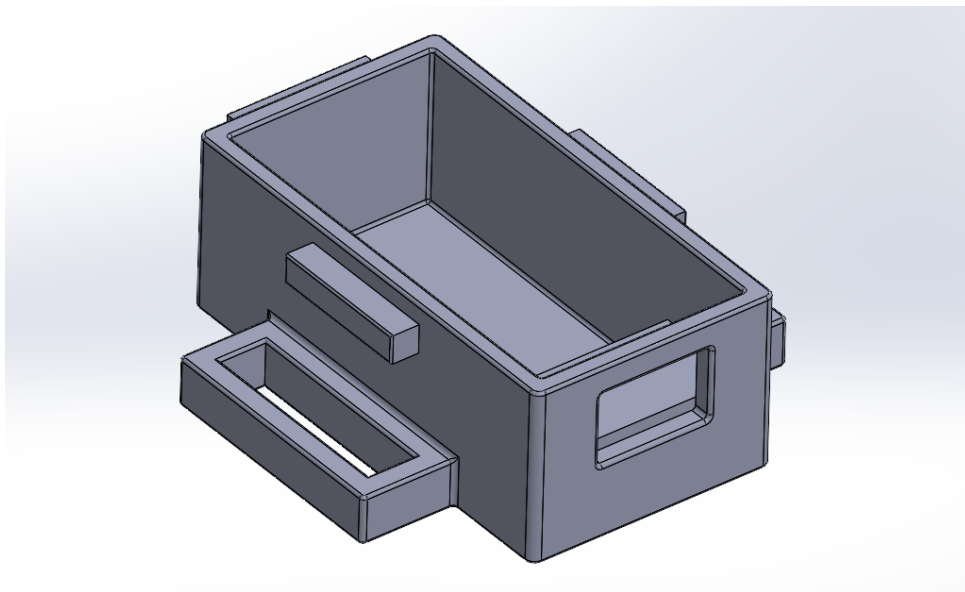


Figure 3.3 Sensor's casing design

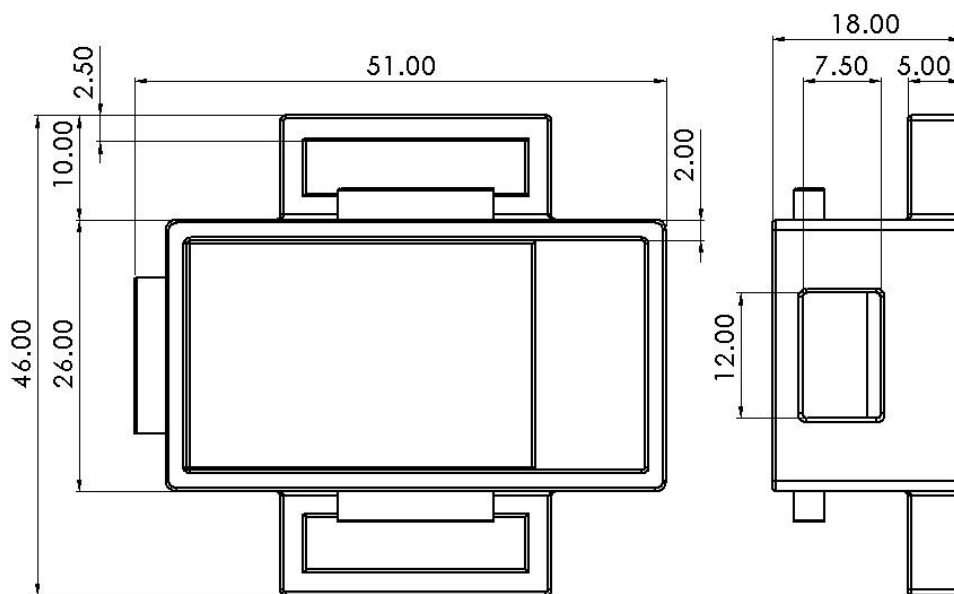


Figure 3.4 Sensor's casing main dimensions, all dimensions are given in millimeters

Once designed, the casing was 3D printed with the Durable Resin [38]. This material allowed keeping the mass low whilst securing the required precision and casing's toughness despite the tiny dimensions (51 x 46 x 18 mm).

The sensor was installed inside the casing with double-sided tape, which, given the low mass of the device, was sufficient to secure firm fixation. Additionally, USB Micro-B Breakout Board [39] was installed in the casing with two 4 mm screws and soldered to the pins of the TinyTag. This allowed convenient and time-efficient charging and data collection from the sensor via Micro USB without removing it off the casing.

The last step to complete the mounting system was adding the straps. For this, Velcro straps, with a width of 20 mm, were fixed to the casing body with metal clips. The final version of the system used in this master thesis can be seen in Figure 3.5

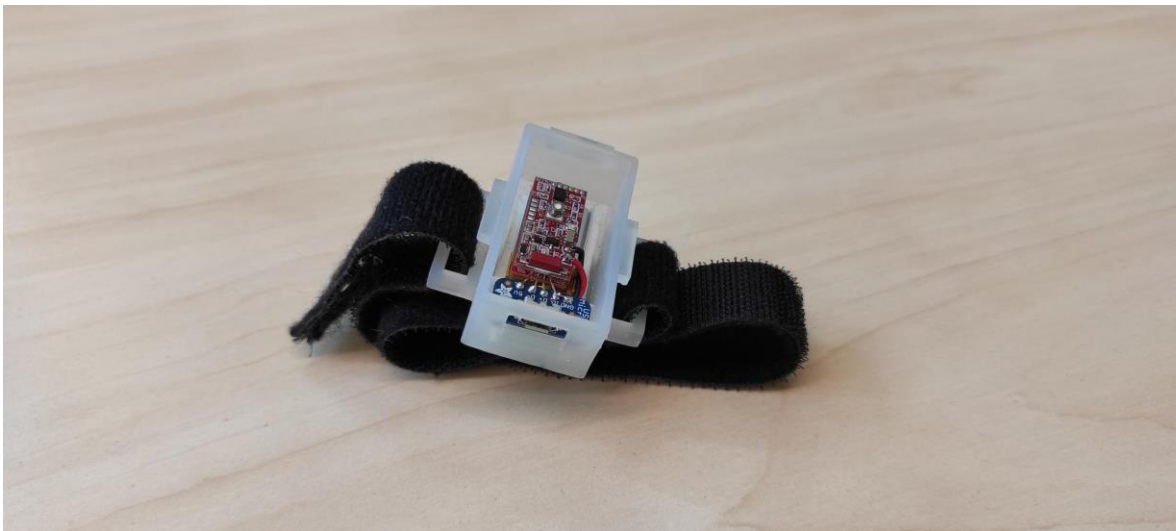


Figure 3.5 Final assembly of the sensor's casing

Buffer materials for mechanical damper

At first, in order to develop a data filtering system, an approach with a mechanical damper was considered. Buffer materials were fulfilling the role of the damper, aiming to reduce the noise in the high-frequency signal and limit soft-tissue artifacts (skin movement). For that, three different materials were used, with two distinct thicknesses for each. Two of them were silicon-based compounds Zhermack Elite Double 16 [40] and Zhermack Elite double 22 [41]. The remaining third material utilized in the experiments was Sorbothane, which is a synthetic viscoelastic urethane polymer with remarkable shock absorption properties and a very high damping coefficient [42], [43].

According to the datasheets available on the Zhermack website, Zhermack Elite double 16 has the hardness of 16 on the Shore A scale (see Figure 3.6), and second Zhermack silicon has the hardness of 22 on the same scale. As to the Sorbothane, according to the manufacturer’s datasheet, the 2.5mm and 6.4mm Sorbothane sheets have the hardness of 40 and 30 on the Shore 00 scale respectively. Materials that belong to Shore 00 hardness scale have a spring force of 113 g, whereas materials from Shore A on the same scale have a spring force of 822 g [44].

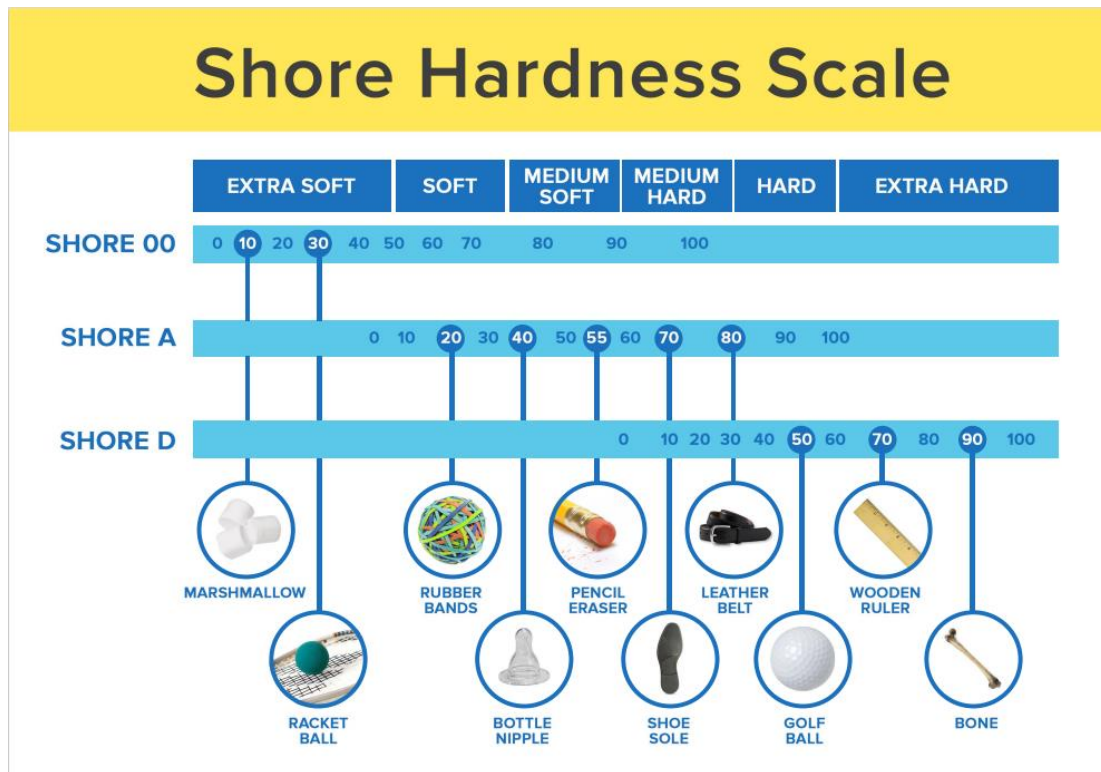


Figure 3.6 Share Hardness Scale [44]

Whilst Sorbothane is commercially available in the shape of sheets of different thicknesses, Zhermack Elite double products are distributed as liquids and have to be cast to obtain a solid form. For that, custom molds were designed and 3D printed (see Figure 3.7), using the same material as for the casing.

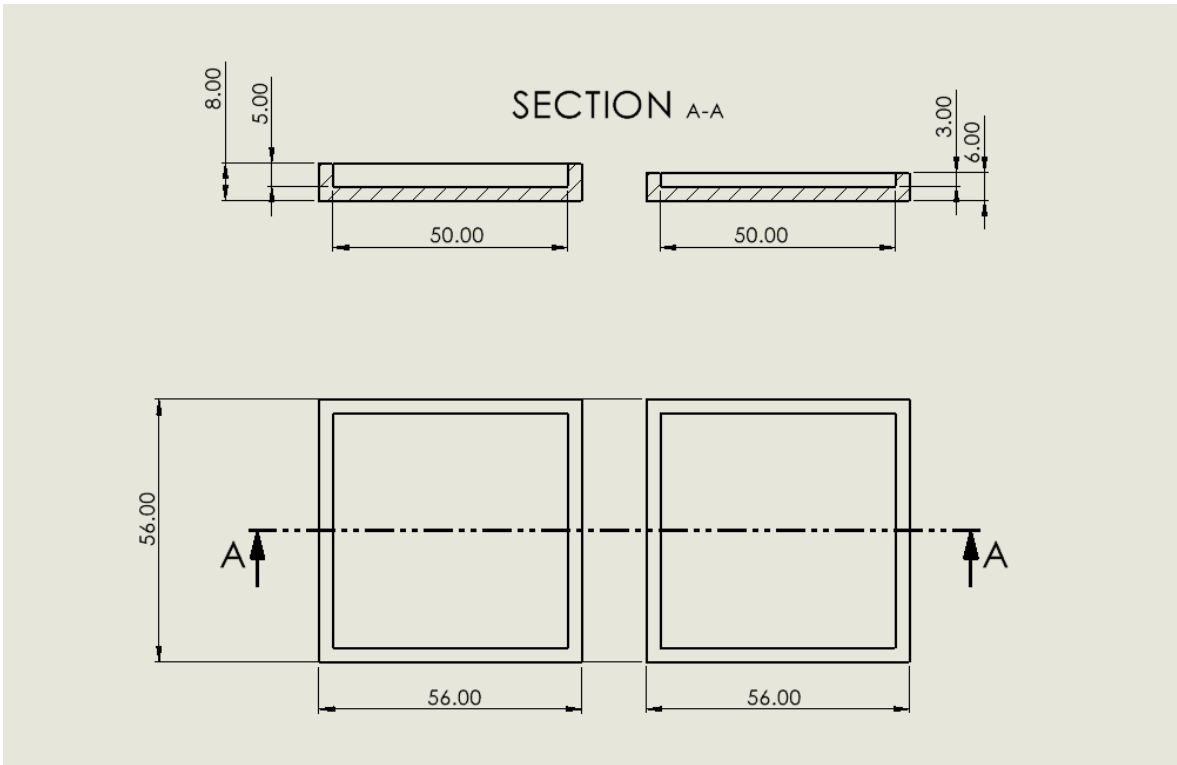


Figure 3.7 Designs of the molds, all dimensions are given in millimeters

Once molds were printed, Zhermack silicones were cast in two different thicknesses of 3 and 5 mm. As to the sorbothane, the sheets of thickness 2.5 mm and 6.4 mm were used, which allowed producing rectangular pads matching the dimensions of the mold. Concluding, three different materials with two thicknesses were prepared for testing the efficiency of a mechanical filter approach (see Figure 3.8).



Figure 3.8 3D printed molds and materials used in the Pilot study 2

1 – Sorbothane (2.5 mm), 2 – Zhermack ED 16 (3 mm), 3 – Zhermack ED 22 (3 mm), 4 – Zhermack ED 16 (5 mm), 5 – Zhermack ED 22 (5 mm), 6 – Sorbothane (6.4 mm).

3.2 Methodology

In order to target the research questions of the master thesis, the data acquisition process was based on collecting n – number of repetitions of the selected task, where n depends on the stage of the project. Tests were performed with the sensor placed at three different locations, described in detail below, on the right lower limb.

Throughout this master thesis three data acquisition processes were carried out, strictly connected to a specific aim of the project:

- 1) Data acquisition for the TinyTag validation for gait analysis purposes versus camera-based method – Pilot study 1
- 2) Data acquisition for evaluation of dumping materials efficiency – Pilot study 2
- 3) Data acquisition from the study group of 10 people to evaluate the algorithm performance – Experimental study

The summary of all experiments performed, specifying involved equipment, the number of subjects, and the amount of repetitions is presented in Table 3.1.

Table 3.1 Summary of the data acquisitions that were performed with involved measurement systems, number of participants, and number of repetitions

Data acquisition	Systems		Subjects	Repetitions		
	IMU	Camera		Ankle	Shank	Foot
Pilot study 1	x	x	1 male	5	5	5
Pilot study 2	x		1 male	30	30	30
Experimental study	x		5 male	50	50	50
			5 female	50	50	50

Whereas three separate data collection processes were performed, to keep continuity and assure gait pattern repeatability, the set-up shown in Figure 3.9 was implemented for every experiment. It consisted of 4.5 meters walkway, the beginning, and end of which were marked with double-sided tape, and a 40 cm gap before the beginning mark of the walkway, also marked with a tape.

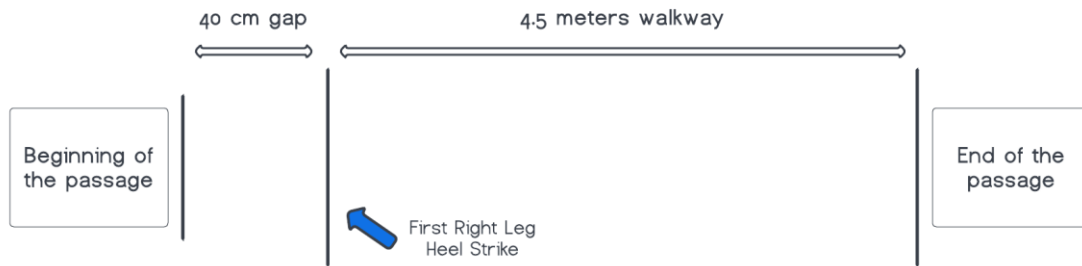


Figure 3.9 Data acquisition set-up

The subject was initiating each data acquisition whilst located at the beginning of the gap (see Figure 3.9 for better visualization). After turning on the sensor, the subject was starting to walk with the left leg and continued towards the end of the walkway at the self-selected pace. Once reaching the end of the passage subject was asked to turn off the sensor. Within each dataset, four full steps were completed, three of which (the first 3 complete gait cycles) were the area of interest for further analysis. For simplification, within this master thesis, a complete gait cycle duration (i.e. stride time) is considered to be a period of time between the TO and the consequent TO of the same lower limb.

During all the experiments data were collected from three different locations – shank, ankle, and top of the foot (further referred to as foot). These locations were chosen based on the outcomes of the literature review, as they provide the most suitable data for comprehensive clinical gait analysis.

On the shank, the sensor was located at the mid-point between the head of the fibula and the lateral malleolus (see Figure 3.10 and Figure 3.11). To position the sensor in the mid-point, measurements were done for each subject before the beginning of the experiment. On the ankle, the sensor was placed above the lateral malleolus (see Figure 3.12). Lastly, the sensor was located on top of the foot, over the laces (see Figure 3.13).



Figure 3.10 Distance measurement for the sensor positioning on the shank



Figure 3.11 Sensor attached to the shank with arrows representing axes of the accelerometer

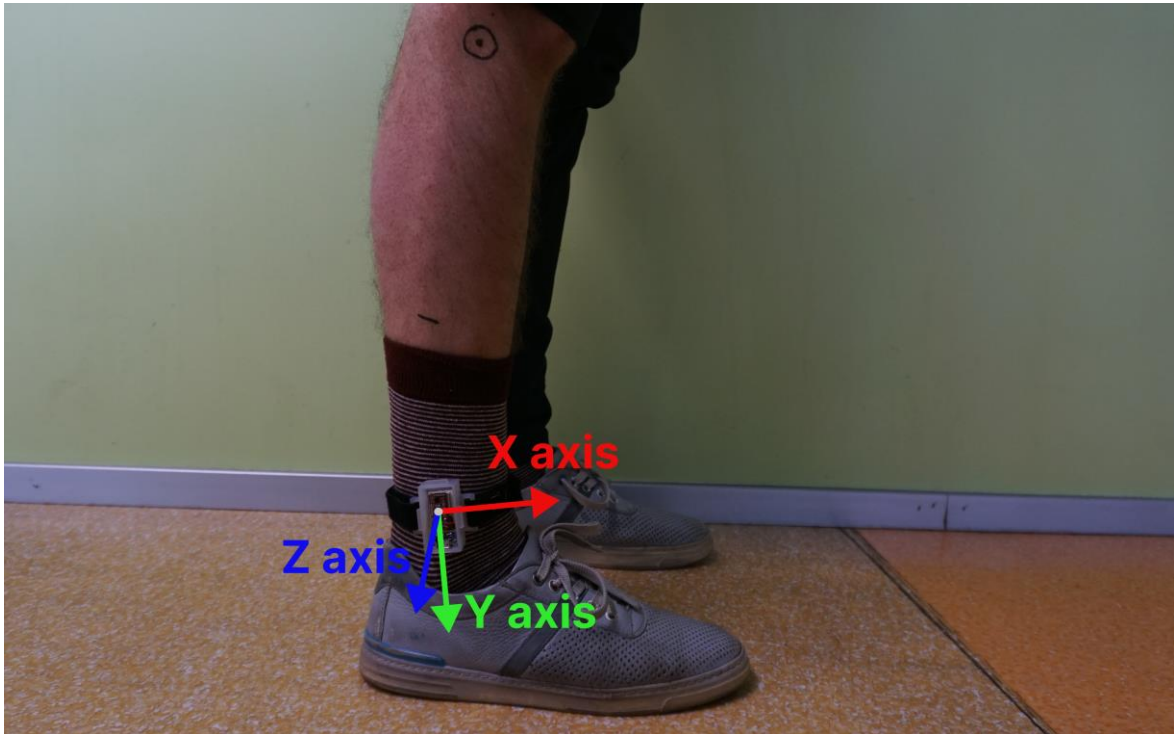


Figure 3.12 Sensor attached to the ankle with arrows representing axes of the accelerometer



Figure 3.13 Sensor attached to the foot with arrows representing axes of the accelerometer

Pilot study 1

As mentioned in section 2.4, the first objective of the thesis was to validate the TinyTag sensor's suitability for gait analysis. For that, a camera-based method was chosen as a ground truth reference. The set-up of this study is presented below (see Figure 3.14) and consisted of the set-up shown in Figure 3.9 with the addition of the

camera. The video recordings were made with Sony ALPHA 6000 (24.3MP, 30fps). The camera was located at the distance sufficient to capture all 4 strides, at the height of 80cm. Every experiment was initiated by turning on a camera and sensor. Once both were recording, the subject was asked to walk at a self-selected speed until the end of the passage and switch off the sensor. After that, the camera recording was also stopped.

For Pilot study 1, five repetitions with the sensor placed at each of the locations were made by a single subject. Thus, 15 datasets were collected in total.

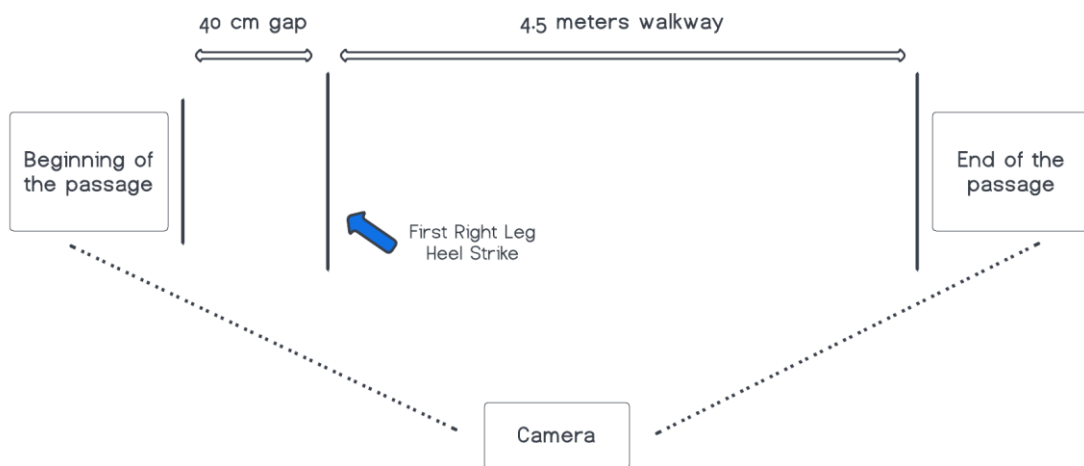


Figure 3.14 Set-up for the Pilot study 1

Validation of the system was made based on the comparison of two parameters:

- 1) The ability of the sensor to record all the major gait events (heel strikes and toe-offs) occurrences

Since the number of gait events was predefined by the experiment set-up, it was manually checked how many gait events were properly identified.

- 2) Durations of the three complete gait cycles

Gait cycle durations were calculated for both measurement systems. The recordings from the camera were processed with Kinovea [45] (version 0.9.5). Kinovea is the standard software for frame-by-frame video processing that allowed extracting the timestamps at which consequent gait events occurred. The results were summarised in the Excel table to further calculate strides duration.

Based on the timestamps of the gait events the stride duration, or the duration of a single complete gait cycle, was calculated for each repetition. As mentioned before, the duration of the single gait cycle is a period of time from the TO event to the consequent TO event of the same leg. Thus, the step duration (SD) was calculated following the equation:

$$SD = TO_{i+1} - TO_i, \quad (3.1)$$

where: SD – stride duration, TO – toe-off gait event, i – number of gait event.

Next, a similar procedure was completed for accelerometer data. To estimate the strides duration from TinyTag's acceleration data, it was necessary to pre-process the '.txt' file obtained from the sensor in Excel, combining acceleration in all three axes collected by the sensor included in the logger. For that, total acceleration or the magnitude of the acceleration vector was calculated following the formula:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}, \quad (3.2)$$

where: a – magnitude of the acceleration vector, a_x – acceleration on x-axis, a_y – acceleration on y-axis, and a_z – acceleration on z-axis.

The Excel files were afterward processed with the open-source scientific environment Spyder IDE [46]. To identify peaks corresponding to the gait event of interest a custom Python code was prepared. The peaks were identified with the "find_peaks" function from the "scipy.signal" add-in library. For this study threshold for the peaks was specified manually for each dataset.

Timestamps obtained from the data processing with Spyder IDE were captured in centiseconds, therefore they had to be converted into seconds for further analysis. Thus, equation 3.1 for stride duration calculation transformed into:

$$SD = \frac{TO_{i+1} - TO_i}{100}, \quad (3.3)$$

where: SD – stride duration, TO – toe-off gait event, i – number of gait event.

Having SD calculated with two methods, the comparison was made by finding the difference in the results. For that, SD calculated with TinyTag data was subtracted from SD calculated with a camera-based method.

Pilot study 2

The second data acquisition process was made in order to collect the data for mechanical damper development. For that, 3 different materials, with 2 distinct thicknesses were used, along with the TinyTag sensor. The sensor was attached to the subject's lower limb using the custom-made casing described in 3.1 (Figure 3.5). The data-acquisition set-up was the same as described in section 3.2. All the repetitions were made by the same person, to keep the consistency of the gait pattern. Having that, there were 6 batches of 15 datasets (5 repetitions per sensor location) for each material mentioned in chapter 3 (Buffer materials for mechanical damper). Thus, in total there were 90 datasets.

The data was processed similarly as described in the previous subsection. First, the magnitude of the acceleration vector was calculated following equation (3.1) and furtherly processed with Spyder IDE to identify the peak occurrences corresponding to the gait events.

To evaluate the impact of the buffer material on the desired outcome the “*number of peaks per single dataset*” approach was taken. Meaning, that the number of peaks identified in data with buffer material was compared to the number of peaks in datasets without a buffer material (referred to as control study). Since within one single experiment three complete gait cycles were studied, ideally, a single dataset should consist of seven consequent peaks. Four corresponding to the TO events, and three to the HS gait events.

Experimental study

Lastly, in order to test the performance of the developed gait events identification algorithm 10 volunteers were involved. The group of 10 volunteers consisted of five male and five female subjects. A summary of the demographic parameters of the study group is provided in Table 3.2. Each volunteer of the study group had no previous history of injuries in the lower limbs.

Table 3.2 Demographic parameters of the study group

Gender	Age	Mass	Height
5 male	23 - 37	65 - 92 kg	165 - 183 cm
5 female	28 - 39	55 - 65 kg	162 - 170 cm

All subjects followed the set-up described in 3.2 and performed 10 repetitions per every sensor location. This resulted in 30 datasets per person and consequently 300 datasets in total.

Every dataset was then processed in Excel to calculate the magnitude of the acceleration vector (following equation 3.2). Then the gait events identification algorithm was applied.

Since the set-up of every experiment was predefined by a concrete number of steps made by every subject, the efficiency evaluation of the algorithm was based on the number of gait events identified correctly by the algorithm. All three locations were

studied separately to evaluate whether the sensor location influences the reliability of the algorithm.

3.3 Numerical filter

Based on the results of the mechanical damper approach the numerical filter method was also considered. This approach is oriented toward manipulation on the dataset at the post-processing stage, rather than impacting the data acquisition process itself. For that, Savitzky-Golay (S-G) data smoothing filter was applied. S-G is a digital filter that is meant for smoothing oversampled signals corrupted by noise. This filter is widely used in signal processing as it allows a significant smoothing of noisy data while preserving the height and width of the original signal. It is also commonly applied in gait analysis [28], [47], [48]. S-G filter's smoothing effect is achieved by fitting the subsequent data sets of a certain window length with the low-degree polynomial function. Then by convolution of these polynomials, the smoothed signal is achieved. [49]–[51]

Since this filter is broadly used in data processing, it is available in Spyder IDE as a part of the signal processing library "*scipy.signal*" and is called "*savgol_filter*". When implemented, as a function input it requires a dataset to be analyzed, along with 2 parameters: window length and order of the polynomial. To achieve meaningful results these two parameters have to be carefully considered; therefore, an additional study was executed in order to find suitable parameters.

At first polynomial order was chosen experimentally. Linear and quadratic functions turned out to be insufficient for smoothing accelerometer data. Third-order polynomial, on the other hand, proved to be a suitable choice. The cubic function assures desirable data smoothing without losing important information. Thus, for all the data processing with the S-G filter, a third-order polynomial was chosen.

Differently, to evaluate the influence of the window length on the filtered data, the distance between the peak occurrence (on the timeframe axis) in the raw data was compared to the occurrence of the same peak in the filtered data with different window sizes. The result of subtraction of the filtered timestamp from the original peak timestamp was calculated for several window lengths. The window length parameter for the final algorithm was chosen based on the findings from this study.

To evaluate the efficacy of the numerical filter, a similar method as in Pilot study 2 was considered, examining the number of peaks per dataset.

3.4 Automated gait events identification algorithm

This section contains a detailed explanation of the idea behind the automated gait event identification algorithm as well as tools that were used to achieve it. At first, the following subsections cover theoretical aspects of the utilized tools, and then the algorithm's logic is described.

Matrix profiling

Matrix profiling is a novel approach for automated time-series segmentation. It is based on the fundamental concept of "conservation of pattern", which means that events in a time series that are meaningful and can be segmented tend to be self-similar. The idea was first introduced in 2016 by Eamonn Keogh and colleagues [52], and has been intensively studied and applied in a broad range of research and industrial applications including finance, animal and human kinematics, and genomics. In its essence, the matrix profiling algorithm exploits the idea of comparing the Euclidian distances between two subsequences of a fixed length (motifs) in the given time series. The reference subsequence motif can be generated by recursively increasing the length of the selected time series, or it can be predefined by the user. In this work, the motif was manually defined. The parameters of the motif are the *signal length*, which describes the time frame of the motif, and the *shape*, which describes how data changes within this time frame. The algorithm then returns an array that consists of the Euclidian distance between a given motif and corresponding subsequence at every point in the studied dataset. Extracting the lowest value from this array then will point to the subsequence in the analyzed dataset that is nearest to the motif it was compared to. Extracting the highest value of Euclidian distance will point to the discord or anomaly. In other words, the piece of signal that is unlike any other in the given dataset [53]–[55].

An example of a motif (orange) and a discord (red) taken from an example gait time series used in this study is shown in Figure 3.15. As can be seen from the figure, the "Motif" is repeated several times, whereas the "Discord" is unique and is not repeated.

This principle was exploited in this master thesis, since the human gait has a highly repetitive pattern, especially when multiple datasets of a single subject are analyzed.

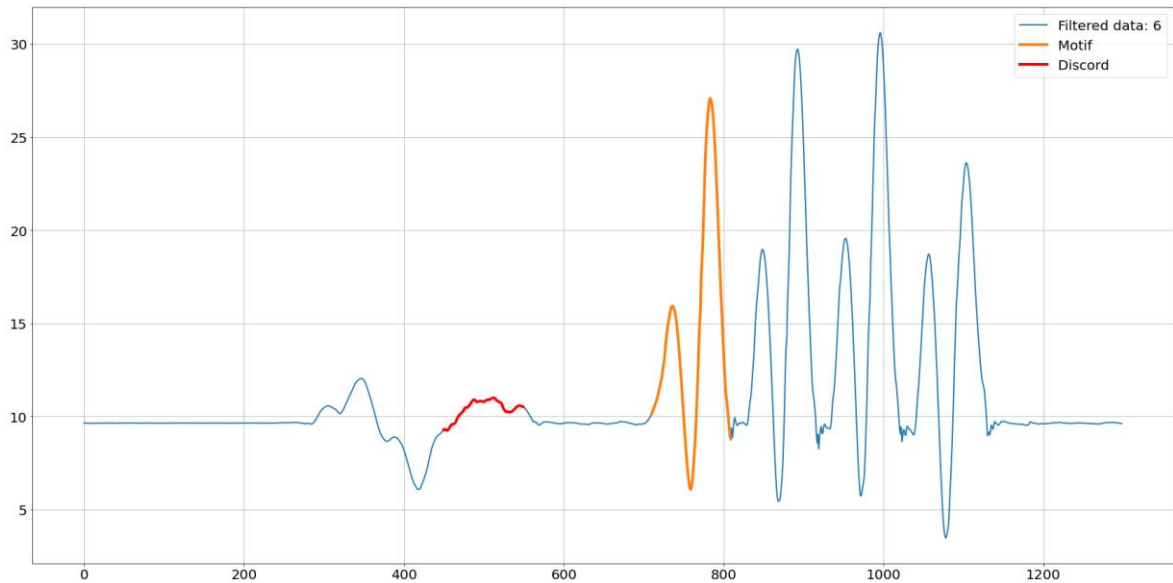


Figure 3.15 An Examples of a “Motif” and “Discord” found in the real-world human gait data

Peak identification

As it was concluded in the literature review, peak identification is the most common approach in gait analysis for identifying gait events. Therefore, this principle was also exploited in this master thesis since a major outcome of the automated gait events identification algorithm has to be occurrences of the gait events within a given dataset. For peaks identification algorithms a threshold value is usually required. Within this master thesis, different threshold values were used in the final version of the algorithm. The threshold value was mainly based on the sensor location of the studied dataset.

Novel gait analysis data processing

For this thesis work, the newly developed automated data processing method was used. It included 3 data processing algorithms:

- 1) Filtering the raw data with the S-G data processing tool
- 2) Matrix profiling, for highlighting common gait patterns (motifs) in the separate datasets
- 3) Peaks identification, for specifying time occurrences of the gait events in analyzed datasets

In contrast to existing methods which require extensive user inputs, the proposed method in this work allows very quick analysis of the large massifs of data. The logic behind the algorithm was following the waterfall implementation of the ideas described above. At first, data were filtered with pre-set S-G parameters and checked with matrix profiling to identify 3 consequent strides in the analyzed dataset. Once done,

the three steps, that were the area of interest of the study, were separated from the whole dataset. Then the peaks identification function, which also comes as a built-in library in the Spyder IDE and is called “*find_peaks*”, was applied to the separated dataset. The threshold value was specified based on the analyzed location: for the sensor located on the shank, it was equal to 12 m/s², and for the ankle and foot 13 m/s².

The same algorithm was applied separately to all 3 batches (one batch per location) of data from a single study subject. Two different approaches of matrix profiling were tested, namely – *General Motif* and *Subject-Based Motif*. Within this master thesis, a *Motif* is identified as a signal that represents a human gait pattern. Since the purpose of the research was to find and identify three consequent strides, or seven gait events, within its boundaries, the motif included three strides.

For the general motif approach, the motif was defined manually based on the single dataset from Pilot study 2. Every location had a separate motif. Then, once applied to the analyzed dataset, the function looked for the portion of the gait cycle that is most similar to the predefined motif. The outcome of the algorithm was a graphical representation of the found motif overlaid over the raw data, along with an array of gait events timestamps and a number of peaks found in the identified gait pattern match (Figure 3.16).

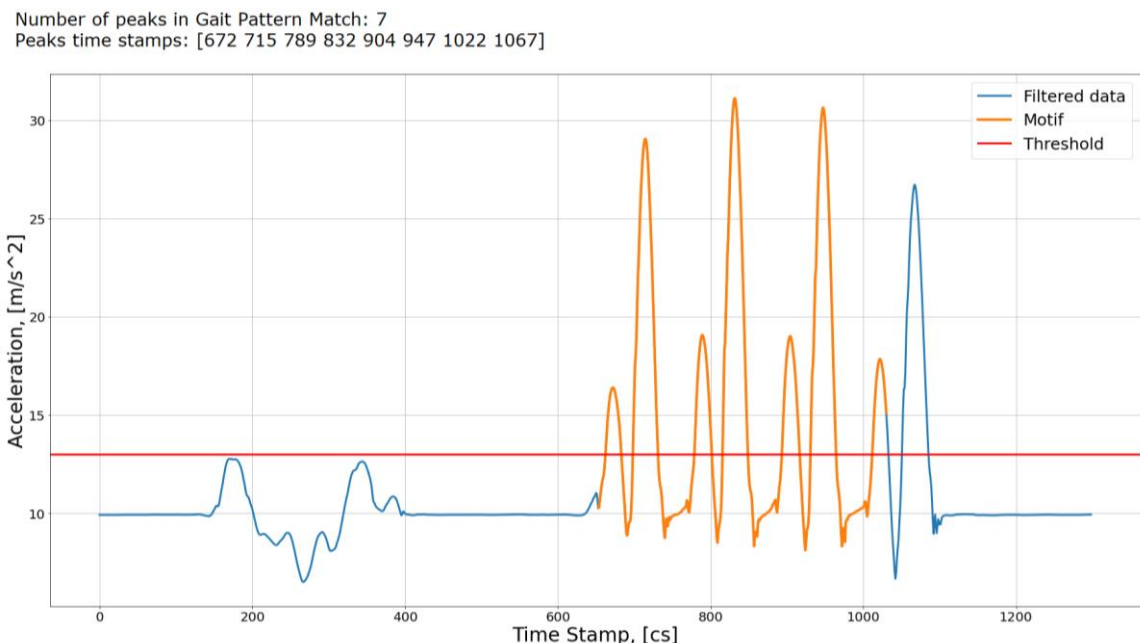


Figure 3.16 An example of a general motif approach output

For the second approach, the motif was specified for each subject separately, based on the first dataset from the batch of the studied subject. Once the program is executed, the first filtered dataset of the currently analyzed batch is graphically

presented to the user, along with peaks identified within an analyzed dataset (Figure 3.17). The program then asks the user to input two values, which will define the beginning and the end of the motif. Once these values are introduced, the algorithm runs through all datasets and presents the results in the manner identic to the general motif approach seen previously (Figure 3.16). The efficiency of both approaches was concluded based on whether the algorithm is able to find all seven gait events correctly.

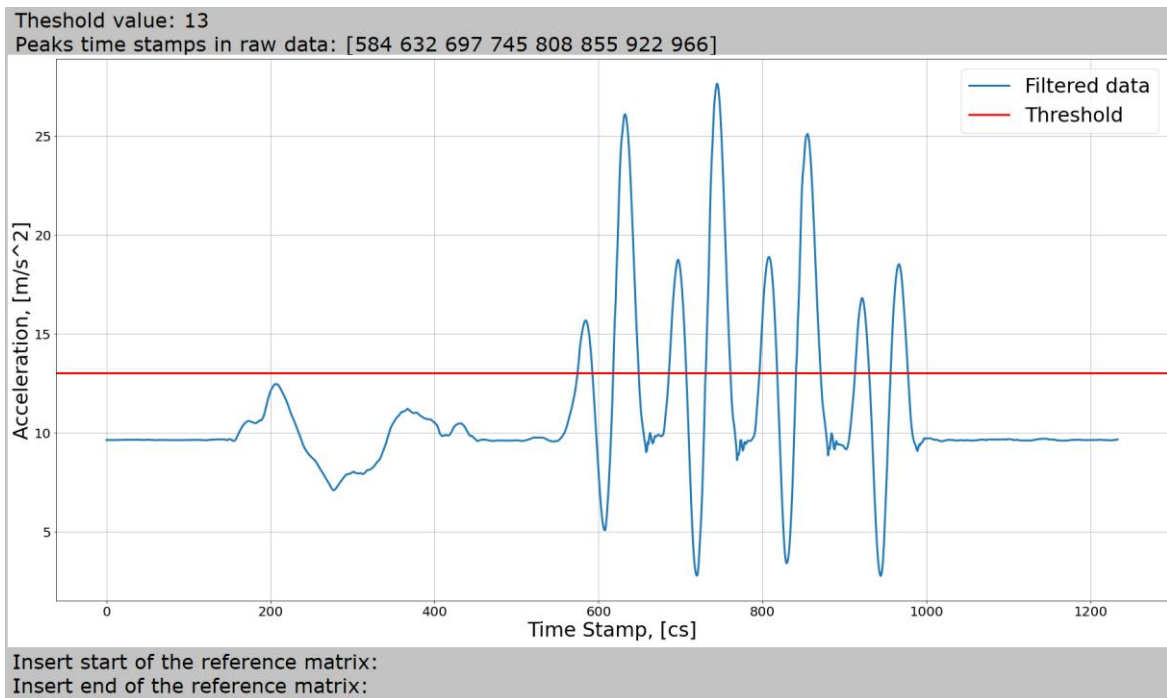


Figure 3.17 An example of a subject-based motif definition process

A schematic representation of both approaches is captured in Figure 3.18.

All the major Python codes that were used throughout this MSc thesis are available on GitHub, following the link given in APPENDIX 2. Along with that, in the same appendix, a code for a subject-based motif method is available as an example.

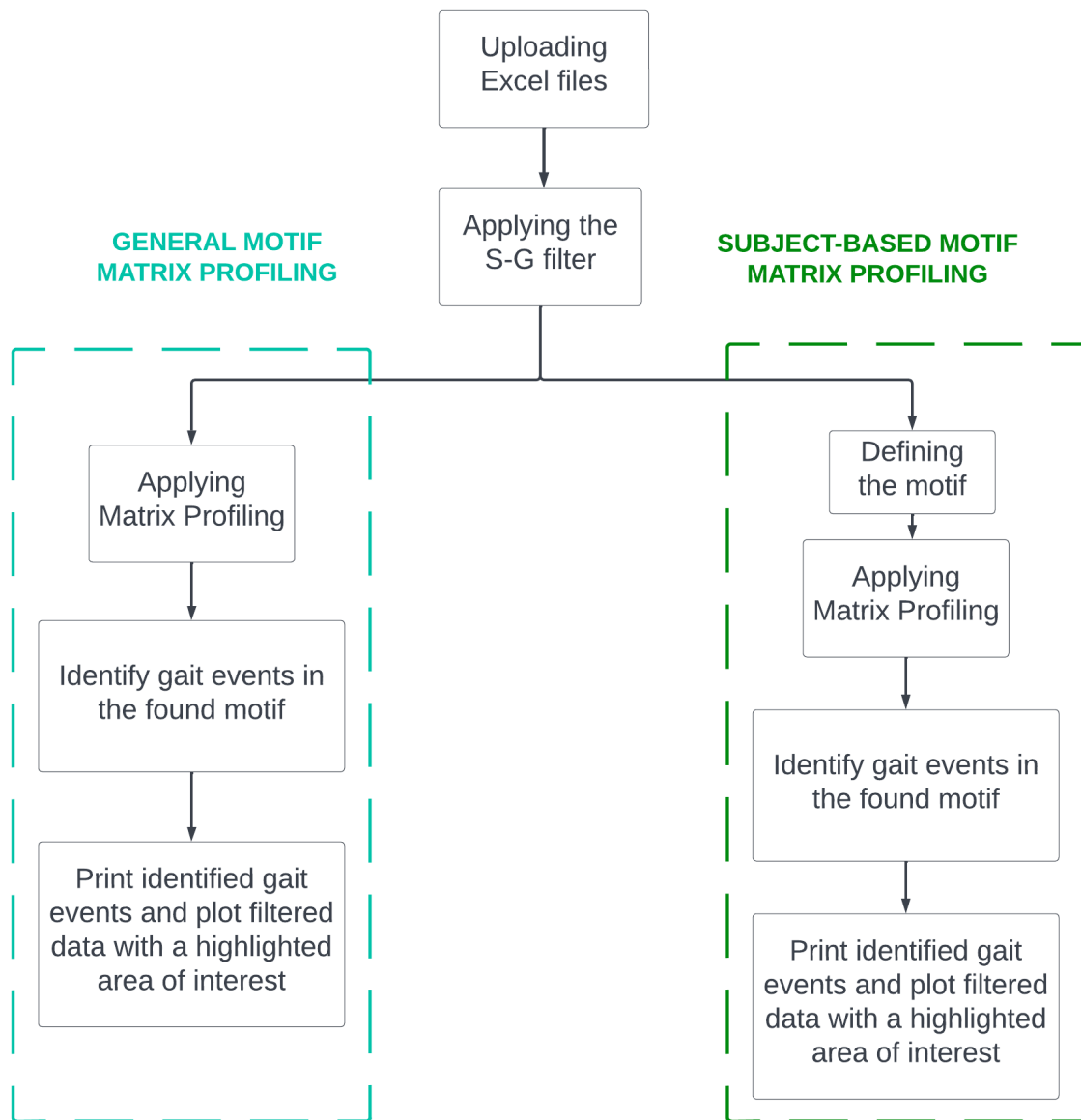


Figure 3.18 Schematic representation of the gait event identification algorithm with *general motif* and *subject-based motif* approaches

4. RESULTS AND DISCUSSION

This chapter presents the results of the studies described in Chapter 3 along with their discussion. Similar to the previous chapter, it is organized in a manner coherent with the timeline of the thesis project.

4.1 TinyTag validation for gait analysis

To validate the TinyTag adequacy for gait analysis evaluation of two parameters was performed. Using recordings from the camera as a reference, the number of HS and TO events occurring was calculated, as well as stride duration.

At first, the occurrence of the events in the videos was labeled, reporting in a table the timestamps at which the certain gait events occurred. An example of the HS is shown in Figure 4.1, and an example of the TO gait event is presented in Figure 4.2.

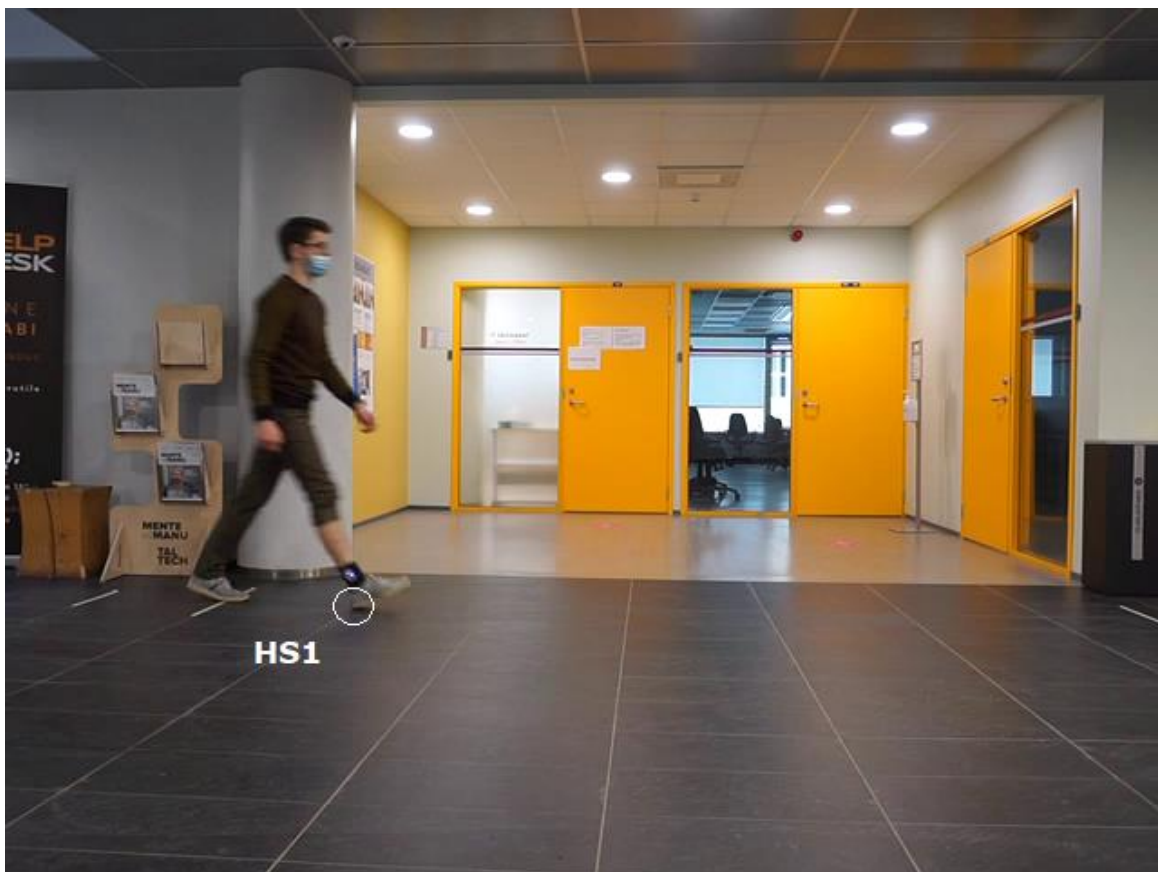


Figure 4.1 An example of the heel strike event of the right leg extracted from the video recordings



Figure 4.2 An example of the toe-off event of the right leg extracted from the video recordings. The results of the video labeling are summarized in the table given in APPENDIX 1. Based on the results, with equation (3.1), stride durations were calculated. The summary of these calculations is presented in Table 4.1.

Table 4.1 Summary of the stride durations calculated based on the video recordings

Sensor location and number of repetition	Stride duration in [s]		
	Stride1	Stride2	Stride3
SHANK_1	1.23	1.14	1.20
SHANK_2	1.20	1.14	1.16
SHANK_3	1.17	1.17	1.20
SHANK_4	1.16	1.14	1.20
SHANK_5	1.17	1.10	1.13
ANKLE_1	1.20	1.10	1.14
ANKLE_2	1.14	1.10	1.13
ANKLE_3	1.13	1.14	1.13
ANKLE_4	1.11	1.10	1.17
ANKLE_5	1.13	1.10	1.17
FOOT_1	1.14	1.10	1.13

Table 4.1 continued

Sensor location and number of repetition	Stride duration in [s]		
	Stride1	Stride2	Stride3
FOOT_2	1.10	1.14	1.13
FOOT_3	1.20	1.10	1.17
FOOT_4	1.14	1.10	1.14
FOOT_5	1.16	1.14	1.13

After that, the same procedure was done for the accelerometer data. Sensor data were processed in Excel to calculate the magnitude of the acceleration vector (equation 3.2). Then the Excel file was imported to the Spyder IDE in order to identify the peak occurrences. For convenience, the x-axes of the datasets were transferred to centiseconds or cs (1cs= 0.01s). An example of such a procedure is shown in Figure 4.3.

Peaks identified in "ANKLE_2" dataset

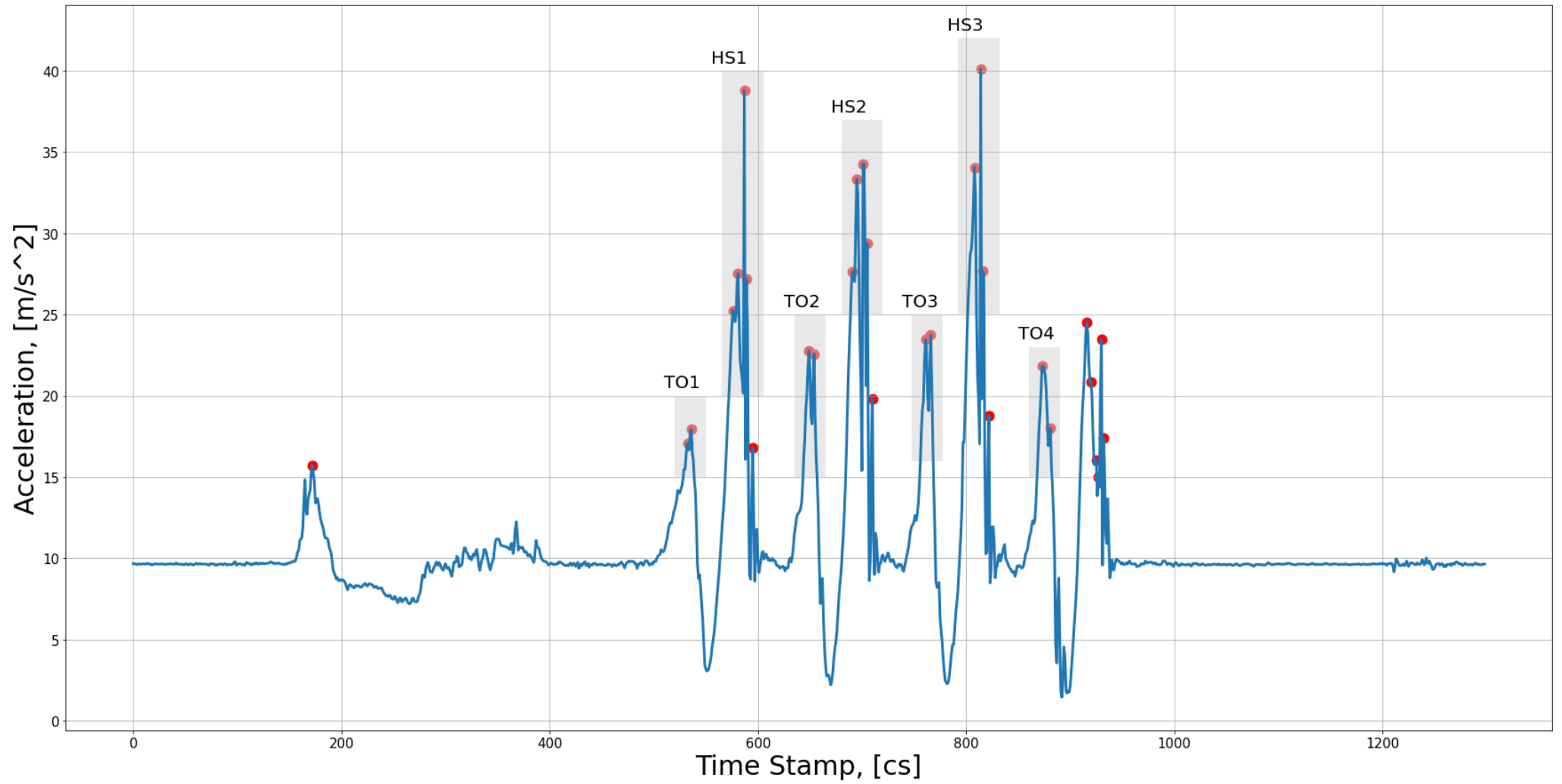


Figure 4.3 Example of the peaks identification process in Spyder IDE

As can be seen from the figure, due to the high frequency and sensitivity of the sensor the raw data from the accelerometer has multiple peaks per single gait event. To the best of the author's knowledge, there is no agreement among researchers regarding which of the peaks should be identified as a gait event. Therefore, within this master thesis, the highest peak was identified as the gait event. The summary of the gait events identified with the TinyTag is presented in the table in APPENDIX 1.

Then, with equation 3.3 stride durations for accelerometer data were found. The results are summarised in Table 4.2.

Table 4.2. Summary of the stride durations calculated based on the TinyTag data

Sensor location and number of repetition	Stride duration in [s]		
	Stride1	Stride2	Stride3
SHANK_1	1.22	1.15	1.18
SHANK_2	1.19	1.16	1.16
SHANK_3	1.16	1.17	1.19
SHANK_4	1.15	1.13	1.16
SHANK_5	1.19	1.11	1.12
ANKLE_1	1.13	1.17	1.07
ANKLE_2	1.11	1.12	1.14
ANKLE_3	1.10	1.11	1.12
ANKLE_4	1.10	1.12	1.12
ANKLE_5	1.09	1.11	1.14
FOOT_1	1.12	1.14	1.08
FOOT_2	1.18	1.03	1.13
FOOT_3	1.14	1.11	1.20
FOOT_4	1.14	1.07	1.17
FOOT_5	1.14	1.13	1.09

Based on the stride duration calculations the comparison between the two approaches was performed, by finding a difference between respective stride durations for each experiment. The results are summarised in Table 4.3.

Table 4.3. Summary of the comparison between the camera and IMU data-based methods for gait events identification

Sensor location and Nº of repetition	Difference in Stride duration [s]		
	Stride1	Stride2	Stride3
SHANK_1	0.01	-0.01	0.02
SHANK_2	0.01	-0.02	0.00

Table 4.3 continued

Sensor location and Nº of repetition	Difference in Stride duration [s]		
	Stride1	Stride2	Stride3
SHANK_3	0.01	0.00	0.01
SHANK_4	0.01	0.01	0.04
SHANK_5	-0.02	-0.01	0.01
ANKLE_1	0.07	-0.07	0.07
ANKLE_2	0.03	-0.02	-0.01
ANKLE_3	0.03	0.03	0.01
ANKLE_4	0.01	-0.02	0.05
ANKLE_5	0.04	-0.01	0.03
FOOT_1	0.02	-0.04	0.05
FOOT_2	-0.08	0.11	0.00
FOOT_3	0.06	-0.01	-0.03
FOOT_4	0.00	0.03	-0.03
FOOT_5	0.02	0.01	0.04
Mean absolute error			0.03

As can be seen from the calculations, on average, the difference between Stride durations calculated based on two different methods is 0.03 s. Since both negative and positive values were occurring, for total mean error an absolute value of the subtraction result was used. Higher deviations were noticed in three datasets (ANKLE_1, FOOT_2, and FOOT_3) with a difference reaching 0.11 seconds in the worst-case scenario for one particular dataset. Despite that, in most datasets, the difference remained very low, around ± 0.01 s.

The frame rate of the camera is 30 FPS, resulting in a camera mean error of 0.033 seconds ($1/30 = 0.033$ s). The calculated mean absolute error of the TinyTag (0.03 s) is lower than the camera framerate. Therefore, the TinyTag sensor is able to collect adequate data for gait analysis in comparison to the camera-based method, the Gold standard method in the field.

4.2 Mechanical damper

As was noticed during the TinyTag accelerometer data processing and can be seen in Figure 4.3, the number of peaks recorded by the sensor per single gait event generally varies from two to five peaks. Having that, in addition to a lack of agreement among researchers concerning which of these peaks should be considered as gait event

occurrence, it was attempted to reduce the number of peaks by adding a buffer material aiming at damping the noise and limiting soft-tissue artifacts and excessive vibrations. As described in section 3.2 (Pilot study 2), in total 90 datasets were gathered using three different materials with two distinct thicknesses. To evaluate the impact of the buffer material on the data acquisition process the *number of peaks per dataset* approach was considered. The summary of the study is provided in Table 4.4.

As can be seen from the table, in the control study, which represents an analysis of the data with no buffer material, around 26 peaks, on average, are found in a single dataset. Knowing that seven gait events are of interest within this study, it results in roughly four peaks per single event. The lowest value was noticed in Ankle data, with about 23 peaks per dataset, whereas the worst results were highlighted in the foot data, reaching around 29 peaks.

Based on the results of the peak identification captured in the table below, it is evident that the use of buffer materials did not introduce any decrease in the number of peaks per single dataset. In fact, it can be noted from the last row of the table, that provides the average of the peaks identified with every material, the number of peaks increased with the addition of the buffer material. Additionally, ankle sensor location provided the lowest number of peaks per dataset in all conditions, averaging this time at around 27 peaks in 30 studied samples. The highest number was instead noted in the shank data, with 30 peaks per dataset on average.

Having these results, it was concluded that the mechanical damper approach is insufficient for the data smoothing and noise reduction that is required for the development of the automated gait events identification algorithm.

Table 4.4 Summary of the mechanical damper development study. Where: Control Study data that was gathered without buffer material, while other columns correspond to studied buffer materials with their thickness given in the brackets.

Material	Control Study	Zhermack ED 16 (3 mm)	Zhermack ED 16 (5 mm)	Zhermack ED 22 (3 mm)	Zhermack ED 22 (5 mm)	Sorbothane (2.4 mm)	Sorbothane (6.4 mm)
Ankle1	23	25	29	26	28	25	29
Ankle2	24	29	26	28	28	25	30
Ankle3	23	26	32	29	28	24	29
Ankle4	24	27	27	25	29	25	26
Ankle5	24	26	30	29	26	24	28
Shank1	21	24	27	32	28	28	27
Shank2	26	29	26	31	32	33	36
Shank3	28	28	29	28	32	32	29
Shank4	25	31	29	32	32	32	30
Shank5	26	29	28	31	32	31	32
Foot1	29	26	26	27	28	26	28
Foot2	30	30	27	27	29	28	29
Foot3	30	28	29	29	27	27	28
Foot4	27	27	26	31	29	25	27
Foot5	28	26	26	26	30	25	28
Average	25.9	27.4	27.8	28.7	29.2	27.3	29.1

4.3 Numerical filter and its parameters

As it was concluded in the previous section, the results of the mechanical damper were highly unsatisfying. Therefore, data smoothing based on the Savitzky-Golay filter was considered. However, as was already mentioned earlier in 3.3, this filter requires two parameters (window length and polynomial order) as an input. It was stated earlier that third-order polynomial was found to be the optimal choice. This section contains results, and their discussion, of the studies that were executed in order to find the optimal window length parameter for a numerical filter. On top of that, results and a discussion of numerical filter performance evaluation are presented.

Window length for Savitzky-Golay filter

One of the required S-G parameters is window length which is identified as a portion of data points that the filter will attempt to fit in one polynomial. During the experiments, it was noted that the larger the window length the bigger is the peak shift to either positive or negative direction from the original (unfiltered) timestamp. Thus, this study was executed to evaluate the influence of the window length on the peak occurrence shift, furtherly referred to as a *phase shift*, in order to define the optimal window length for the final gait analysis algorithm. The graphical representation of two particular cases of the phase shift, which are drawn as an example, is shown in Figure 4.4 for the ankle and Figure 4.5 for the shank. On the graphs, the original data is shown in blue and all the data filtered with different window lengths are consequently marked. The peaks are marked with the dots, and as was mentioned earlier, at the given gait event range (grey box on the plots), the highest acceleration is considered to be the gait event occurrence. As can be seen from the graphs, the larger gets the window length the larger gets the phase shift. The same study was applied to every dataset gathered during the Experimental study, therefore 300 datasets were analyzed. The summary of this study is represented by box plots in Figure 4.6, Figure 4.7, and Figure 4.8.

Each entity on the box plots summarizes the phase shift with increasing window length across all the studied locations. Thus, it includes 100 datasets per location. The smallest variability occurred in the ankle data, whereas foot and shank showed similar results. Across all the locations, the phase shift distribution was rather consistent and was increasing gradually with the growth of the window length. From the graphs, it is seen that once window length approaches the value of 59 cs (0.59 s) the phase shift gets up to the region of 0.1 s. Based on the results of the TinyTag validation for gait analysis (Table 4.2), it can be seen that on average stride duration is 1.13 seconds. Having that,

0.1 s is roughly 10% of the stride duration. Therefore, it was decided that window length above 49 cs would introduce too dramatic changes into raw data, to keep the gait analysis results reliable.

It can also be noted from Figure 4.4 that for window length values below 29, smoothed data has more than one peak for gait events. However, for other sensor locations, window length 29 also was producing more than a single peak for HS events, as can be seen from Figure 4.5. Since the target of the data filtering was to achieve one peak per gait event, window lengths below 39 were concluded to be insufficient, despite producing desirable results for ankle location. Based on the results of this study window length of 39 cs was chosen (see Figure 4.5, purple) and afterward adopted for the final gait analysis algorithm. This window length, while smoothing data to the point of desired single peak per gait event, kept phase shift in the range of roughly 5 cs or 0.05 s, as can be seen from the box plots.

Phase shift evaluation, Ankle dataset 1, Subject 1

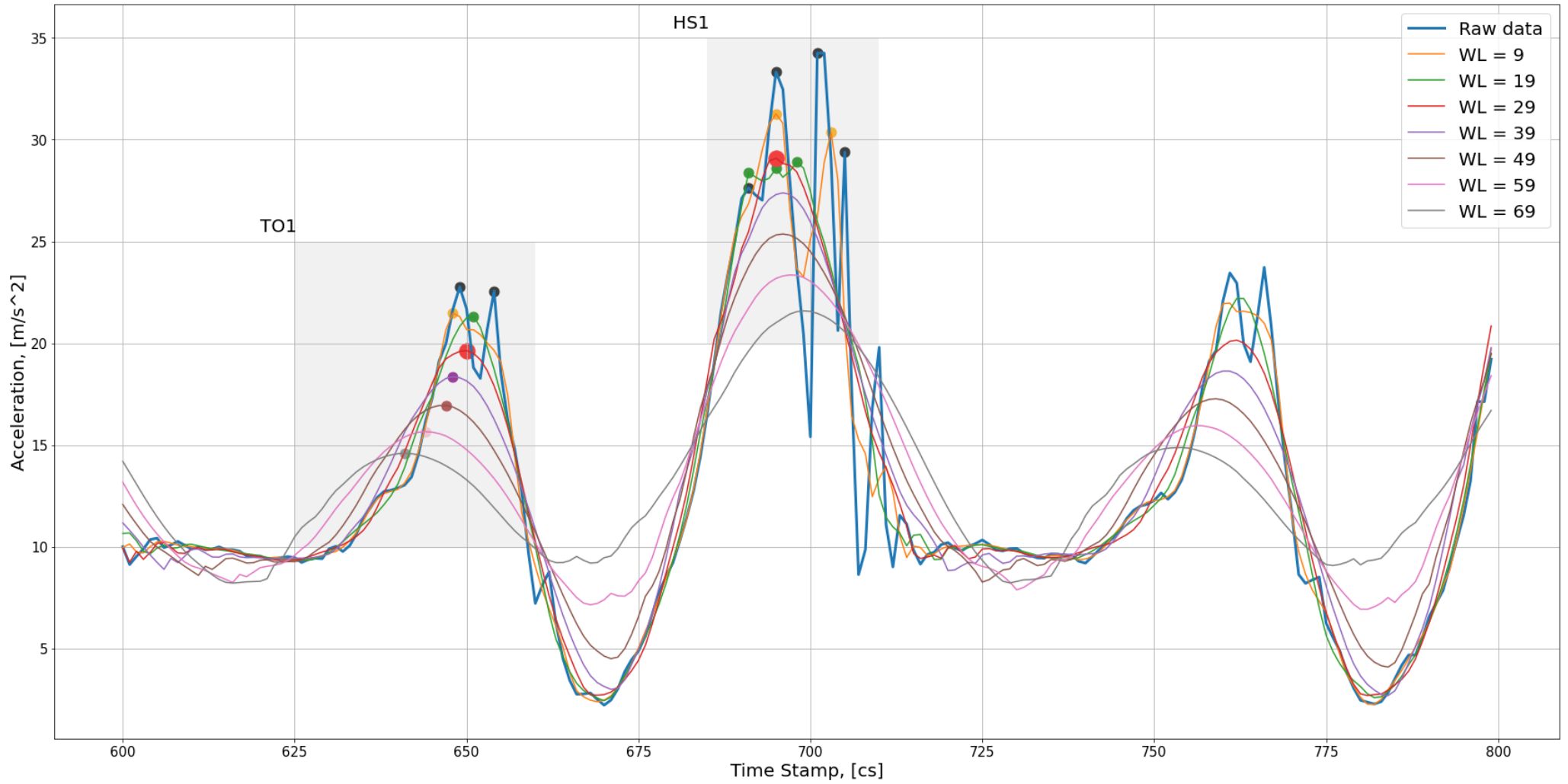


Figure 4.4 Example of the phase shift caused by *window length* parameter of the *S-G filter* in ankle dataset 1 of Subject 1

Phase shift evaluation, Shank dataset 4, Subject 1

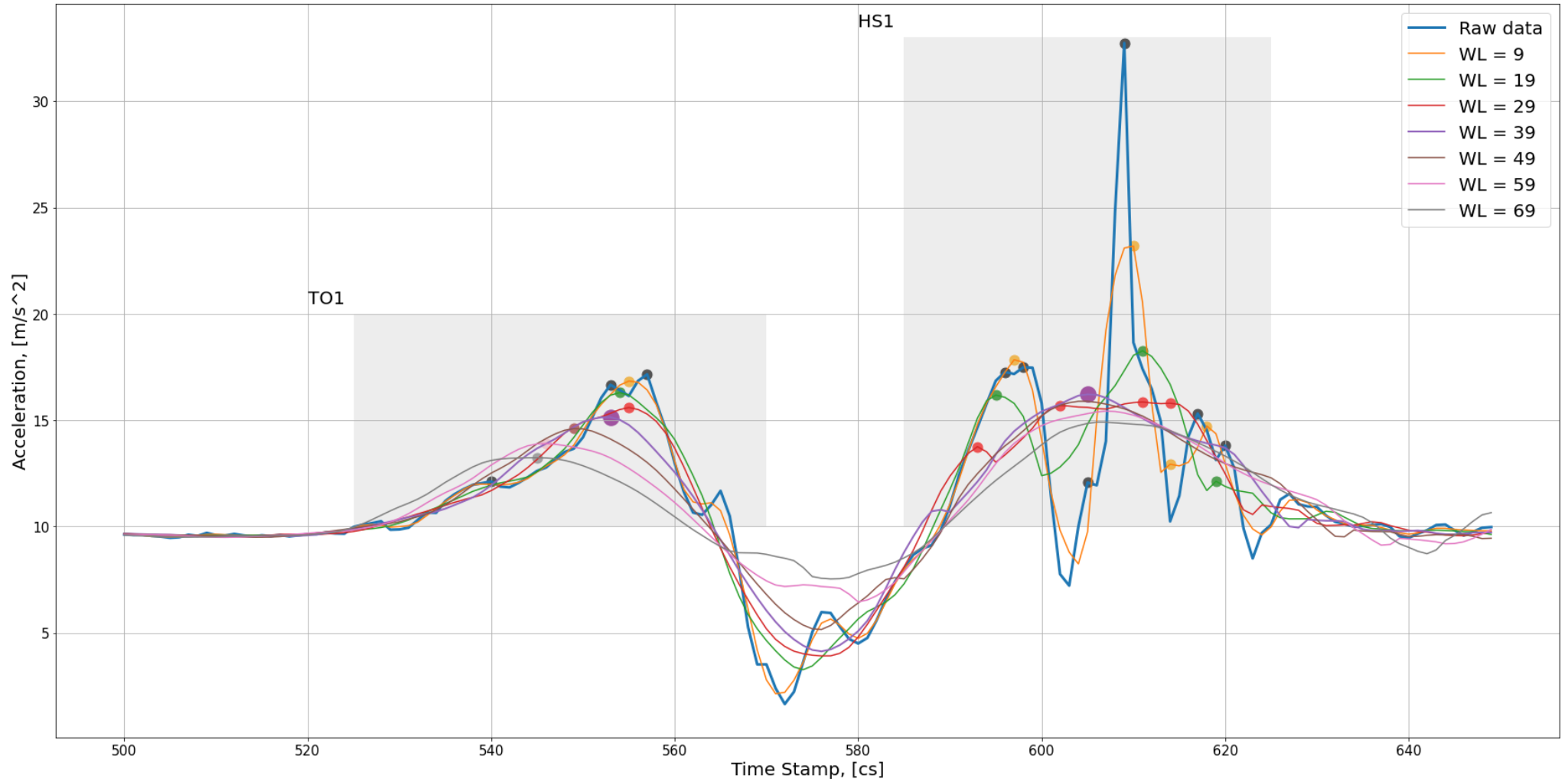


Figure 4.5 Example of the phase shift caused by *window length* parameter of the *S-G filter* in shank dataset 1 of Subject 4

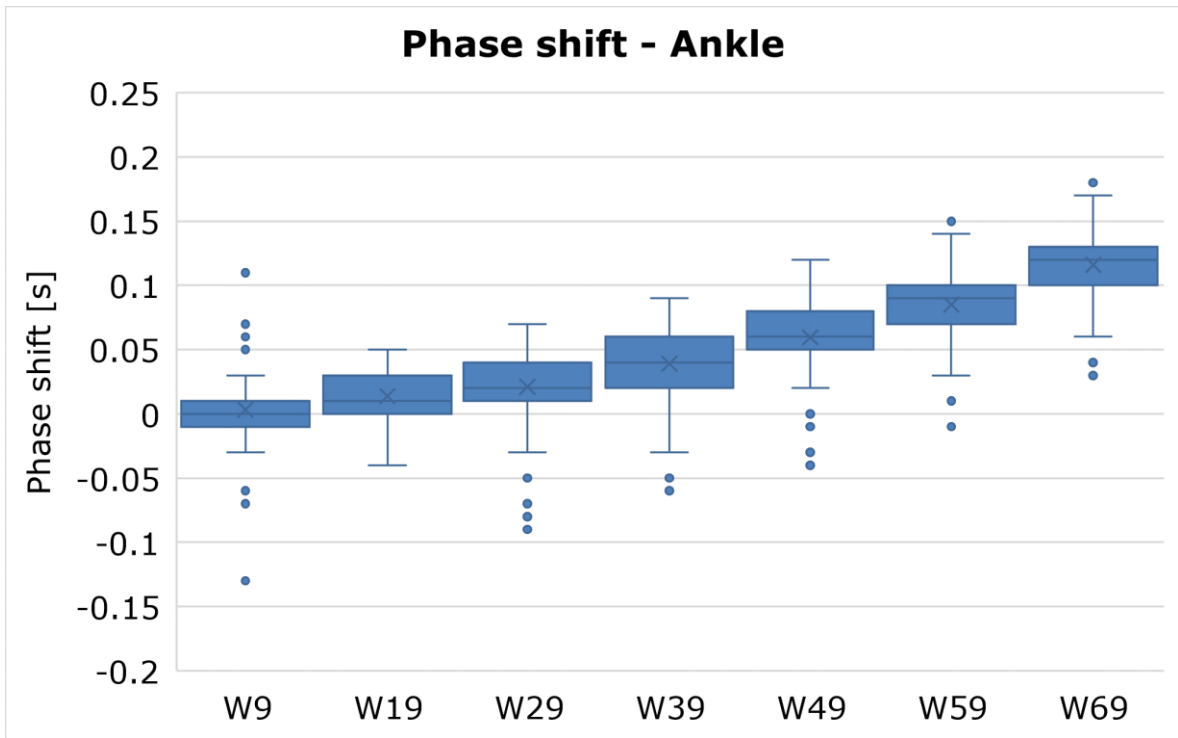


Figure 4.6 Window Length Phase shift for sensor positioned on the ankle, where W9 – W69 correspond to the studied window length for the S-G filter

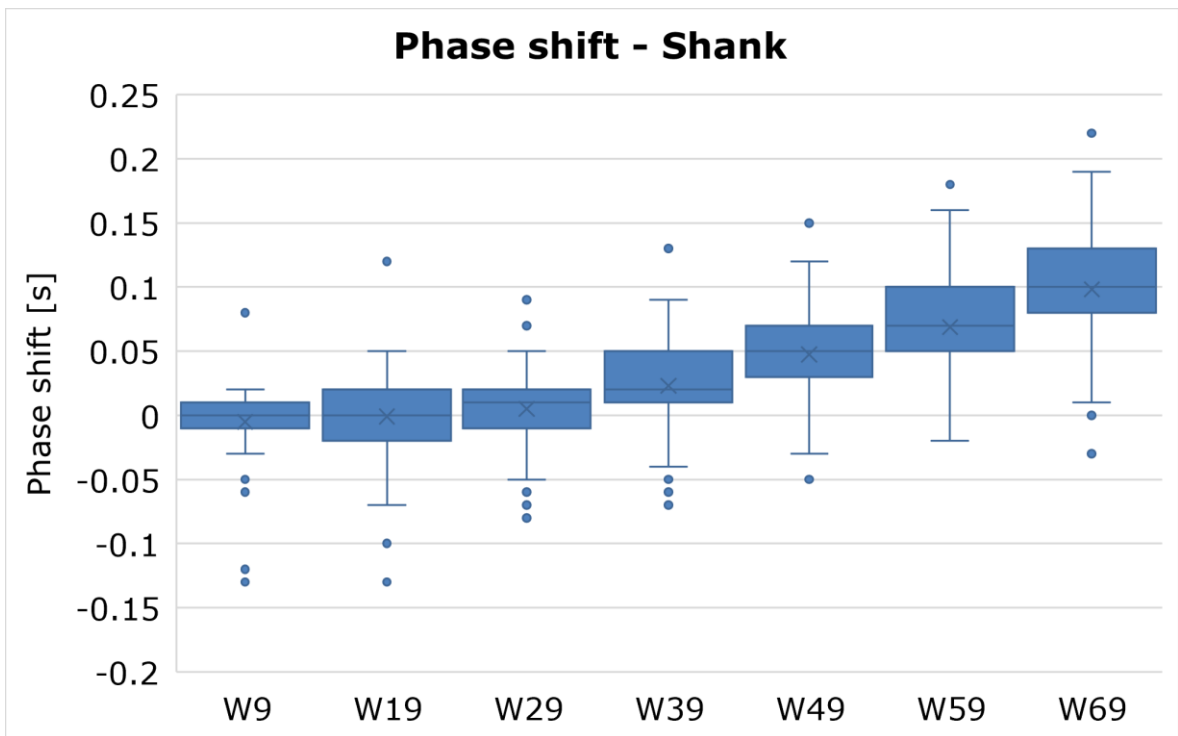


Figure 4.7 Window Length Phase shift for sensor positioned on the shank, where W9 – W69 correspond to the studied window length for the S-G filter

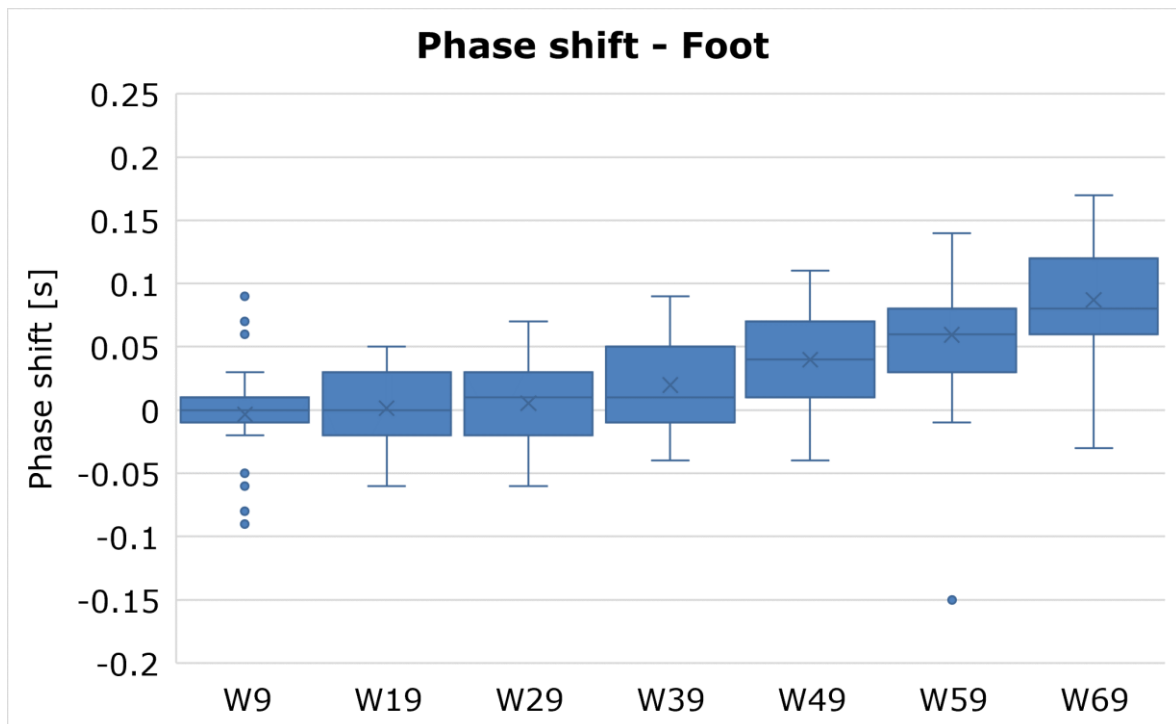


Figure 4.8 Window Length Phase shift for sensor positioned on the foot, where W9 – W69 correspond to the studied window length for the S-G filter

Numerical filter performance evaluation

To evaluate numerical filter *numbers of peaks per dataset* were monitored. For that, datasets acquired during Pilot Study 2 were used. The summary of this study is presented in Table 4.5, and for better visualization, the results of the mechanical damper smoothing were kept.

For this study S-G filter with a window length of 39 cs with a third-order polynomial was applied. As can be seen from the table, the numerical filter introduces compelling changes in the outcome. In all cases, the number of identified peaks is close to the desired seven peaks, with exceptions in some datasets with the sensor located on the shank and one dataset from the foot location, whereas for mechanical damper number of peaks varies in a range from 24 to 36 peaks.

Based on the results of this particular study it was also highlighted that, with the given set-up, the ankle could be the most suitable location for gait analysis with wearable IMU, since in all conditions the number of peaks identified with the numerical filter was equal to seven. On top of that, since the number of peaks in the control study doesn't differ significantly from studies with buffer materials, for further data acquisition they were not used.

Table 4.5 Summary of the numerical filter development study. Where: Control Study – data that was gathered without buffer material, other columns of first row – studied buffer material with its thickness given in the brackets, Mech. D – sum of gait events identified with mechanical damper, Num. F. – sum of gait events identified with numerical filter

Material	Control Study		Zhermack ED 16 (3 mm)		Zhermack ED 16 (5 mm)		Zhermack ED 22 (3 mm)		Zhermack ED 22 (5 mm)		Sorbothane (2.4 mm)		Sorbothane (6.4 mm)	
	Mech. D	Num. F	Mech. D	Num. F	Mech. D	Num. F	Mech. D	Num. F	Mech. D	Num. F	Mech. D	Num. F	Mech. D	Num. F
Ankle1	23	7	25	7	29	7	26	7	28	7	25	7	29	7
Ankle2	24	7	29	7	26	7	28	7	28	7	25	7	30	7
Ankle3	23	7	26	7	32	7	29	7	28	7	24	7	29	7
Ankle4	24	7	27	7	27	7	25	7	29	7	25	7	26	7
Ankle5	24	7	26	7	30	7	29	7	26	7	24	7	28	7
Shank1	21	7	24	7	27	7	32	7	28	7	28	7	27	7
Shank2	26	7	29	7	26	8	31	7	32	8	33	7	36	7
Shank3	28	7	28	8	29	8	28	7	32	7	32	7	29	7
Shank4	25	7	31	8	29	7	32	7	32	7	32	7	30	7
Shank5	26	7	29	8	28	9	31	7	32	7	31	7	32	7
Foot1	29	7	26	7	26	7	27	7	28	7	26	7	28	7
Foot2	30	7	30	7	27	7	27	7	29	7	28	7	29	7
Foot3	30	8	28	7	29	7	29	7	27	7	27	7	28	7
Foot4	27	7	27	7	26	7	31	7	29	7	25	7	27	7
Foot5	28	7	26	7	26	7	26	7	30	7	25	7	28	7
Average	26.7	7.1	29.5	7.2	27.8	7.3	28.7	7.0	29.2	7.1	27.3	7.0	29.1	7.0

4.4 Efficiency of the developed algorithm

The efficiency of the automated gait events identification algorithm was evaluated based on the number of gait events the algorithm was able to identify correctly with reference value being predefined by the data acquisition set-up. As shown in Figure 3.18, for matrix profiling, two approaches were tested, based on *general* and *subject-based motif*. To further estimate the accuracy of the developed algorithm and evaluate the influence of the sensor's location on the algorithm's accuracy, results are presented individually for each location. Two approaches with general and subject-based motif were evaluated separately; however to better visualize the results they are presented within the same table.

The results of this study are captured in Table 4.6 for the ankle, Table 4.7 for the shank, and Table 4.8 for the foot. In the tables, every row represents the summary of 10 repetitions per location. Columns, on the other hand, report in how many repetitions, out of 10 performed, all seven gait events were identified correctly. For example, in Table 4.6, Subject 7 has 100% in "General Motif" and 100% in "Subject-Based Motif". This means, that for every repetition this particular study subject performed, all gait events were identified correctly. On the contemporary, for Subject 3 in Table 4.8, the general motif approach managed to identify all seven gait events correctly in 7 out of 10 repetitions performed, whereas for the subject-based motif approach, within the same 10 datasets of this subject all seven gait events were identified correctly in all 10 repetitions. Therefore, if "Correct detection" is 100% then all seven gait events were identified correctly in 10 all repetitions. If however, it is 10%, then only in 1 dataset all gait events were found, and so on.

Table 4.6 Matrix profiling accuracy evaluation for a sensor located on the ankle

ANKLE	Correct detection	
	General Motif	Subject-Based Motif
Subject 1	50%	60%
Subject 2	100%	100%
Subject 3	50%	100%
Subject 4	20%	80%
Subject 5	70%	90%
Subject 6	20%	60%
Subject 7	100%	100%

ANKLE	Correct detection	
	General Motif	Subject-Based Motif
Subject 8	50%	100%
Subject 9	70%	90%
Subject 10	100%	100%
Mean	63%	88%

Table 4.7 Matrix profiling accuracy evaluation for a sensor located on the shank

SHANK	Correct detection	
	General Motif	Subject-Based Motif
Subject 1	30%	60%
Subject 2	60%	60%
Subject 3	20%	90%
Subject 4	100%	100%
Subject 5	100%	100%
Subject 6	100%	100%
Subject 7	100%	100%
Subject 8	40%	70%
Subject 9	80%	60%
Subject 10	100%	100%
Mean	73%	84%

Table 4.8 Matrix profiling accuracy evaluation for a sensor located on the foot

FOOT	Correct detection	
	General Motif	Subject-Based Motif
Subject 1	20%	80%
Subject 2	80%	90%
Subject 3	70%	100%
Subject 4	80%	100%
Subject 5	70%	100%
Subject 6	80%	90%
Subject 7	100%	100%

Table 4.8 continued

FOOT	Correct detection	
	General Motif	Subject-Based Motif
Subject 8	20%	90%
Subject 9	70%	80%
Subject 10	100%	80%
Mean	69%	91%

From the results of this study, it can be noted that overall, the subject-based motif approach shows better efficiency than the general motif. For the general motif case, the best accuracy was found to be when the sensor is located on the shank, with the foot being only 4% lower. Among subject-based motif results, foot resulted being the most accurate, showing on average 91% of gait events being correctly detected, with all seven gait events being identified correctly in every repetition in four datasets. Additionally, for every repetition of each subject, a *number of peaks per dataset* was calculated (similarly to *Mechanical Damper* and *Numerical filters* comparison in Table 4.5). Results of the calculations show that for ankle, on average 7.2 peaks per dataset were identified, whereas for shank and foot 8.53 and 7.7 peaks respectively were found.

Outcomes of this study point out that with a given set-up, subject-based motif results in the best gait events identification accuracy. With regards to the location, the ankle and foot are showing similarly high accuracy, 88 and 91% respectively, whereas the ankle shows a lower score in the *number of peaks per dataset*.

To conclude, matrix profiling is a very efficient approach for data mining; however, its simplicity also introduces some flaws for such specific tasks as gait events identification. This data processing approach simply finds the lowest Euclidian distance, without taking into consideration the specific goals of the task. An example of this is highlighted in Figure 4.9. As can be seen from the figure, indeed, the algorithm has found the motif that is matching general gait cycle pattern (orange). However, since the area of interest of this study is seven consequent gait events starting with the first TO of the right leg (green rectangle), it can be noted that the algorithm missed the first stride (TO1 and HS1). Having that, further improvements to the algorithm are required.

Unsuccessfully identified Motif

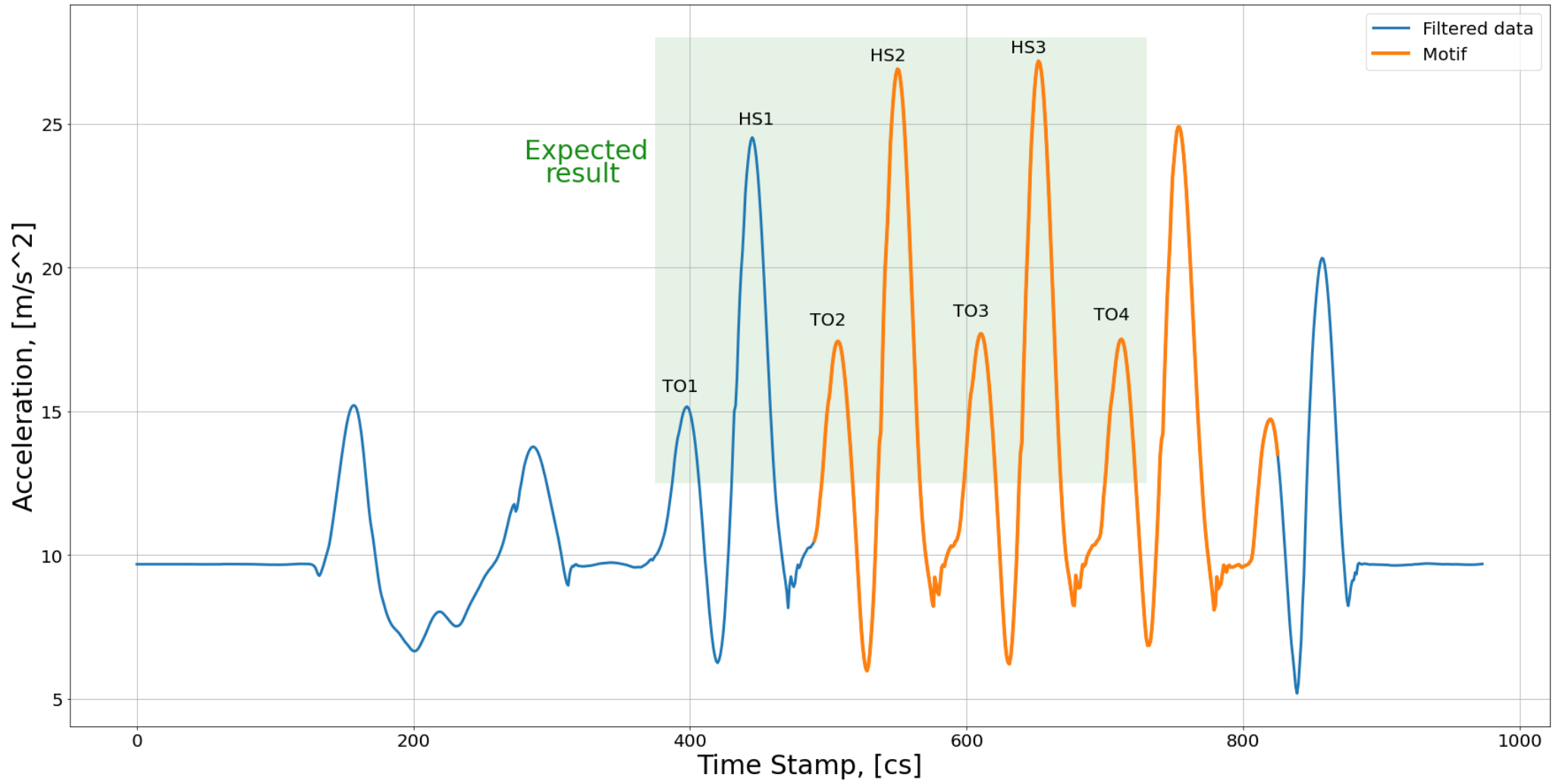


Figure 4.9 An example of matrix profiling finding a motif that does not match the purpose of the study

5. CONCLUSIONS

The main objective of this work was to develop an automated gait event detection system implementing a new miniature IMU sensor developed at TalTech. First, a custom mounting system, consisting of 3D printed durable resin casing and straps was designed in order to assure firm fixation of the sensor and its flexible application in different locations on the lower limb of the human body. Next, the sensor was validated for human gait analysis, utilizing a camera-based method as a ground truth reference. The results showed that stride duration calculated based on these two methods had a mean absolute difference of only 0.03 seconds, which with a given average stride duration of 1.13 seconds represents an insignificant deviation.

Once validated, a data filtering system was developed in order to smooth the raw IMU data, which was too noisy for an efficient automated gait event identification system development. For this, two approaches were attempted, namely a mechanical damper and a numerical filter. Experiments were made with a sensor located on three different positions: shank, ankle, and foot. With a mechanical damper, it was attempted to smooth the data by implementing buffer materials between the sensor casing and the human body that aimed to limit the noise and decrease the influence of the soft-tissue (skin movement) artifacts. Thus, as the result, it was expected to achieve seven peaks per dataset that describe three consequent strides. The results of this attempt, however, proved buffer materials being unable to achieve this task. Whereas in the control study, which did not utilize buffer materials, the average number of peaks was 25.9, datasets with buffer materials resulted in 27.3 – 29.2 peaks per dataset.

The mechanical damping approach was found to have no positive impact on the recorded data. Instead of mechanical damping, a Savitzky-Golay numerical filter was implemented to smooth the data. Identically, a cross-comparison of the number of peaks per dataset in filtered and unfiltered data was made. The numerical filter method was applied to the same datasets as for the mechanical damper and the results highlighted significant improvements. The number of peaks decreased from 27.3 – 29.2 to 7.0 -7.3 peaks per dataset. Additional analysis was made in order to identify optimal parameters for the S-G filter, and it was found that for the developed algorithm a third-order polynomial with a window length of 39 centiseconds (0.39 seconds) is the preferable choice.

Lastly, an automated gait events identification signal processing method was developed combining numerical filtering of the data, a novel data mining algorithm called matrix profiling, and a peak identification function. The newly developed

algorithm was implemented on the datasets that were obtained from the study group consisting of 10 healthy subjects. Three different sensor locations were studied, namely shank, ankle, and foot. Thus, in total 300 datasets (10 repetitions for each location per every study subject) were analyzed. Two different approaches of the data processing were considered: a general motif, where the same motif was applied to all datasets, and a subject-based motif, where a motif was specified for each subject and study location separately. Results of this study highlighted that overall, the subject-based motif method is significantly more efficient, showing better accuracy for every location. It was also noted that ankle and foot locations provide the highest efficiency for the subject-based motif method, resulting in 88 and 91% of correct detection respectively. Another important finding was the number of peaks per dataset, whereas, with desired seven peaks per dataset, the lowest score was highlighted in the ankle, with an average of 7.2 peaks per dataset across the analyzed batch of data. For foot and shank, this number was 7.7 and 8.53 respectively. Having that, it was concluded that overall, with a given experiment set-up a subject-based motif approach with the sensor located on the ankle or foot is a preferable method for automated gait event identification gait analysis due to the lowest number of peaks per dataset and highest correct detection.

Whereas gait analysis with wearable sensors has been known for over two decades, this work proposes a gait analysis approach that significantly decreases time-consuming data processing. It also provides data-supported suggestions concerning the sensor attachment system and its location on the lower limb that results in about 90% accuracy of gait events detection utilizing the proposed automated data processing approach.

Despite the proposed data processing method showing high efficiency for gait events identification in a differentiated group of healthy subjects' data, they rarely are a target of clinical analysis. Therefore, in future works to further validate the developed data processing system a study with the group of subjects with abnormal gates is required. On top of that, while the subject-based motif method is showing great efficiency, it still requires, though limited, input from the user. Thus, improvements to achieve further automatization are to be considered.

6. KOKKUVÕTE

Käesoleva töö peamine eesmärk oli arendada automaatne kõnnisündmuste (*gait event*, ingl. k.) tuvastamise süsteem, mis rakendab TalTechis välja töötatud uut miniatuurset IMUga andurit (inertsiaalandurit). Kõigepealt kavandati ja 3D-prinditi - korpusest ja rihmadest koosnev spetsiaalne kinnitussüsteem, et tagada anduri stabiilne kinnitamine ja selle pindlik kasutamine erinevates kohtades inimkeha alajäsematel. Seejärel valideeriti andur inimese kõnnialalüüsiks, kasutades kaamerapõhist meetodit, mis on alustõeks. Tulemused näitasid, et nende kahe meetodi alusel arvutatud sammu kestuse keskmine absoluutne erinevus oli ainult 0,03 sekundit, mis keskmise sammu kestust 1,13 sekundit arvestades on ebaoluline kõrvalekalle.

Pärast valideerimist töötati välja andmete filtreerimissüsteem, et siluda IMU toorandmeid, mis olid liiga mürarikkad, et arendada tõhusat automaatset kõnnisündmuste tuvastamise süsteemi. Selleks prooviti kasutada kahte lähenemisviisi, nimelt mehaanilist summutust ja numbrilist filtrit. Katsed tehti kolmel erineval positsioonil asuva anduriga: säärel, pahklul ja jalalabal. Mehaanilise filtri abil püüti andmeid siluda, rakendades anduri korpuse ja inimkeha vahel puhvermaterjali, mille eesmärk oli piirata müra ja vähendada naha liikumise artefaktide mõju. Sellega loodeti saavutada seitse piiki andmekogumi kohta, mis kirjeldavad kolme järjestikust sammu. Katse tulemused näitasid siiski, et puhvermaterjalid ei suuda antud ülesannet täita. Kui võrdlusuuringus, kus puhvermaterjale ei kasutatud, oli keskmine piikide arv 25,9, siis puhvermaterjalidega andmekogumite puhul saadi 27,3 – 29,2 piiki andmekogumi kohta.

Leiti, et mehaaniline summutamine ei mõjuta salvestatud andmeid positiivselt ja selle asemel kasutati andmete silumiseks Savitzky-Golay numbrilist filtrit. Samamoodi võrreldi piikide arvu andmekogumi kohta filtreeritud ja filtreerimata andmetes. Numbrilise filtri meetodit rakendati samade andmekogumite suhtes, kui mehaanilise summutuse puhul ja tulemused näitasid märkimisväärset paranemist. Piikide arv vähenes 27,3-29,2 piigilt 7,0-7,3 piigini andmekogumi kohta. S-G-filtri optimaalsete parameetrite kindlaksmääramiseks tehti täiendav analüüs ja leiti, et väljatöötatud algoritmi jaoks on eelistatavad valikud 3. järgu polünoom ja akna pikkus 0,39 sekundit.

Lõpuks töötati välja automaatne kõnnisündmuste tuvastamise signaalitöötlusmeetod, mis ühendab andmete numbrilise filtreerimise, uudse andmekaevealgoritmi, mida nimetatakse maatriksprofiiliks, ja piikide tuvastamise. Välja töötatud uut algoritmi

rakendati andmekogumite suhtes, mis saadi 10 tervest isikust koosnevast uurimisrühmast. Uuriti kolme erinevat sensori asukohta, nimelt säärt, pahklud ja jalalaba. Seega analüüsiti kokku 300 andmekogumit (10 kordust iga asukoha kohta ja iga uuritava kohta). Andmete töötlemisel kaaluti kahte erinevat lähenemisviisi: üldine lähenemine, mille puhul rakendati sama meetodit kõigi andmekogumite suhtes, ja subjektipõhine lähenemine, mille puhul määrati lahendus iga subjekti ja uuringukoha jaoks eraldi. Selle uuringu tulemused näitasid, et üldiselt on subjekti- ja asukohapõhine meetod oluliselt tõhusam, näidates iga asukoha puhul paremat täpsust. Samuti täheldati, et anduri asukohad hüppeliigesel ja jalalabal on kõige tõhusamad subjektipõhise lähenemise puhul, tulemuseks on vastavalt 88% ja 91% korrektset tuvastamist. Teine oluline leid oli piikide arv andmekogumi kohta. Madalaim tulemus oli pahkluu puhul, kus analüüsitud andmepartiis oli keskmiselt 7,2 piiki andmekogumi kohta (soovitud tuvastatavate piikide arv andmekogumi kohta oli 7). Jalalaba ja sääre puhul oli see arv vastavalt 7,7 ja 8.53, Selle põhjal järeldati, et antud katse ülesehituse puhul on üldjuhul eelistatavam meetod kõnnisündmuste automaatseks tuvastamiseks kõnnianalüüsis subjektipõhine lähenemisviis, mille korral sensor asub pahkluu või jalalaba peal. See lähenemine annab kõige vähem piike andmekogumi kohta ja kõige rohkem õigeid tulemusi.

Kuigi kõnnianalüüs kantavate anduritega on tuntud juba üle kahe aastakümne, pakutakse käesolevas töös välja kõnnianalüüsi lähenemisviisi, mis vähendab märkimisväärselt aeganõudvat andmetöötlust. Samuti antakse andmetel põhinevaid soovitusi anduri kinnitussüsteemi ja selle asukoha kohta alajäsemetel, mille tulemuseks on umbes 90% täpsus kõnnisündmuste tuvastamisel, kasutades pakutud automaatset andmetöötlusmeetodit.

Hoolimata sellest, et väljapakutud andmetöötlusmeetod näitab suurt tõhusust kõnnisündmuste tuvastamisel tervete isikute andmetes, on nad harva kliinilise analüüsi sihtmärgiks. Seetõttu on edaspidi vaja väljatöötatud andmetöötlussüsteem täiendavaks valideerimiseks uuringut ebanormaalse kõnnakuga subjektide rühmaga. Lisaks sellele, kuigi subjektipõhine meetod näitab suurt tõhusust, nõuab see siiski kasutaja sisendit, kuigi piiratud määral. Seega tuleb kaaluda täiustusi, et saavutada edasine automatiseerimine.

7. LIST OF REFERENCES

- [1] M. W. Whittle, *Gait Analysis: An Introduction*. Butterworth-Heinemann, 2014.
- [2] L. S. Vargas-Valencia, A. Elias, E. Rocon, T. Bastos-Filho, and A. Frizera, "An IMU-to-Body Alignment Method Applied to Human Gait Analysis," *Sensors*, vol. 16, no. 12, Art. no. 12, Dec. 2016, doi: 10.3390/s16122090.
- [3] Y. Celik, S. Stuart, W. L. Woo, and A. Godfrey, "Gait analysis in neurological populations: Progression in the use of wearables," *Med. Eng. Phys.*, vol. 87, pp. 9–29, Jan. 2021, doi: 10.1016/j.medengphy.2020.11.005.
- [4] J. Taborri, E. Palermo, S. Rossi, and P. Cappa, "Gait Partitioning Methods: A Systematic Review," *Sensors*, vol. 16, no. 1, Art. no. 1, Jan. 2016, doi: 10.3390/s16010066.
- [5] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward Pervasive Gait Analysis With Wearable Sensors: A Systematic Review," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1521–1537, Nov. 2016, doi: 10.1109/JBHI.2016.2608720.
- [6] G. Yang, W. Tan, H. Jin, T. Zhao, and L. Tu, "Review wearable sensing system for gait recognition," *Clust. Comput.*, vol. 22, no. 2, pp. 3021–3029, Mar. 2019, doi: 10.1007/s10586-018-1830-y.
- [7] T. Stöckel, R. Jacksteit, M. Behrens, R. Skripitz, R. Bader, and A. Mau-Moeller, "The mental representation of the human gait in young and older adults," *Front. Psychol.*, vol. 6, p. 943, 2015, doi: 10.3389/fpsyg.2015.00943.
- [8] D. Jarchi, J. Pope, T. K. M. Lee, L. Tamjidi, A. Mirzaei, and S. Sanei, "A Review on Accelerometry-Based Gait Analysis and Emerging Clinical Applications," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 177–194, 2018, doi: 10.1109/RBME.2018.2807182.
- [9] M. Pistacchi *et al.*, "Gait analysis and clinical correlations in early Parkinson's disease," *Funct. Neurol.*, vol. 32, no. 1, pp. 28–34, Apr. 2017, doi: 10.11138/FNeur/2017.32.1.028.
- [10] A. Muro-de-la-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications," *Sensors*, vol. 14, no. 2, Art. no. 2, Feb. 2014, doi: 10.3390/s140203362.
- [11] F. Petraglia, L. Scarcella, G. Pedrazzi, L. Brancato, R. Puers, and C. Costantino, "Inertial sensors versus standard systems in gait analysis: a systematic review and meta-analysis," *Eur. J. Phys. Rehabil. Med.*, vol. 55, no. 2, pp. 265–280, Apr. 2019, doi: 10.23736/S1973-9087.18.05306-6.
- [12] M. Fusca, F. Negrini, P. Perego, L. Magoni, F. Molteni, and G. Andreoni, "Validation of a Wearable IMU System for Gait Analysis: Protocol and Application to a

New System," *Appl. Sci.*, vol. 8, no. 7, Art. no. 7, Jul. 2018, doi: 10.3390/app8071167.

[13] R. Caldas, M. Mundt, W. Potthast, F. Buarque de Lima Neto, and B. Markert, "A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms," *Gait Posture*, vol. 57, pp. 204–210, Sep. 2017, doi: 10.1016/j.gaitpost.2017.06.019.

[14] L. C. Benson, C. A. Clermont, E. Bošnjak, and R. Ferber, "The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review," *Gait Posture*, vol. 63, pp. 124–138, Jun. 2018, doi: 10.1016/j.gaitpost.2018.04.047.

[15] B. Zhang, S. Jiang, D. Wei, M. Marschollek, and W. Zhang, "State of the Art in Gait Analysis Using Wearable Sensors for Healthcare Applications," in *2012 IEEE/ACIS 11th International Conference on Computer and Information Science*, May 2012, pp. 213–218. doi: 10.1109/ICIS.2012.100.

[16] S. Sprager and M. B. Juric, "Inertial Sensor-Based Gait Recognition: A Review," *Sensors*, vol. 15, no. 9, Art. no. 9, Sep. 2015, doi: 10.3390/s150922089.

[17] "G-WALK | Wearable inertial system," *BTS Bioengineering*.
<https://www.btsbioengineering.com/products/g-walk-inertial-motion-system/>
(accessed Nov. 29, 2021).

[18] "Gait Analysis." <https://www.xsens.com/gait-analysis> (accessed Nov. 29, 2021).

[19] K. Ben Mansour, N. Rezzoug, J. Jacquier-Bret, and P. Gorce, "Validation of a low-cost wearable accelerometer for temporal gait parameter quantification," *Comput. Methods Biomech. Biomed. Engin.*, vol. 17 Suppl 1, pp. 160–161, Jan. 2014, doi: 10.1080/10255842.2014.931623.

[20] D. Trojaniello, A. Cereatti, and U. Della Croce, "Accuracy, sensitivity and robustness of five different methods for the estimation of gait temporal parameters using a single inertial sensor mounted on the lower trunk," *Gait Posture*, vol. 40, no. 4, pp. 487–492, Sep. 2014, doi: 10.1016/j.gaitpost.2014.07.007.

[21] D. Trojaniello, A. Ravaschio, J. M. Hausdorff, and A. Cereatti, "Comparative assessment of different methods for the estimation of gait temporal parameters using a single inertial sensor: application to elderly, post-stroke, Parkinson's disease and Huntington's disease subjects," *Gait Posture*, vol. 42, no. 3, pp. 310–316, Sep. 2015, doi: 10.1016/j.gaitpost.2015.06.008.

[22] S. D. Din, A. Hickey, N. Hurwitz, J. C. Mathers, L. Rochester, and A. Godfrey, "Measuring gait with an accelerometer-based wearable: influence of device location, testing protocol and age," *Physiol. Meas.*, vol. 37, no. 10, pp. 1785–1797, Sep. 2016, doi: 10.1088/0967-3334/37/10/1785.

- [23] A. Hickey, S. D. Din, L. Rochester, and A. Godfrey, "Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis," *Physiol. Meas.*, vol. 38, no. 1, pp. N1–N15, Dec. 2016, doi: 10.1088/1361-6579/38/1/N1.
- [24] M. Zago *et al.*, "Gait evaluation using inertial measurement units in subjects with Parkinson's disease," *J. Electromyogr. Kinesiol.*, vol. 42, pp. 44–48, Oct. 2018, doi: 10.1016/j.jelekin.2018.06.009.
- [25] H. Zhao, Z. Wang, S. Qiu, J. Li, F. Gao, and J. Wang, "Evaluation of Inertial Sensor Configurations for Wearable Gait Analysis," in *Big Data, Cloud Computing, and Data Science Engineering*, R. Lee, Ed. Cham: Springer International Publishing, 2020, pp. 197–212. doi: 10.1007/978-3-030-24405-7_13.
- [26] W.-C. Hsu *et al.*, "Multiple-Wearable-Sensor-Based Gait Classification and Analysis in Patients with Neurological Disorders," *Sensors*, vol. 18, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/s18103397.
- [27] L. Carcreff *et al.*, "What is the Best Configuration of Wearable Sensors to Measure Spatiotemporal Gait Parameters in Children with Cerebral Palsy?," *Sensors*, vol. 18, no. 2, Art. no. 2, Feb. 2018, doi: 10.3390/s18020394.
- [28] A. R. Anwary, H. Yu, and M. Vassallo, "Optimal Foot Location for Placing Wearable IMU Sensors and Automatic Feature Extraction for Gait Analysis," *IEEE Sens. J.*, vol. 18, no. 6, pp. 2555–2567, Mar. 2018, doi: 10.1109/JSEN.2017.2786587.
- [29] A. Peters, B. Galna, M. Sangeux, M. Morris, and R. Baker, "Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review," *Gait Posture*, vol. 31, no. 1, pp. 1–8, Jan. 2010, doi: 10.1016/j.gaitpost.2009.09.004.
- [30] M. Suzuki, H. Mitoma, and M. Yoneyama, "Quantitative Analysis of Motor Status in Parkinson's Disease Using Wearable Devices: From Methodological Considerations to Problems in Clinical Applications," *Park. Dis.*, vol. 2017, p. e6139716, May 2017, doi: 10.1155/2017/6139716.
- [31] G. Pacini Panebianco, M. C. Bisi, R. Stagni, and S. Fantozzi, "Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements," *Gait Posture*, vol. 66, pp. 76–82, Oct. 2018, doi: 10.1016/j.gaitpost.2018.08.025.
- [32] "Assessing the Accuracy of an Algorithm for the Estimation of Spatial Gait Parameters Using Inertial Measurement Units | Proceedings of the 4th International Conference on Movement Computing."
<https://dl.acm.org/doi/abs/10.1145/3077981.3078034> (accessed Nov. 29, 2021).
- [33] S. Khandelwal and N. Wickström, "Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios

- using the MAREA gait database," *Gait Posture*, vol. 51, pp. 84–90, Jan. 2017, doi: 10.1016/j.gaitpost.2016.09.023.
- [34] I. Mileti *et al.*, "Gait partitioning methods in Parkinson's disease patients with motor fluctuations: A comparative analysis," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, May 2017, pp. 402–407. doi: 10.1109/MeMeA.2017.7985910.
- [35] E. Allseits, J. Lučarević, R. Gailey, V. Agrawal, I. Gaunaurd, and C. Bennett, "The development and concurrent validity of a real-time algorithm for temporal gait analysis using inertial measurement units," *J. Biomech.*, vol. 55, pp. 27–33, Apr. 2017, doi: 10.1016/j.jbiomech.2017.02.016.
- [36] J. C. Pérez-Ibarra, H. Williams, A. A. G. Siqueira, and H. I. Krebs, "Real-Time Identification of Impaired Gait Phases Using a Single Foot-Mounted Inertial Sensor: Review and Feasibility Study," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Aug. 2018, pp. 1157–1162. doi: 10.1109/BIOROB.2018.8487694.
- [37] L. Zhou *et al.*, "Validation of an IMU Gait Analysis Algorithm for Gait Monitoring in Daily Life Situations," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Jul. 2020, pp. 4229–4232. doi: 10.1109/EMBC44109.2020.9176827.
- [38] "Buy Durable Resin," *Formlabs*. <https://formlabs.com/store/durable-resin/> (accessed Apr. 22, 2022).
- [39] A. Industries, "USB Micro-B Breakout Board." <https://www.adafruit.com/product/1833> (accessed Apr. 22, 2022).
- [40] "Elite Double 16 Fast," *Zhermack*. <https://www.zhermack.com/en/product/elite-double-16-fast/> (accessed Apr. 21, 2022).
- [41] "Elite Double 22," *Zhermack*. <https://www.zhermack.com/en/product/elite-double-22/> (accessed Apr. 21, 2022).
- [42] "Sorbothane Visco-Elastic Material Properties - FAQ." <https://www.sorbothane.com/material-properties.aspx> (accessed Apr. 21, 2022).
- [43] "Sorbothane," *Wikipedia*. Mar. 15, 2022. Accessed: Apr. 27, 2022. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Sorbothane&oldid=1077360813>
- [44] "Durometer Shore Hardness Scale Explained | AeroMarine," *Aeromarine Products Inc.*, Jul. 30, 2020. <https://www.aeromarineproducts.com/durometer-shore-hardness-scale/> (accessed May 02, 2022).
- [45] "Kinovea." <https://www.kinovea.org/> (accessed May 16, 2022).
- [46] "Home — Spyder IDE." <https://www.spyder-ide.org/> (accessed Apr. 22, 2022).

- [47] X. S. Papageorgiou *et al.*, "Experimental validation of human pathological gait analysis for an assisted living intelligent robotic walker," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Jun. 2016, pp. 1086–1091. doi: 10.1109/BIOROB.2016.7523776.
- [48] F. Crenna, G. B. Rossi, and M. Berardengo, "Filtering Biomechanical Signals in Movement Analysis," *Sensors*, vol. 21, no. 13, Art. no. 13, Jan. 2021, doi: 10.3390/s21134580.
- [49] Abraham. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures.," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/ac60214a047.
- [50] R. W. Schafer, "What Is a Savitzky-Golay Filter? [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011, doi: 10.1109/MSP.2011.941097.
- [51] "Savitzky–Golay filter," *Wikipedia*. Jan. 17, 2022. Accessed: Apr. 25, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Savitzky%E2%80%93Golay_filter&oldid=1066338331
- [52] C.-C. M. Yeh *et al.*, "Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, Spain, Dec. 2016, pp. 1317–1322. doi: 10.1109/ICDM.2016.0179.
- [53] S. Law, "Part 1: The Matrix Profile," *Medium*, Dec. 31, 2020. <https://towardsdatascience.com/the-matrix-profile-e4a679269692> (accessed Apr. 27, 2022).
- [54] S. Law, "Part 7: Fast Pattern Searching with STUMPY," *Medium*, Dec. 31, 2020. <https://towardsdatascience.com/part-7-fast-pattern-searching-with-stumpy-2baf610a8de1> (accessed Apr. 27, 2022).
- [55] T. Marrs, "Introduction to Matrix Profiles," *Medium*, Feb. 28, 2020. <https://towardsdatascience.com/introduction-to-matrix-profiles-5568f3375d90> (accessed May 09, 2022).

APPENDIX 1

This Appendix contains tables that were obtained during data labeling and served as a basis for Stride Duration calculations described in 4.1. Table 1 contains timestamps that were obtained from video processing with Kinovea. Based on these results stride durations in Table 4.1 were calculated. Table 2 in turn, contains timestamps of the gait events that were obtained from the IMU data. Based on these results stride durations in Table 4.2 were calculated.

Table 1 Timestamps of the Gait Events obtained from the video recordings

Sensor location and Nº of repetition	Occurrence of the gait event [s]						
	TO1	HS1	TO2	HS2	TO3	HS3	TO4
SHANK_1	10.98	11.48	12.21	12.68	13.35	13.85	14.55
SHANK_2	23.99	24.46	25.19	25.66	26.33	26.79	27.49
SHANK_3	12.61	13.08	13.78	14.25	14.95	15.42	16.15
SHANK_4	11.48	11.94	12.64	13.11	13.78	14.25	14.98
SHANK_5	12.08	12.54	13.25	13.68	14.35	14.81	15.48
ANKLE_1	14.68	15.18	15.88	16.32	16.98	17.45	18.12
ANKLE_2	9.54	9.98	10.68	11.11	11.78	12.24	12.91
ANKLE_3	11.98	12.44	13.11	13.58	14.25	14.68	15.38
ANKLE_4	12.14	12.58	13.25	13.71	14.35	14.81	15.52
ANKLE_5	7.21	7.67	8.34	8.78	9.44	9.91	10.61
FOOT_1	8.07	8.54	9.21	9.68	10.31	10.78	11.44
FOOT_2	8.04	8.51	9.14	9.61	10.28	10.71	11.41
FOOT_3	12.61	13.11	13.81	14.25	14.91	15.38	16.08
FOOT_4	11.84	12.34	12.98	13.45	14.08	14.55	15.22
FOOT_5	11.38	11.91	12.54	13.01	13.68	14.15	14.81

Table 2 Timestamps of the Gait Events obtained from the IMU data

Sensor location and Nº of repetition	Occurrence of the gait event [s]						
	TO1	HS1	TO2	HS2	TO3	HS3	TO4
SHANK_1	1849	1904	1971	2025	2086	2143	2204
SHANK_2	728	785	847	903	963	1017	1079

Table 2 continued

Sensor location and Nº of repetition	Occurrence of the gait event [s]						
	TO1	HS1	TO2	HS2	TO3	HS3	TO4
SHANK_3	601	655	717	772	834	888	953
SHANK_4	557	609	672	724	785	838	901
SHANK_5	595	651	714	766	825	879	937
ANKLE_1	536	587	649	701	766	814	873
ANKLE_2	678	728	789	842	901	954	1015
ANKLE_3	718	768	828	881	939	992	1051
ANKLE_4	608	658	718	764	830	881	942
ANKLE_5	589	640	698	750	809	861	923
FOOT_1	738	781	850	894	964	1003	1072
FOOT_2	665	705	783	825	886	929	999
FOOT_3	699	740	813	852	924	965	1044
FOOT_4	630	671	744	783	851	895	968
FOOT_5	579	627	693	733	806	846	915


```

#upload the data into Python environment
for Ref_file in Ref_files:
    if Ref_file.endswith(".xlsx") and Ref_file.startswith(Ref_Location):
        Ref_df = pd.read_excel(os.path.join(Ref_path, Ref_file))
        Ref_df.columns = ['time','stamp','battery',
'pressure','temperature','ax','ay','az','gx','gy','gz','mx','my','mz','averagea']
        Ref_acceleration[i] = Ref_df.averagea
        time[i] = Ref_df.time
        Ref_Acceleration_39[i] = Ref_SavGol_39(i)
        i += 1

#set the threshold based on the location of the sensor
if Ref_Location == 'Ankle':
    y = 13
elif Ref_Location == 'Shank':
    y = 12
elif Ref_Location == 'Foot':
    y = 13

#find peaks in filtered data
peaks,_ = find_peaks(Ref_Acceleration_39[0], height=y)
peaks_array_filtered[11]=peaks

#a function to find boundaries for the motif
def Ref_Peaks(i):
    print('Threshold value: ' + str(y))
    print ("Peaks time stamps in raw data: ", peaks_array_filtered[11])
    plt.plot(Ref_Acceleration_39[i], linewidth = '3', label = 'Filtered data')
    plt.axhline(y, linewidth = '3', color = 'r',label = 'Threshold')
    #plt.suptitle(name + ' ' + Ref_Location + ' ' + str(i), fontsize='30')
    plt.legend(fontsize = 30)
    plt.xlabel('Time Stamp, [cs]', fontsize=30)
    plt.ylabel('Acceleration, [m/s^2]', fontsize = 30)
    plt.grid(True, 'both')
    plt.show()

#-----Motif Definition-----#

#Cut out the Motif
Ref_Peaks(0)
print ("Start: ",end="")
start = int(input()) #specify first gait event
start = start - 20
print ("End: ",end="")
end = int(input()) #specify last gait event
end = end + 20

gaitCyclePattern = Ref_Acceleration_39[0][start:end] #define the motif
print ('Start: ' + str(start), 'End: ' + str(end))

#-----Matrix Profiling-----#

def SavGol_39 (i):
    #applying the Sav_Gol filter
    ACC_filtered = savgol_filter(df.averagea, 39, 3)
    #the filtered signal is stored into an array

```

```

    acceleration_39[i] = ACC_filtered
    return acceleration_39[i]
def PlotMatch(i):
    plt.plot(acceleration[i], label = 'Filtered data: ' + str(i+1))
    plt.plot(range(indx_array[i],indx_array[i] + len (gaitCyclePattern)),
acceleration[i][indx_array[i]:indx_array[i]+len(gaitCyclePattern)], label = 'Motif',
linewidth = 4)
    plt.suptitle('Match in ' + name + "'s dataset number: " + str(i+1), fontsize='30')
    plt.grid(True, 'both')
    plt.legend()
    plt.show()

path = Ref_path
files = os.listdir(path)
name = path.split('\\')
name = name[-1]
#Upload the data and apply Sav_Gol filter upon it
i = 0
for file in files:
    if file.endswith(".xlsx") and file.startswith(location):
        df = pd.read_excel(os.path.join(path, file))
        df.columns = ['time','stamp','battery',
'pressure','temperature','ax','ay','az','gx','gy','gz','mx','my','mz','averagea']
        acceleration[i] = SavGol_39(i)
        i += 1

#Run the matrix profiling through all datasets
i = 0
for i in range (0,10):
    distance_profile = stumpy.mass(gaitCyclePattern, acceleration[i])
    distance_profile_array[i] = distance_profile
    indx = np.argmin(distance_profile)
    indx_array[i] = indx
    i += 1

#Chop out the Motif
i = 0
for i in range(0,10):
    profile_match = acceleration[i][indx_array[i]:indx_array[i]+len(gaitCyclePattern)]
    profile_match_array[i] = profile_match
    i+=1

#find peaks in the chopped Motif
i = 0
for i in range (0,10):
    peaks,_ = find_peaks(profile_match_array[i], height=y)
    peaks_array_filtered[i]=peaks
    NumOfPeaks[i] = len(peaks)
    i+=1
#plot every dataset with peaks array
i = 0
for i in range (0,10):
    PlotMatch(i)
    print('Peaks: ', peaks_array_filtered[i])
    print('Number of Peaks in Profile Match: ', NumOfPeaks[i])
    i += 1

```