Thesis

One Step At a Time: Newly Proposed Gait Event Detection Using Position Benchmarked Against Existing Acceleration-Based Methods

Y.O.U.P. Mickers

Technische Universiteit Delft

TU Delft Delft University of Technology

Challenge the future

Thesis

One Step At a Time:
Newly Proposed Gait Event Detection Using Position Benchmarked Against Existing Acceleration-Based Methods

by

Y.O.U.P. Mickers

in partial fulfillment of the requirements for the degree of

Master of Science

in Mechanical Engineering,
Biomechanical Design (specialisation: BioRobotics)

at the Delft University of Technology,
Mekelweg 2,
2628 CD Delft

to be defended publicly on Friday December 4th, 2020 at 14:00.

Student number:	4246713	
Supervisor:	Prof. Dr. Ing. H. Vallery	TU Delft
Thesis committee:	Prof. Dr. Ing. H. Vallery	TU Delft
	Dr. Ir. drs. M. Kok,	TU Delft
	Ir. M.M. Reijne,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Delft University of Technology

PREFACE

I'm glad that I can finally present to you the final work of my master in Mechanical Engineering. This thesis started out at Motek Medical, where the original goal was to create an algorithm able to detect gait events on data coming from the Rysen [1] in rehabilitation setting. I would like to thank them for their original idea and their support throughout the initial phase of my thesis. Unfortunately, due to the COVID-19 pandemic we were not able to do the experiments that were planned (the experiment plan is included in appendix A). As such, we started looking at alternatives in the form of existing data sets, on which we would be able to benchmark acceleration-based and position-based gait detection algorithms. Several data sets were considered and analysed, but none were usable for our use case (for an overview of the data sets that were considered and the reason they were discarded, we would like to refer you to appendix B). The data set by Mundt [2] finally offered us a path to continue and I would like to thank her for her willingness to share the data set and generously answer all of my questions.

This work has taught me invaluable lessons, in writing, analysis and perseverance. None of it would have been possible without my supervisor, Prof. Dr. Ing. H. Vallery, to whom I would like to express my gratitude for her valuable insights and unrelenting support. In the end, we've successfully managed to benchmark and compare the acceleration-based algorithms, create position-based gait event detection methods, show their performance improvement and propose a method for detecting left and right steps. In short, showing that the full gait cycle can adequately be monitored with nothing more than tracking the upper-body position. I hope that one day, parts of this work might be used to improve the RYSEN again, completing the circle...

Youp Mickers, November 21st, 2020

Version	Date	Reviewers	Remarks
0.1	01 October 2020	Author	Request for feedback Friday meeting.
0.2	17 October 2020	Marco Reijne	First draft paper.
0.3	25 October 2020	Author	Incorporated feedback Marco (sections introduction/methodology).
0.4	7 November 2020	Author	Statistics added.
1.0	20 November 2020	Author	Final report.
1.1	21 November 2020	Author	Small improvements, frontpage added.

TABLE I
VERSIONING.

One Step At a Time: Newly Proposed Gait Event Detection Using Position Benchmarked Against Existing Acceleration-Based Methods

Youp Mickers

***Abstract*—Gait event detection allows for insight into one's gait pattern, an invaluable aid in rehabilitation. Current methods often rely on measured acceleration and rarely on position measurements [3]–[6]. In this paper we propose 4 novel gait detection methods based on the position of the Center of Mass (one approach being causal and thus suitable for real-time use) and compare them to 4 existing state-of-the-art acceleration-based methods. All algorithms are benchmarked on an existing data set (overground walking, 23 participants, 1772 steps), comparing the detection rate, false positive rate and the mean and (intra- and interparticipant) standard deviation of the timing error for Heel Strikes and Toe-Offs. We show that position-based algorithms give well-balanced results and are able to outperform the acceleration-based algorithms in all five metrics. Additionally, we propose and compare several methods for detecting left and right steps, thereby enabling quantification of the full gait cycle.**

Keywords: Gait Event Detection | Gait Phase Analysis | Acceleration-based | Position-based | Toe-Off | Heel Strike

I. Introduction

Gait contains a wide variety of information, both in everyday life and in the clinical setting and as such it can give us insight in a person's health. Baker [7] identified four reasons for gait analysis: diagnosis of a disease or injury, to assess the severity of a disease or injury, to monitor the progression of a disease or injury and to monitor the effect of an intervention or treatment. Additionally, knowing the precise progress of the gait cycle allows one to use this information in AR/VR (Augmented/Virtual Reality) applications or specialized balance exercises, such as disturbing a person at a specific phase of their gait cycle.

Knowing the current phase within the gait cycle through a straightforward method greatly improves its usability. Where Inertial Measurement Units (IMUs) are often used for this purpose, position-based methods have received less attention. Position measurements are often less noisy and we expect better accuracy due to this. Alternatively, they might be easier to obtain due to the specific setup (such as in the case of the RYSEN [1]). The European counterpart to GPS, Galileo, already offers sub-cm positioning to paying customers and systems with similar levels of precision focused on indoor environments are becoming available as well and might replace currently used motion capture systems in the nearby future.

A literature research revealed that the most sought-after events for gait analysis in the clinical setting were the Heel Strike (HS) and the Toe-Off (TO) (see figure 1). Algorithms extracting these events from upper-body acceleration exist [3]–[6]. To the authors' knowledge, no such algorithms have been proposed yet relying on position measurements of the upper-body. Position-based detection might offer superior performance, as position measurements tend to contain less noise than acceleration measurements. Additionally, it is a first step towards algorithms that fuse position and acceleration information.

Faculty of Mechanical, Maritime, and Materials Engineering - Biomechanical Engineering. TU Delft, 2628 CD Delft, The Netherlands

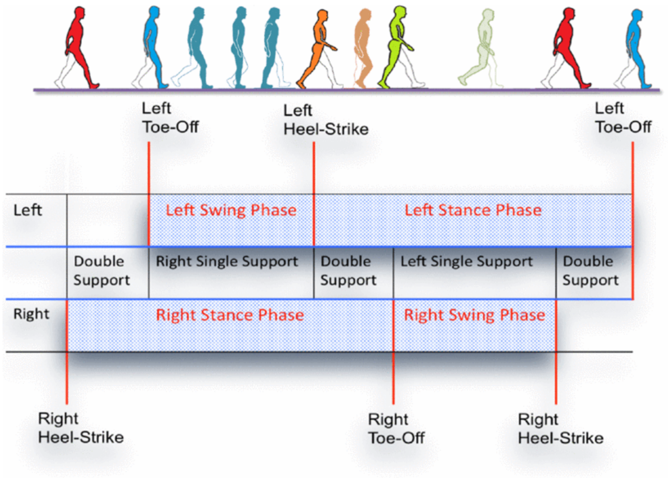

Fig. 1. Overview of the gait phases in healthy gait [8] (©2016 IEEE, reused with permission)

The goal of this paper was therefore to come up with one or more algorithms which are able to detect steps based on the position pattern of the Centre of Mass (CoM). Additionally, the constraint can be made that such an algorithm is causal, allowing use in on-line (possibly real-time) scenarios. To fully quantify the gait, a method classifying steps as either left or right is also proposed (the algorithms are explained in appendix C). The following research questions are defined and will be answered:

- How do the acceleration-based algorithms perform compared to one another?
- Is it possible to create a position-based algorithm to accurately quantify the gait cycle?
  - Can position-based algorithms improve gait event detection compared to acceleration-based methods?
  - Is such a method possible whilst being causal?
- Is a position-based algorithm's performance sensitive to the marker placement?

We will propose four algorithms, referred to as pos-AP, pos-Vert, pos-Fused and pos-RT (pos-RT being causal). Their performance will be compared to four existing state-of-the-art acceleration-based algorithms, by Zijlstra [3], Gonzalez [4], McCamley [5] and Shin [6] (see appendix D for more details), based on the metrics defined in section II-C. The performance will be benchmarked on an existing data set, shared by Mundt [2]. Note that of the acceleration-based algorithms, only the one by McCamley detects Toe-Offs. In addition to the detection of the gait events, a mechanism is proposed to label each step as either a left or a right step (the process of which will be referred to as *side detection*). This information allows for a reasonable view of the temporal characteristics of one's gait.

II. Method

A. Proposed algorithms

The proposed algorithms will be shortly introduced here. For a more extensive explanation, the reader is referred to appendix C.

1) Pos-AP: Pos-AP is based on the velocity in the anterior-posterior direction for the HS, assuming the forward velocity reaches a maximum at the moment of HS. A minimum in the AP acceleration is considered a TO.

2) Pos-Vert: Pos-Vert is based on the assumption that the CoM vertical movement resembles a sinusoid, reaching its lowest point at the moment of HS in normal gait. TO is detected by finding the minimum vertical acceleration following a HS.

3) Pos-Fused: Pos-Fused was created to mitigate the high number of false positives stemming from the pos-AP algorithm. It does this by running both pos-AP and pos-Vert and only reporting a HS if this is found by both algorithms, then taking the timing of the HS by pos-AP. TO is detected by the minimum in AP acceleration following a HS.

4) Pos-RT: Pos-RT is created to be causal and thus applicable in online situations. A maximum in the AP velocity is taken as HS (with the added heuristic that the AP acceleration was sufficiently positive over the previous samples). TO is taken as the zero-crossing of the AP acceleration from positive to negative.

5) Side Detection: Two versions of side detection are proposed, both based on the modeling of the mediolateral swaying motion as an (inverted) pendulum. By looking at the cumulative value of the signal over the preceding samples we can see whether the velocity or acceleration is predominantly in one direction, which it turn allows for classifying the steps.

B. Existing algorithms

The existing algorithms will be shortly introduced here. For a more extensive explanation, the reader is referred to appendix D.

1) McCamley: McCamley [5] is based on the acceleration in the vertical direction. The signal is integrated and then differentiated using a Gaussian Wavelet Transform (GWT), effectively smoothing the signal, after which the minima are taken as HS. After a further differentiation using GWT, the maxima are taken as TO.

2) Zijlstra: Zijlstra [3] was one of the first algorithms detecting HS using upper-body acceleration measurements. A low-pass filtered (2 Hz) version of the AP signal is used, with the peak preceding each zero-crossing (from positive to negative) being taken as a HS.

3) Shin: Shin [6] is the only causal acceleration-based algorithm and the only algorithm not requiring separated acceleration components, instead using the magnitude of the total acceleration. The total acceleration is smoothed (using a sliding window summation) and then differentiated, after which each zero-crossing (negative to positive) is taken as a HS.

4) Gonzalez: Gonzalez [4] is based on Zijlstra and thus relies on the AP signal. It adds some further heuristics based on the vertical acceleration.

C. Outcome measures

As mentioned, we are primarily interested in two types of events; heel strikes and toe-offs. For both of them, we will assess the quality of their detection based on five metrics, being the detection rate (also known as sensitivity), false positives rate, the mean (absolute) error in milliseconds and the variance of the timing error (intra- and inter-participant).

1) Detection rate: The detection rate (also known as the sensitivity) is defined as the ratio of the number of correctly detected events to the number of total events. An event is considered to be correctly detected if the absolute error is smaller than 300 ms. A higher detection rate is considered positive.

2) False positive rate: The false positive rate is defined as the ratio of the number of false positives to the total number of events. The often used counterpart of sensitivity, specificity, cannot be used directly here, as there are no negative events in the traditional sense. This concept is used instead. A false positive is defined as being more than 300 ms from an event, or if multiple events are detected for an event (with each extraneous event counting as one false positive). It is favourable to have a low false positive rate.

3) Timing (mean and variance): The timing metrics that are of interest are considered to be the mean error and the standard deviation of the error. This results in effectively two variances; intra-participant (how much do the timing errors differ within a single participant) and inter-participant (how consistent is the mean timing error between different participants). In case multiple detected events are within the window that is considered matching (being false positives), the timing of the detected event closest to the actual event is taken. A low Mean Absolute Error (MAE) is considered desirable, as are low variances.

D. Comparisons

Comparing the algorithms will be done according to following benchmarking strategies. For all comparisons, the steps of each participant will be grouped as a single case. All metrics are thus calculated per person, with the exception of the *effect of gait velocity*.

1) Asymmetric performance acceleration-based methods: During the exploratory phase it was found that the asymmetric attachment of the IMU results in a noticeable performance difference between left and right steps. As such, it was decided that the results will be shown as a whole and additionally separated per side. In order to make the comparison between acceleration-based and position-based algorithms (as the latter does not depend on an asymmetrically located marker) fairer, all graphs will show the performance for left/right steps separately, as well as combined.

2) Performance heel strike detection: All algorithms will be compared on their performance of the heel strike detection. A heel strike is defined as correctly detected if the absolute error is below 300 ms (based on a minimum step duration of 0.6 seconds [9]).

3) Performance toe off detection: The algorithms that detect toe-offs (all four position-based algorithm and McCamley), will be compared on their performance of the toe-off detection. A TO is defined as correctly detected if the absolute error is below 300 ms (based on a minimum step duration of 0.6 seconds [9]).

4) Performance side detection: The side detection is a binary classification (left or right) and the metric used will be the number of events correctly classified. That is, for each heel strike, we run the algorithm (described in appendix C) to detect whether it is a left or a right step. The number of correctly classified steps is then divided by the total steps. As the number of steps on each side will be roughly equal and their classification can be considered equally important (misclassifying right steps is no worse than misclassifying left steps), it is not necessary to apply a weighting, or use concepts akin to the F1-score [10]. This is done for the approach based on the mediolateral velocity and for the approach based on the mediolateral acceleration, acquired by differentiating the position respectively once and twice. Both approaches will be tested at 5 different durations, being 100, 150, 200, 250, and 300 ms, resulting in a total of 10 conditions.

5) Effect of gait velocity: According to the original paper the data set is split over 5 gait velocities (every participant has 10 trials per velocity), at 0.8 m/s, 1.1 m/s, 1.4 m/s, 1.7 m/s and 2.0 m/s ($\pm$ 10%). To measure the effect of gait velocity on the performance of the algorithm, each velocity will be handled as a separate condition and metrics will be calculated per velocity instead of per participant.

The information stating which gait velocity was aimed at was however not included in the data set. Efforts to extract this information using the displacement of certain markers over time also did not lead to a conclusive and accurate categorization and as such, this comparison will be skipped.

6) Robustness to marker attachment: The robustness of the position-based algorithms to the exact location of the marker will be benchmarked by comparing the performance on the virtual marker representing the CoM to a marker further away from the CoM. The marker that will be used is located on the RASI (indicated in figure 2).

E. The Dataset

The dataset was shared by Mundt [2] and contains 1112 trials of 23 participants (12 female, 28.1$\pm$6.0 years, 72.3$\pm$12.7 kg, 1.77$\pm$0.07 m) walking a 10m walkway. For an overview of other data sets that were considered and the reason they were discarded, we would like to refer you to appendix B.

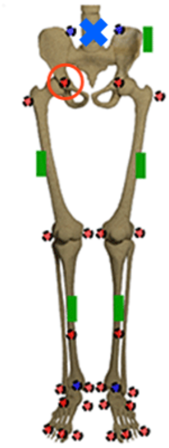

Fig. 2. Overview of the marker and accelerometer locations. The red (front) and blue (back) dots represent the MoCap markers, the green boxes the attached IMUs. The blue cross is the virtual point created by taking the middle of the two directly adjacent markers. The red circled marker is the RASI and used for testing the robustness to marker location *(by Mundt, licensed under CC BY 4.0)* [2].

1) Marker and IMU attachment: From each trial, the following measurements are available;

- tri-axial acceleration at 5 points (tri-axial IMUs)
- position of 28 markers (motion capture)
- tri-axial force data from zero, one or two force plates

The IMUs were positioned on the left and right shank, left and right knees and the left hip and sampled at 100 Hz (as shown in figure 2), meaning IMU that is used is not attached symmetrically, like the MoCap markers that are tracked. The markers are located according to an anatomical model based on the recommendations of the International Society of Biomechanics (ISB) [11] and the motion capture system (VICON^{TM} MX F40, Oxford, UK) is sampling at 100 Hz. The CoM is taken as the middle between the RPSI and LPSI marker.

2) Data processing: The number of force plates varies between trials and every force plate captures a single step, meaning that every trial contains zero, one or two steps. In total, the set contained 1772 steps and after an initial exploratory phase analysing the suitability of the data set, the set was divided into 894 steps for the analysis phase (with which the algorithms were developed) and 878 steps for the validation phase. This division ensured the algorithms were not tuned specifically to the data set and was done by using the random method from the Python Numpy-package by;

- Sorting all trials first by subject, then by trial number (both ascending)
- Seeding Numpy.random with 101 using *np.random.seed(101)*
- Assigning the sorted items a number (0 or 1) using *np.random.randint(2)*
- Splitting the items in the analysis (0) and control group (1)

All signals were synchronized according to the method described in the original paper [2]. Those signals were then fed to the algorithms, with no filtering applied for the acceleration and position data and 6 Hz low-pass filtering applied for the force plates.

3) Golden standard extraction: The ground truth of the events was extracted from the force plates, and as such is independent from both the motion capture and the acceleration measured by the IMUs. For the heel strike extraction, the following methodology is used: The peak value of the force is detected, after which the moments where the value first reaches 10% and 90% of this peak are taken (thresholds commonly used to define the rise time). The moment of heel strike is then defined as the middle of these moments. For the toe-off event extraction, the moment where the signal drops to 10% of the peak again is taken.

F. Statistical analysis

Four differences will be statistically investigated for their significance, all at the p≤0.05-level. This will be done using the three timing-related metrics, being the mean timing error, the intra-participant variance and the inter-participant variance. These metrics will be considered as having a normal distribution, an assumption we will inspect visually and test using a Shapiro-Wilk test. After this a pair-wise Welch's t-test [12] (thereby incorporating potentially unequal variances) will be done for the mean timing error and the intra-participant variance, while the inter-participant variance will compared using a Bartlett test [13]. We will apply a standard Bonferroni correction to minimize the chance of making a type I error, meaning a difference is considered statistically significant when a p-value lower than 0.0167 is found.

The following four comparisons will be examined for their significance:

- Left steps vs. right steps for the acceleration-based algorithms
- Comparison HS detection between pos-Fused/pos-RT and the acceleration-based methods
- Comparison TO detection between pos-Fused/pos-RT and the acceleration-based methods
- Virtual CoM marker vs RASI marker position-based algorithms

III. RESULTS

In this section, the first paragraph will analyze the asymmetric performance of the acceleration-based algorithms, after which the second and third paragraph will treat the detection of heel strikes and toe-offs in all algorithms. Side detection is treated in the fourth paragraph, while the fifth paragraph will analyze the robustness of the position-based algorithms.

A. Asymmetric performance acceleration-based methods

As can be seen in figure 3, the asymmetric attachment of the IMU results in a noticeable performance difference between left (red) and right (blue) steps for the acceleration-based algorithms (the IMU is positioned on the left hip). As such all subsequent graphs will, in addition to the combination (purple), also show both separately. This will allow for a fairer comparison between acceleration-based and position-based algorithms (the latter depends on a symmetrically located virtual marker).

B. Performance heel strike detection

Looking at the acceleration-based algorithms in figure 3 it is immediately visible that Gonzalez [4], detects only slightly more than half of the heel strikes for most participants (49.7±13.0%). Shin has a higher detection rate (99.4±1.3%), but also generates a higher amount of false positives (24.4±9.6%). Both Zijlstra and McCamley have high detection rates and low false positive rates. McCamley has a slightly higher detection rate (99.7±0.8% vs 97.2±3.6%), a smaller MAE (Mean Absolute Error) (-3.14±18.7 ms vs -32.3±16.3 ms), and a lower intra-participant variance for the error (std: 56.3±13.6 ms vs 58.2±14.2 ms), but Zijlstra generates a lower number of false positives (3.9±4.6% vs 1.5±2.2%).

The non-causal position-based algorithms outperformed the acceleration-based algorithms on most metrics. Pos-AP has the second-lowest inter-participant variance at 7.81 ms (and lowest intra-participant variance, 14.74 ms), but generated a large number of false positives (33.7±7.6%). The pos-Vert generates a very low number of false positives (0.90±2.2%), but has a higher variance. The pos-Fused manages to combine their properties, while also having a superior MAE, although the (intra-participant and inter-participant) variance is slightly higher than for pos-AP.

Finally, the causal position-based algorithm pos-RT performs well on detection rate and timing (having the lowest MAE and the second-lowest intra-participant variance). This is at the cost of a slightly higher false positive rate (3.55±6.49 %).

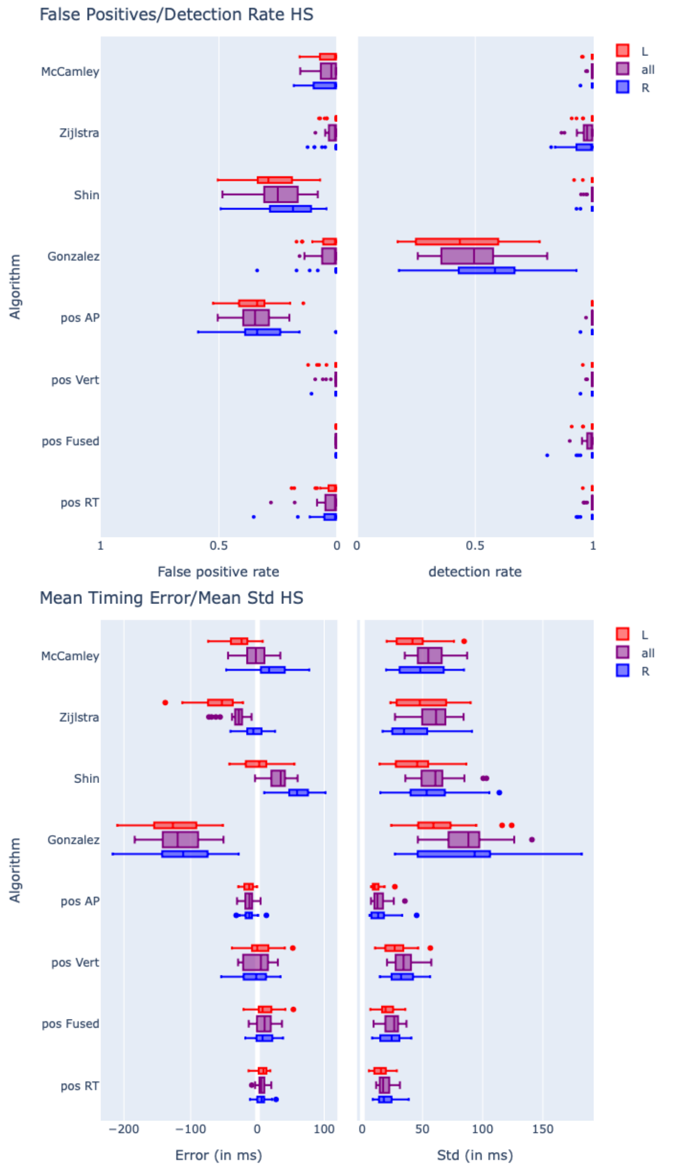

Fig. 3. Comparison of all position-based and acceleration-based algorithms for HS. In the top row, the *false positive rate* and *detection rate*. The bottom row shows the *mean timing error* and the *std of the timing error*

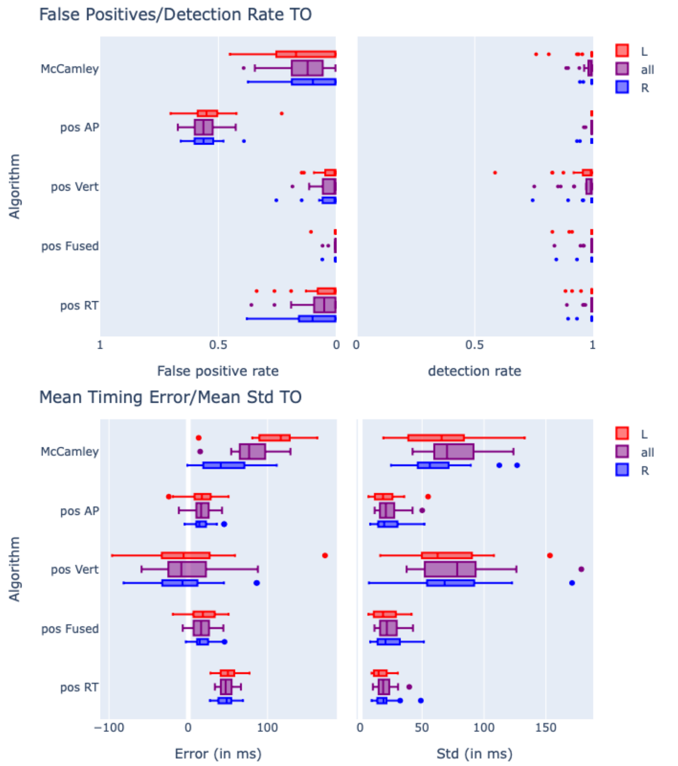

Fig. 4. Comparison of McCamley and the position-based algorithms for TO. In the top row, the *false positive rate* and *detection rate*. The bottom row shows the *mean timing error* and the *std of the timing error*

C. Performance toe off detection

Of the acceleration-based algorithm, only McCamley includes TO detection as can be seen in figure 4. For the non-causal position-based algorithms a comparable situation arises as with the HS detection. Pos-AP has a lower variance, but generates a large number of false positives. Again, the pos-Fused manages to combine the properties of the pos-AP and pos-Vert. The position-based algorithms outperform McCamley on most metrics and shift the mean error from 79.3±24.3 ms to -0.81±34.8 – 48.28±8.83 ms.

Looking at pos-RT, the detection rate is high (99.1±2.33%), albeit a relatively high false positive rate is observed (7.2±9.0%). The timing is interesting, having the highest MAE of all position-based algorithms (48.3±8.83 ms) and simultaneously the lowest inter-participant variance of all algorithms (8.83 ms).

D. Performance side detection

In figure 5 it can be seen that velocity-based side detection has higher classification rates over a longer period (300 ms results in a 98.8-99.5% correct rate), although there is little difference with the 200 ms (98.0-99.6%) and 250 ms (98.4-99.7%) situation. On the other hand, acceleration-based side detection has higher classification rates when a shorter period is used (100 ms results in a 79.4-83.2% correct rate). Comparing the two, velocity-based side detection classify more steps correctly than acceleration-based detection.

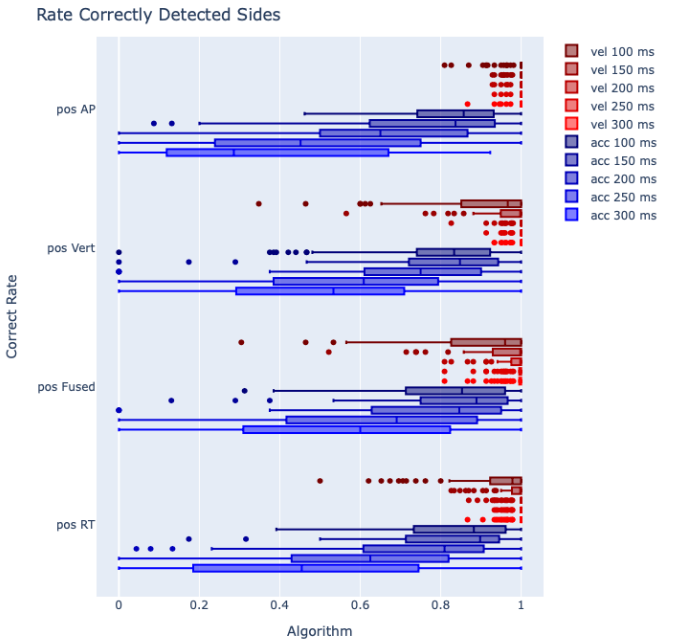

Fig. 5. Comparing the performance of velocity-based and acceleration-based side detection at various timings.

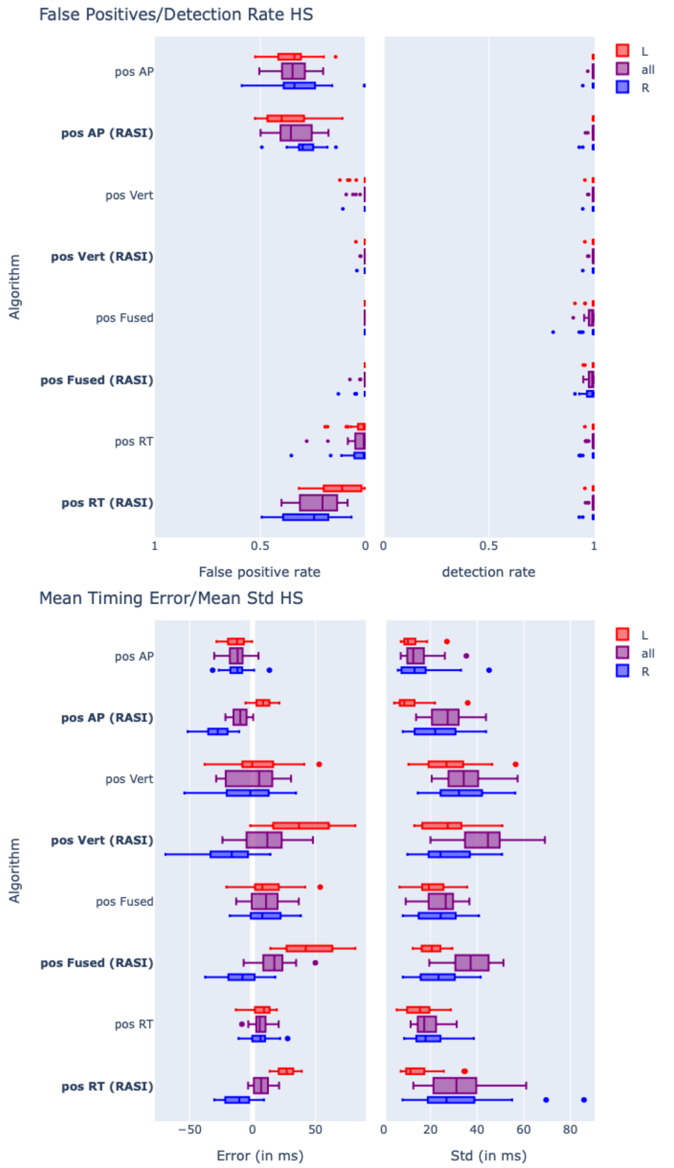

Fig. 6. Comparing the robustness of the algorithms by using a marker located further from the CoM (RASI). In the top row, the *false positive rate* and *detection rate*. The bottom row shows the *mean timing error* and the *variance of the timing error*

E. Robustness to marker attachment

In order to test whether the proposed algorithms are robust to the exact location of the markers, they are run on one of the other markers attached to the pelvis (RASI, located on the front right). It can be seen (figure 6) that the position-based algorithms still work and have comparable detection and false positive rates, with the exception of an increase in the false positive rate for pos-RT (going from 3.55±6.49% to 21.2±9.7%). The timing of the errors tells a different story and a strong asymmetry arises between left and right steps. Interesting to note is the timing error of the left step for pos-AP, which shows both a very low mean error and a lower variance (intra-participant and inter-participant) for the RASI marker than for the central marker (a similar pattern is seen with right steps and the LASI marker).

F. Statistical analysis

In this section 4 comparison are evaluated for their statistical significance. A result is found to be significant at p≤0.05-level and involves 3 metrics, meaning the required p-value is set to 0.0167 after a Bonferroni-correction.

1) Left vs. right steps: Looking at table II in appendix E it is visible that for 3 acceleration-based algorithms (the exception being Gonzalez), there are significantly different results for left vs. right steps. This is in all cases due by the mean error, although Zijlstra additionally has a significantly different inter-participant variance between left and right steps.

2) Performance heel strike detection: From table III in appendix E, it can be seen that there is a significant difference in performance between all 28 algorithm pairs. In 22 pairs this is caused by the mean errors, in 24 pairs by the intra-participant variance and in 17 cases by the inter-participant variance.

Comparing pos-Fused with the acceleration-based methods we can see that the mean error and intra-participant variance are significantly different from all acceleration-based algorithm, and the inter-participant variance from Gonzalez. For pos-RT, all three metrics (mean error, intra- and interparticipant) variance are significantly different from all acceleration-based algorithms, with the exception of the inter-participant variance with McCamley.

3) Performance toe-off detection: From table IV in appendix E, it can be seen that there are significant differences in performance in TO detection for 9 out of 10 algorithm pairs, the exception being Pos-AP vs Pos-Fused. For 7 pairs this is caused by the mean error, for 6 pairs by the intra-participant variance and 6 pairs show significant differences in the inter-participant variance. Comparing pos-Fused and pos-RT with the only acceleration-based algorithm (McCamley), we can see that for both on all three metrics the differences can be considered statistically significant.

4) Performance marker location: Comparing the performance of the position-based algorithms on the virtual CoM vs. the asymmetric RASI marker, table V in appendix E shows the performance can be considered significantly different based on the intra-participant variance for all 4 algorithms. None of the other differences in metrics can be considered significant at the $p \leq 0.05$-level.

IV. Discussion

A. Research questions

Using the results we will now answer the defined research questions.

How do the acceleration-based algorithms perform compared to one another?
Of the acceleration-based algorithms, Shin and Gonzalez are largely unsuitable for general use. Gonzalez due to the low detection rate, Shin due to high number of false positives. McCamley and Zijlstra would be advisable, depending on the exact needs of the application. McCamley has a better detection rate, a lower mean error and lower variances. Additionally, McCamley is also able to extract heel strikes. Zijlstra has a slightly lower false positive rate, which might be more favourable in certain applications.

Is it possible to create a position-based algorithm to accurately quantify the gait cycle?
Position-based methods were able to detect both gait events (HS and TO) that were considered necessary. Additionally, position-based methods were able to classify steps as either left or right, scoring ≥98% with all algorithms, when the *velocity, 300 ms*-method is used. This means the full gait cycle can indeed be quantified using position-based methods.

Can position-based algorithms improve gait event detection compared to acceleration-based methods?
In general, the position-based algorithms show great potential and outperform the acceleration-based algorithms on all metrics. Especially the increased consistency is remarkable. Both within a participants and between participants, the variance is much lower, which is arguably more important than the mean timing error (as the latter is easily corrected for in off-line applications). Pos-AP can be considered unsuitable due to the high amount of false positives, but its timing error indicate anterior-posterior movements might still be the most precise, as was also shown in the literature review. A way is needed to remove false positives, as is done in pos-Fused, improving the false positive rate, at the cost of a small increase in variance. For the TO, similar patterns are seen as with HS. We can conclude that pos-Fused is able to detect gait events better than acceleration-based methods.

Is such a method possible whilst being causal?
Pos-RT is causal and shows good results for HS, although the false positive rate is somewhat higher than for the non-causal position-based methods (and comparable to the acceleration-based methods). The performance on TO results in a similar number of false positives and the mean error is relatively high compared to other position-based algorithms, but the (inter- and intra-participant) variance shows good consistency. The *side detection* proposed is also causal.

Is a position-based algorithm's performance sensitive to the marker placement?
From the results it can be seen that the position-based algorithm are still able to detect gait events with a high accuracy when an asymmetric marker location is used. The performance in timing is somewhat degraded, but we have shown that the position-based algorithms performances are robust to a strongly asymmetric (RASI) attachment, albeit with a decreased performance in timing.

B. Real-time usage

An important boundary condition for real-time usage of an algorithm is causality. As three of the algorithms (pos-AP, pos-Vert, pos-Fused) rely on bidirectional filtering for their analysis, they depend on future values, making them non-causal. Pos-RT relies on a rolling average over 5 samples, followed by a peak detection over 5 samples (in both directions). As such, the delay is 10 samples (100 ms at the 100 Hz sampling rate used), which would be usable in a real-time context. We expect this delay could be further improved, for example by increasing the sampling rate, or by further tuning the peak detection.

The side detection relies only on past values and does not apply any form of peak detection. This makes the necessary delay 0 ms and means it can be used in real-time.

C. Force Plate Filtering

The force plate data has been (bi-directionally) low-pass filtered, smearing the signal somewhat. As a result of this limitation in the data set, the average rise time (going from 10% to 90% of the maximum force) over all steps is 12.8 samples (128 ms). This is relatively large and results in an uncertainty in extracting the exact moment of the gait events. The effect is largest in the mean error, as this could "shift" the results of all algorithms. The spread of the means (inter-participant) and the standard deviation (intra-participant) are not affected though and it can be seen that position-based algorithms tend to perform more consistent from them.

D. Location marker vs. IMU

An important issue with the data set is the asymmetric attachment of the IMU versus the symmetric position of the virtual marker. From the results it can clearly be seen that steps on the opposite side of the IMU are generally detected better, affecting both detection rate and timing error, the extent of this effect varying with the algorithms. Although we try to compensate for this by comparing both the average and best-case scenarios for the acceleration-based algorithms, there is simply no way to know if the acceleration-based algorithm would have performed better with a more symmetric attachment. In the opposite sense, we have shown that the position-based algorithms performances are robust to a strongly asymmetric (RASI) attachment.

E. Further Research

This is to author's knowledge the first time gait event detection is done using position-based methods. For further applications, for example in rehabilitation, several aspects can be explored further.

1) Dynamic Thresholding: A strategy that is used often is thresholding, in which an event is not considered if the signal does not have a certain minimum (or maximum) value. This threshold varies between persons and over different gait velocities (peak acceleration values for example vary greatly with gait velocity [14]), making it unfeasible to pick one value which works over a range of conditions. A solution is dynamic thresholding, as applied by Yang [15], in which the threshold value is based on the maximum value encountered over the previous n samples. This approach is promising, especially in steady-state gait, but the current data set is not feasible to test this on. This due to the fact that often the data signal only extends a short time before the registered gait events, giving no possibility to accurately set the dynamic threshold. Still, we expect the concept could be used to improve the proposed algorithms.

2) Pos-AP asymmetric attachment performance: An interesting result of pos-AP is the increase in timing accuracy when the marker is placed asymmetrical. Steps on the opposite side of the marker attachment are detected with the lowest variance of all measured conditions (algorithm, marker position, step side) and a very low mean error. This could certainly be explored further and a method relying on two markers (RASI and LASI), might be used to improve the algorithm further.

3) Effect of gait velocity: As stated, the original goal was to also compare the performance at various gait velocities as this effects the gait characteristics [15], but this information is lacking from the data set. Still, further research should be done to asses the robustness of the various algorithms. In revalidation applications, it might make sense to choose algorithms that are strong with lower gait velocities, as this is likely to occur.

4) Gait Pathologies and Use in Revalidation: The data set contains only normal walking on a flat plane, which can reasonably be assumed to have reached steady-state. For further use in medical settings, the algorithms could further be validated on a variety of gait pathologies (see also appendix A) and non-transient gait. Examples of the latter would include starting walking from standstill, people performing exercised walks, or people being disturbed at specific moments in their gait cycle.

V. Conclusion

In this paper, we have shown that a single marker tracking the position of the CoM can be used as a viable alternative to an accelerometer for the detection of gait events. Four algorithms were proposed and their performance on the dataset was compared, with position-based algorithms outperforming acceleration-based algorithms on key parameters. Additionally, side detection was found to be reliable, meaning the entire gait cycle can be accurately monitored using position-based measurements.

Within the acceleration-based methods Gonzalez and Shin are considered generally unsuitable; Gonzalez due to the low detection rate (49.7$\pm$13.0%) and Shin due to the high false positive rate (24.4$\pm$9.6%). McCamley seems the best overall method here, with Zijlstra being a close second with different strengths.

Four position-based methods are proposed; pos-AP, pos-Vert, pos-Fused (combining the aforementioned two) and a causal pos-RT. Pos-Fused shows great overall potential, with a detection rate of 98.8% and being the only algorithm with a 0.0% false positive rate. On the timing, it outperforms 3 of the 4 acceleration-based algorithms in mean error and has both a lower intra- and interparticipant variance than all acceleration-based algorithms. Pos-AP has slightly lower (inter- and intra-participant) variances in mean error, but at the cost of a high false positive rate. Pos-RT proves reliable for real-time settings; again, 3 of the 4 acceleration-based algorithms are outperformed in mean error and all acceleration-based algorithms in both intra- and interparticipant variance.

In detecting TO, the position-based algorithms also surpasses acceleration-based methods (McCamley being the only one detecting TO, at 98.5% with 13.7% FP). Pos-Fused again manages to merge the advantages of pos-Vert and pos-AP, detecting 98.8% of the events at a 0.37% false positive rate (the lowest of all algorithms). Pos-RT detects 99.1% with 7.2% false positives. On the timing, both pos-Fused and pos-RT report a lower mean absolute error and much smaller intra- and inter-participant variances than McCamley.

Side detection was examined and shown to work correctly with all four position-based algorithms. In the end the velocity-based method, derived by differentiation, over the previous 300 ms proved the most accurate (98.8-99.5% correct rate). Velocity-based provided higher accuracy with longer times, while acceleration-based profited from shorter times.

The proposed algorithms were shown to be relatively robust to the positioning of the marker, although timing precision was lower for the (strongly) asymmetrical attachment. The exception to this was the pos-AP algorithm, whose precision showed a remarkable increase for the timing of steps on the opposite side.

Concluding, full gait cycle quantification using position-based methods is feasible and potentially superior to acceleration-based methods.

Acknowledgments

I'm indebted to many people for the completion of this project. First and foremost my supervisor Heike Vallery, for whom I cannot express enough gratitude for offering countless hours of invaluable advice and insights and without whom this endeavour would have been far less educative experience. Other people I would like to thank in no particular order;

- Andrew Berry for proofreading my literature review and general remarks.
- Marco Reijne for proofreading and general remarks.
- Yan Zhao, Coen de Weerd, and their colleagues at Motek Medical, for coming up with the initial project and their help in shifting when this became impossible due to Covid-19.
- Marion Mundt, for generously sharing the data on which this project is relying and answering all my questions regarding this data set.

Finally I would also like to thank Aneesh, Ashwin, Nathan and Suzanne, my fellow students who offered advice and/or proofread earlier versions.

References

[1] M. Plooij, U. Keller, B. Sterke, S. Komi, H. Vallery, and J. von Zitzewitz, "Design of RYSEN: An intrinsically safe and low-power three-dimensional overground body weight support," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2253–2260, Jul. 2018. [Online]. Available: https://doi.org/10.1109/lra.2018.2812913

[2] M. Mundt, A. Koeppe, S. David, T. Witter, F. Bamer, W. Potthast, and B. Markert, "Estimation of gait mechanics based on simulated and measured IMU data using an artificial neural network," *Frontiers in Bioengineering and Biotechnology*, vol. 8, Feb. 2020. [Online]. Available: https://doi.org/10.3389/fbioe.2020.00041

[3] W. Zijlstra and A. L. Hof, "Assessment of spatio-temporal gait parameters from trunk accelerations during human walking," *Gait & Posture*, vol. 18, no. 2, pp. 1–10, Oct. 2003. [Online]. Available: https://doi.org/10.1016/s0966-6362(02)00190-x

[4] R. C. González, A. M. López, J. Rodriguez-Uría, D. Álvarez, and J. C. Alvarez, "Real-time gait event detection for normal subjects from lower trunk accelerations," *Gait & Posture*, vol. 31, no. 3, pp. 322–325, Mar. 2010. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2009.11.014

[5] J. McCamley, M. Donati, E. Grimpampi, and C. Mazzà, "An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data," *Gait & Posture*, vol. 36, no. 2, pp. 316–318, Jun. 2012. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2012.02.019

[6] S. H. Shin and C. G. Park, "Adaptive step length estimation algorithm using optimal parameters and movement status awareness," *Medical Engineering & Physics*, vol. 33, no. 9, pp. 1064–1071, Nov. 2011. [Online]. Available: https://doi.org/10.1016/j.medengphy.2011.04.009

[7] B. R, E. A, B. MG, and D. K., "Gait analysis: clinical facts," *Eur J Phys Rehabil*, vol. 52, no. 4, pp. 560–574, Aug. 2016. [Online]. Available: https://doi.org/10.1016/s0021-9290(01)00222-6

[8] S. Chen, J. Lach, B. Lo, and G.-Z. Yang, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 6, pp. 1521–1537, Nov. 2016. [Online]. Available: https://doi.org/10.1109/jbhi.2016.2608720

[9] G. Panahandeh, N. Mohammadiha, A. Leijon, and P. Handel, "Continuous hidden markov model for pedestrian activity classification and gait analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 1073–1083, May 2013. [Online]. Available: https://doi.org/10.1109/tim.2012.2236792

[10] L. Derczynski, "Complementarity, F-score, and NLP evaluation," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 261–266. [Online]. Available: https://www.aclweb.org/anthology/L16-1040

[11] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D. D. D'Lima, L. Cristofolini, H. Witte, O. Schmid, and I. Stokes, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part i: ankle, hip, and spine," *Journal of Biomechanics*, vol. 35, no. 4, pp. 543–548, Apr. 2002. [Online]. Available: https://doi.org/10.1016/s0021-9290(01)00222-6

[12] B. L. WELCH, "THE GENERALIZATION OF 'STUDENTs' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARLANCES ARE INVOLVED," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947. [Online]. Available: https://doi.org/10.1093/biomet/34.1-2.28

[13] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 160, no. 901, pp. 268–282, 1937. [Online]. Available: http://www.jstor.org/stable/96803

[14] T.-H. Hwang, A. O. Effenberg, and H. Blume, "A rapport and gait monitoring system using a single head-worn IMU during walk and talk," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, Jan. 2019. [Online]. Available: https://doi.org/10.1109/icce.2019.8662087

[15] M. Yang, H. Zheng, H. Wang, S. McClean, and D. Newell, "iGAIT: An interactive accelerometer based gait analysis system," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 715–723, Nov. 2012. [Online]. Available: https://doi.org/10.1016/j.cmpb.2012.04.004

[16] J. K. Moore, S. K. Hnat, and A. J. van den Bogert, "An elaborate data set on human gait and the effect of mechanical perturbations," *PeerJ*, vol. 3, p. e918, Apr. 2015. [Online]. Available: https://doi.org/10.7717/peerj.918

[17] 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. [Online]. Available: https://doi.org/10.1016/j.gaitpost.2016.09.023

[18] R. Moe-Nilssen and J. L. Helbostad, "Estimation of gait cycle characteristics by trunk accelerometry," *Journal of Biomechanics*, vol. 37, no. 1, pp. 121–126, Jan. 2004. [Online]. Available: https://doi.org/10.1016/s0021-9290(03)00233-1

Appendix A
Original Experiment Plan

A. Data Collection

Biometric information. Both the length and the weight of the person will be recorded.

Time series will be collected of the following variables:

- IMU in the slingbar: linear acceleration
- IMU in the slingbar: angular rate
- Sensor in the slingbar: absolute angle between the slingbar and the room (direction the patient is facing)
- Internal RYSEN data: force in the cables
- Internal RYSEN data: position of the slingbar, calculated from the lengths of the cables
- Marker positions: using a motion capture system.

All time series will be collected by D-Flow at rates of up to 1000Hz. D-Flow will also handle the synchronization of the signals.

B. Walking patterns

Several walking patterns are planned throughout the room in order to get a good estimate of the RYSEN's influence. Due to the weight being carried by the cables, which are at different angles depending on the location in the room, the exact forces experienced by the patient might vary. As such, both breadthwise strides and lengthwise strides are considered. The Zigzag walking is meant to emulate typical usage of the RYSEN within the rehabilitation process. All courses will be created using flat circles or small cones, further clarified using arrows applied on the floor if necessary.

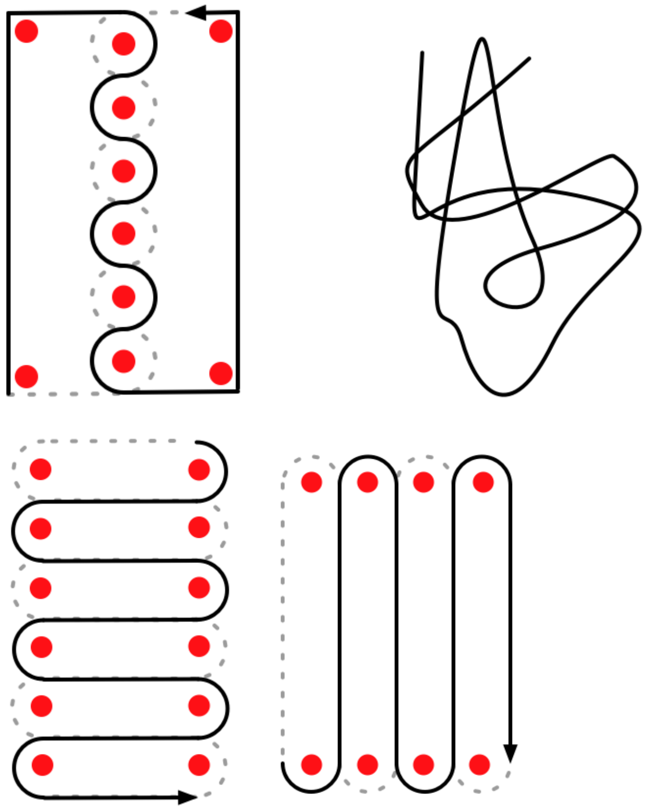

Fig. 7. Different walking patterns through the room. From left to right: (A) Zigzag walking (B) Random walking (C) Breadthwise strides (D) Lengthwise strides

C. Gait types

Several types of gait are considered for the experiment, meant to approximate various forms of pathological gait. Testing how robust the algorithms are to atypical forms of gait, which are often encountered during rehabilitation. We are interested in asymmetric forms of gait (such as a dragging leg), or gait with low accelerations (as most algorithms from literature are acceleration-based). Examples of the latter would be careful walking, which we intend to test by having the participant walk with a full glass of water or a table-tennis ball on a spoon. "Shuffling walk", often seen in patients with Parkinson disease, is also a form of gait which tends to have low accelerations.

- Normal walking
- Careful walking
- Dragging leg
- Shuffling walk (Parkinsonian gait)

Appendix B
Other Datasets

During this project, several other data sets were considered and examined. We will shortly introduce them and explain why they were deemed unsuitable for this purpose.

A. Internal Motek dataset

As mentioned during the introduction, the project started out as a gait detection project for the RYSEN. During preliminary runs of the experiment, it was found out that the RYSEN had a sampling issue which resulted in data being sampled irregularly. Instead of the supposed 1000 Hz, a sampling rate closer to 25 Hz was experienced, with a large variance in frame size (the longest frames being up to 200 ms). When we eventually managed solved this issue together with their team, we were no longer able to do experiments with participants due to the pandemic. As such we turned our efforts to existing data that was collected by Motek over the years.

It was noted that all internal data that was collected by the RYSEN suffered from this issue. Several smoothing, interpolating and filtering techniques were tried, but we were not able to get accurate signals to run the algorithms on.

B. Perturbed walking dataset

The perturbed walking data set by Moore [16] contains treadmill walking by 15 subjects at 3 speeds. Each trial consists of 120 seconds of normal walking, followed by 480 second of ”walking while being longitudinally perturbed during each stance phase with pseudo-random fluctuations in the speed of the treadmill belt” (for a total of ¿5000 gait cycles of unperturbed and 20000 gait cycles of perturbed walking). It contains marker data from 47 markers, sampled by an Osprey camera motion capture system at 1000Hz. Additionally, two 6 DoF force plates are located under the walking area of the R-Mill treadmill, giving the ground truth.

No acceleration data was available, so we tried to estimate the acceleration using double differentiation of a marker's position (including the THEAD marker, the SACR marker and the point between the RPSIS and LPSIS marker). The measurments were noisy, and differentiation often exacerbates noise. This meant we had to either work with very noisy acceleration estimates, or a signal that was so heavily filtered that it contained filtering artifacts such as oscillations. In the end it was decided that a good comparison between position-based and acceleration-based algorithms was not possible this way.

C. MAREA Gait Database

The MAREA Gait database [17] is a data set containing free and treadmill walking of 20 participants (12 males and 8 females, average age: 33.4 $\pm$ 7 years, average mass: 73.2 $\pm$ 10.9 kg, average height: 172.6 $\pm$ 9.5 cm). 11 of the participants completed the indoor treadmill protocol, which consists of:

- 10 minutes of flat walking (4.0 km/h - 8.0 km/h, in increments of 0.4 km/h)
- 12 minutes of sloped walking at self-selected speed (with a slope of 5, 0, 10, 0, 15 and 0 degrees, 2 minutes each)
- 6 minutes of walking at a self-selected pace (3 minutes walking, 3 minutes running)

Ground truth is based on insole forceplates, as the shoes were instrumented with piezo-electric force sensitive resistors (FSRs), fixed at the extreme ends of the sole. 3 IMUs, located at the navel, left wrist and left ankle provide acceleration measurements. The database is not open source, but access can be requested and will be granted for academic research.

While we got several of the algorithms running on the data set, we eventually noted unexpected discrepancies in our results. Upon further inspection, it seemed that the signals (acceleration measurements and the force plate data) were not precisely synchronized with respect to one another. The shift (which we estimate to be up to 100 ms) seemed to be non-constant even within a single trial. The authors did not respond to our request for clarification. As such, we decided that it was not possible to reasonably compare the algorithms on this data set.

APPENDIX C
PROPOSED ALGORITHMS

Four algorithms are proposed in this paper; one based only on the anterior-posterior component of the marker position (AP algorithm), one only on the vertical component of the marker position (vertical algorithm) and one fusing both components (fused algorithm). Finally, one causal algorithm will be proposed (RT algorithm).

A. Pos-AP algorithm

- Filter 5Hz second-order Butterworth filter.
- Differentiate once, giving the velocity in the anterior-posterior direction.
- Any maximum in the velocity is considered a heel strike.
- Differentiate once more, giving the acceleration in the anterior-posterior direction.
- Any minimum in the acceleration is considered a toe-off.

B. Pos-Vert algorithm

- Filter 5Hz second-order Butterworth filter.
- Any minimum in the vertical position is considered a heel strike.
- Differentiate twice, giving the acceleration in the vertical direction.
- For every heel strike, the first minimum in the vertical acceleration following it is considered a toe-off.

C. Pos-Fused algorithm

- Run the AP-algorithm.
- For each HS that is found by the AP-algorithm, if a HS is also found within 120 ms by the vertical algorithm, consider it a Heel Strike.
- For each HS, the first minimum in the anterior-posterior acceleration following it is considered a toe-off.

D. Pos-RT algorithm

- Take the sum over 5 previous samples of the anterior-posterior position (box window).
- Any maximum peak with prominence 5 is taken as a potential heel strike.
- Differentiate the anterior-posterior position twice, resulting in the acceleration.
- Take the sum over 20 previous samples of the anterior-posterior acceleration (box window).

- Any potential peak were the rolling mean of the acceleration is larger than 5 is taken as a heel-strike.
- For each HS, the moment the rolling mean of the acceleration crosses from positive to negative is taken as a TO.

E. Left/Right detection

The detection method is the same for all proposed algorithms, and consists of the following steps. For the velocity-based flavour:

- Differentiate the medio-lateral position, resulting in the medio-lateral velocity.
- For each HS, integrate the medio-lateral velocity over the preceding X ms.
- If the result of this integration is negative, the HS is considered to be part of a right step. If the result is positive, the HS is considered to be a left step.
- Differentiate the medio-lateral position twice, resulting in the medio-lateral acceleration. For the acceleration-based flavour:
- For each HS, integrate the medio-lateral acceleration over the preceding X ms.
- If the result of this integration is positive, the HS is considered to be part of a right step. If the result is negative, the HS is considered to be a left step.

APPENDIX D
ALGORITHM DESCRIPTIONS

A. Moe-Nilssen (removing gravity component)

The algorithm by Moe-Nilssen [18] is a well-known algorithm to remove the static component from an IMUs signal. The resultant acceleration is the true acceleration, so without the gravity component. The algorithm does this by assuming the gravity constant is known and the IMU is mainly in the upright position, assumptions that generally hold when the IMU is attached to a walking human. The algorithm was used by all subsequent algorithms, with the exception of Shin.

B. Zijlstra

[3]

- Filter the Anterior-Posterior acceleration using a 2 Hz Low-Pass filter (4th order Butterworth).
- Each zero-crossing from positive to negative is a Heel Strike.
- The actual moment of Heel Strike is taken as the peak immediately preceding the zero-crossing.

C. Gonzalez

[4]. This method

- Filter the Anterior-Posterior acceleration using a 2 Hz Low-Pass filter (11th order FIR).
- Each zero-crossing from positive to negative is a possible Heel Strike.
- Numerically approximate the positive area preceding the zero-crossing. In case it is smaller than the defined threshold, discard the possible event.
- The same area used to calculate the threshold (so between the two zero-crossings), is considered the search space for the Heel Strike. A peak is considered the moment of Heel Strike if the following conditions are met:
  - Vertical acceleration is higher than gravity.
  - The peak occurs before the vertical acceleration reaches 99% of its local maximum value.
  - If several peak satisfy these conditions the peak closest to the zero-crossing is selected.

D. Shin

[6]

- Take the norm of the total acceleration in all 3 directions:

$$a_{norm} = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

- Employ the sliding window summation over 0.2 seconds (at a sampling rate of 100 Hz, N would be 20):

$$SWS(k) = \sum_{t=k-N+1}^{k} a_{norm}(t)$$

- Employ the acceleration differential technique:

$$a(k) = SWS(k+N) - SWS(k)$$

- The zero-crossing from negative to positive of this signal is taken as the moment of Heel Strike.

E. McCamley

[5] Based on Continuous Wavelet Transform

- Vertical acceleration is integrated.
- The resulting signal is differentiated a Gaussian Continuous Wavelet Transform ().
- The minima of this signal were taken as the Heel Strike
- The signal is differentiated once more using a Gaussian Continuous Wavelet Transform ().
- The maxima of this signal are taken as the Toe-Off.

APPENDIX E
P-VALUES STATISTICAL ANALYSIS

	p-value		
Algorithm	Mean	Intra-participant	Inter-participant
McCamley	**0.00000**	0.26216	0.04248
Zijlstra	**0.00000**	0.12263	**0.00332**
Shin	**0.00000**	0.08186	0.99157
Gonzalez	0.13730	0.02083	0.60892

TABLE II

P-VALUES OF THE COMPARISONS BETWEEN LEFT AND RIGHT STEPS FOR ACCELERATION-BASED ALGORITHMS, COMPARING MEAN TIMING ERROR (WELCH'S T-TEST), INTRA-PARTICIPANT VARIATION (WELCH'S T-TEST) AND INTER-PARTICIPANT VARIATION (BARTLETT TEST FOR VARIANCE). P-VALUES BELOW 0.0167 (0.05/3) ARE CONSIDERED SIGNIFICANT AND ARE PRINTED IN BOLD.

Algorithm	McCamley	Zijlstra	Shin	Gonzalez	Pos-AP	Pos-Vert	Pos-Fused	Pos-RT
McCamley	*Mean*	**0.00000**	**0.00000**	**0.00000**	0.03643	0.61240	**0.01010**	0.02570
	Intra	0.65000	0.21047	**0.00002**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
	Inter	0.53716	0.68419	**0.00328**	**0.00013**	0.80459	0.19210	**0.00001**
Zijlstra			**0.00000**	**0.00000**	**0.00001**	**0.00000**	**0.00000**	**0.00000**
			0.40937	**0.00005**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
			0.83291	**0.00048**	**0.00103**	0.38814	0.48788	**0.00008**
Shin				**0.00000**	**0.00000**	**0.00000**	**0.00008**	**0.00000**
				0.00047	**0.00000**	**0.00000**	**0.00000**	**0.00000**
				0.00095	**0.00052**	0.51349	0.36646	**0.00004**
Gonzalez					**0.00000**	**0.00000**	**0.00000**	**0.00000**
					0.00000	**0.00000**	**0.00000**	**0.00000**
					0.00000	**0.00665**	**0.00004**	**0.00000**
Pos-AP						**0.01026**	**0.00000**	**0.00000**
						0.00000	**0.00005**	0.02290
						0.00006	**0.00796**	0.45322
Pos-Vert							0.04870	0.12519
							0.00025	**0.00000**
							0.12208	**0.00000**
Pos-Fused								0.30212
								0.01305
								0.00086

TABLE III

P-VALUES OF THE PAIR-WISE COMPARISONS BETWEEN ALGORITHMS FOR HS, COMPARING MEAN TIMING ERROR (WELCH'S T-TEST), INTRA-PARTICIPANT VARIATION (WELCH'S T-TEST) AND INTER-PARTICIPANT VARIATION (BARTLETT TEST FOR VARIANCE).

Algorithm	McCamley	Pos-AP	Pos-Vert	Pos-Fused	Pos-RT
McCamley	*Mean*	**0.00000**	**0.00000**	**0.00000**	**0.00001**
	Intra	**0.00000**	0.65799	**0.00000**	**0.00000**
	Inter	**0.01003**	0.10051	**0.00799**	**0.00001**
Pos-AP			0.03358	0.81537	**0.00000**
			0.00000	0.89851	0.12736
			0.00005	0.93284	0.04313
Pos-Vert				0.02541	**0.00000**
				0.00000	**0.00000**
				0.00004	**0.00000**
Pos-Fused					**0.00000**
					0.08667
					0.05215

TABLE IV

P-VALUES OF THE PAIR-WISE COMPARISONS BETWEEN ALGORITHMS FOR TO, COMPARING MEAN TIMING ERROR (WELCH'S T-TEST), INTRA-PARTICIPANT VARIATION (WELCH'S T-TEST) AND INTER-PARTICIPANT VARIATION (BARTLETT TEST FOR VARIANCE). P-VALUES BELOW 0.0167 (0.05/3) ARE CONSIDERED SIGNIFICANT AND ARE PRINTED IN BOLD.

	p-value		
Algorithm	Mean	Intra-participant	Inter-participant
McCamley	0.29986	**0.00000**	0.32283
Zijlstra	0.11368	**0.00788**	0.52989
Shin	0.12309	**0.00000**	0.69552
Gonzalez	0.87261	**0.00040**	0.96757

TABLE V

P-VALUES OF THE COMPARISONS BETWEEN THE VIRTUAL COM MARKER AND THE RASI MARKER FOR POSITION-BASED ALGORITHMS, COMPARING MEAN TIMING ERROR (WELCH'S T-TEST), INTRA-PARTICIPANT VARIATION (WELCH'S T-TEST) AND INTER-PARTICIPANT VARIATION (BARTLETT TEST FOR VARIANCE). P-VALUES BELOW 0.0167 ARE CONSIDERED SIGNIFICANT.