Body-worn sensors, so-called wearables, are getting more and more popular in the sports domain. Wearables offer real-time feedback to athletes on technique and performance, while researchers can generate insights into the biomechanics and sports physiology of the athletes in real-world sports environments outside of laboratories. One of the first sports disciplines, where many athletes have been using wearable devices, is endurance running. With the rising popularity of smartphones, smartwatches and inertial measurement units (IMUs), many runners started to track their performance and keep a digital training diary. Due to the high number of runners worldwide, which transferred their data of wearables to online fitness platforms, large databases were created, which enable Big Data analysis of running data. This kind of analysis offers the potential to conduct longitudinal sports science studies on a larger number of participants than ever before.

In this dissertation, both studies showing how to extract endurance running-related parameters from raw data of foot-mounted IMUs as well as a Big Data study with running data from a fitness platform are presented.
Markus Zrenner

From Raw to Big Data in Endurance Running
Application of Data Science Techniques for Knowledge Creation from Wearable Sensor Data
Markus Zrenner

From Raw to Big Data in Endurance Running

Application of Data Science Techniques for Knowledge Creation from Wearable Sensor Data

Erlangen
FAU University Press
2022
From Raw to Big Data
in Endurance Running
Application of Data Science Techniques for Knowledge Creation from Wearable Sensor Data

Der Weg von Rohdaten zu Big Data beim Ausdauerlauf
Anwendung von Data Science Methoden für den Erkenntnisgewinn aus Daten von Wearables

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur
Erlangung des Doktorgrades Dr.-Ing.

vorgelegt von
Markus Zrenner, M.Sc.
aus Untersiemau
Als Dissertation genehmigt
von der Technischen Fakultät
der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen
Prüfung: 22.02.2022

Vorsitzender des
Promotionsorgans: Prof. Dr. Knut Graichen

Gutachter: Prof. Dr. Bjoern Eskofier
Prof. Dr. Arnold Baca
Abstract

Body-worn sensors, so-called wearables, are getting more and more popular in the sports domain. Wearables offer real-time feedback to athletes on technique and performance, while researchers can generate insights into the biomechanics and sports physiology of the athletes in real-world sports environments outside of laboratories. One of the first sports disciplines, where many athletes have been using wearable devices, is endurance running. With the rising popularity of smartphones and smartwatches, many runners started to track their performance and keep a digital training diary. Due to the high number of runners worldwide, which transferred their data of wearables to online fitness platforms, large databases were created, which enable Big Data analysis of running data. This kind of analysis offers the potential to conduct longitudinal sports science studies on a larger number of participants than ever before.

In this work, I contributed to this research field in two ways. On the one hand, new wearable devices can extend the parameter set of current fitness platforms to provide a holistic picture of both the biomechanics and the physiology of endurance runners. While the physiology of a runner can already be described with the data of Global Positioning System (GPS) and heart rate sensors, foot-mounted Inertial Measurement Units (IMUs) can be used to determine biomechanical parameters. In this work, I evaluated which algorithm is best suited to compute spatio-temporal and kinematic parameters from the IMU raw data in running as well as where to place the sensor on a running shoe to get the most accurate metrics.

In the first analysis, I identified and implemented four different algorithms which can be used to determine stride length and average stride velocity. To evaluate the algorithms, I acquired two different data sets. The results showed that a digital signal processing approach, which reconstructs the foot trajectory during one stride using a zero-velocity update, outperformed the other algorithms. The mean error of stride length was $0.022 \pm 0.157$ m and the mean error of average stride velocity was $0.032 \pm 0.274$ m/s. This high accuracy shows that foot-mounted IMUs can be used to monitor and evaluate the movement of runners outside of laboratories.
Additionally, I investigated the IMU sensor position on the foot. During literature research, I identified four positions on or within a running shoe: on the heel cap, on the lateral side below the ankle, on the instep above the laces, and within a cavity in the sole of the running shoe below the arch of the foot. To evaluate the influence of the sensor position, I compared both the raw data of individual sensors during a stride as well as the accuracy of spatio-temporal and kinematic parameters. The results showed that both the raw data as well as the accuracy of the parameters differed between the sensor positions. The reconstruction of the trajectory worked best with the data of the sensor in the cavity because it was exposed to the least motion during Midstance (MS) and hence yielded the smallest error in the zero-velocity update.

On the other hand, I contributed to the research field of Big Data from wearables by showing the potential of data from fitness platforms for sports science studies. In a Big Data study, I retrospectively investigated the response to training of recreational marathon athletes using wearable data. The training characteristics of 6771 marathon athletes were analyzed, who tracked both the training as well as a marathon race on the adidas Runtastic fitness platform using GPS and heart rate sensors. The response to training was quantified by the improvement of the athletes’ best 10 km performance throughout the training process. The results showed that intensity polarization had a positive effect on the training response. However, no training pattern which works well for all athletes could be identified because the response to training is highly individual.

My work in this thesis exemplifies that Big Data from wearable sensors has a large potential to bring new insights to the sports science community. This is not limited to insights into exercise physiology, which is already possible using the data of GPS and heart rate sensors, but also into biomechanics. Foot-mounted IMUs can accurately determine spatio-temporal and kinematic parameters, which enables the monitoring of biomechanics in-field at a large scale. However, the results of this thesis also show that in Big Data analysis the source of the parameters and their accuracy have to be considered because manufacturers might use different algorithms or data of different sensor positions to compute the metrics.
Kurzzusammenfassung


In der Analyse der verschiedenen Algorithmen wurden vier Möglichkeiten für die Berechnung der spatio-temporalen Parameter Schrittlänge und Schrittgeschwindigkeit in der Literatur identifiziert und implementiert, sowie mit zwei im Rahmen dieser Arbeit erhobenen Datensätzen evaluiert. Es zeigte sich, dass ein Algorithmus, der die Trajektorie des Fußes während eines Schrittes mit einem Zero-Velocity Update rekonstruiert, die genauesten Ergebnisse lieferte. Der mittlere Fehler der Schrittgeschwindigkeit betrug bei diesem Algorithmus $0.032 \pm 0.274 \text{ m/s}$ und der mittlere Fehler der Schrittlänge $0.022 \pm 0.157 \text{ m}$. Die hohe Genauigkeit der Algorithmen
Abstract

zeigt, dass man mit am Fuß befestigten IMUs biomechanische Parameter außerhalb des Labors bestimmen und bewerten kann.


Meine Beiträge in dieser Arbeit zeigen, dass Big Data von am Körper tragbaren Sensoren Sportwissenschaftlern in Zukunft neue Erkenntnisse bringen können. Diese neuen Erkenntnisse werden nicht auf die Physiologie von Läufern limitiert sein, die schon heute mit GPS- und Herzratensensoren im Feld untersucht werden kann. Zusätzlich können auch biomechanische Parameter von Läufern ausgewertet werden, die sich akkurat mit am Fuß tragbaren IMUs bestimmen lassen. In diesen Analysen ist allerdings auf die Sensordatenquelle zu achten, da unterschiedliche Sensorhersteller verschiedene Algorithmen mit Daten von unterschiedlichen Sensorpositionen nutzen, was einen Einfluss auf die Genauigkeit der abgeleiteten Parameter hat.
Acknowledgements

At first, I would like to thank the company and the funding agencies which allowed me to have five wonderful years in research by providing financial support. I received initial funding from the adidas AG, which kick-started my PhD. Afterward, I was supported by the Center Digitisation.Bavaria (ZD.B) within the InnovationLab framework. Lastly, I had the opportunity to work in the Connected Movement project funded by the Association for German Engineers (VDI).

Even though the InnovationLab did not directly influence my scientific work, I am nevertheless very grateful to have had the chance to bring the InnovationLab for Wearable and Ubiquitous Computing to life. However, this would have not been possible without the support of many others. Thanks to the ZD.B, the Innolab coaches from the other universities in Bavaria, the companies which provided projects, all the external lecturers, and the FAU Digital Tech Academy for helping us to offer a great course for our students. Also thank you to all my MaD Lab colleagues, who were Scrum coaches for the student teams and had a great influence on the fantastic prototypes by providing their expertise. In this regard, the biggest thank yous go out to Philipp and Matthias. Philipp, thanks for supporting me throughout the full Innolab time. It was loads of fun organizing this course with you, even though we had our little problems with finding the way to our project partners or booking wrong hotels in Würzburg. Matthias, thanks for taking over and bringing the Innolab to the next level.

Also, thank you to all my colleagues from adidas, who worked with me in the Servicefactory and the Connected Movement project. It was a pleasure to get a glimpse into the way of working in the Innovation & Technology, the Future Sport Science, and the Advanced Concepts team as well as into all their fantastic and innovative projects. From the adidas people, the impact of three persons on my time at adidas and this thesis stands out: Thank you to Ulf, Burkhard, and Christian for the fantastic projects that you provided me with, for welcoming me into your teams and for giving valuable input to my scientific work.

The next thank you goes out to all my colleagues at the MaD Lab. It was really not easy leaving the Lab, because I enjoyed (almost) every single day I spent there. This was mostly due to all the fantastic colleagues I was
lucky enough to work with - colleagues that most certainly became friends. A special shout-out goes to my former office members: Martin, Nils, and Robert - thanks for making even the stressful days fun, the serious and definitely not serious discussions, and all the other fun memories.

Besides, I would like to thank Björn Eskofier. Björn, thank you for giving me the chance to work with all those great people, but also for your guidance and support throughout my time at the Lab. I always saw you rather as a mentor than a boss, which tells a lot about the great way you lead people.

And last, but definitely not least - thank you to my family and friends who supported me not only in my PhD time but also before. It was always good to have people around me who help to forget work in the evening or on the weekend. Thank you Mum and Dad for all your support from the start, for giving me advice when needed but also to let me find my own way. A way, which led me to the last person I would like to thank - my wife Miri. Thanks for always being there for me. Without you the last five years would definitely not have been as great as they were - I am super happy to have you!
Contents

Abstract ................................................................. iii
Acknowledgements .................................................. vii
List of Abbreviations ............................................. xiii
List of Figures ....................................................... xv
List of Tables ....................................................... xxiii

I Introduction ....................................................... 1
1 Motivation ......................................................... 3
  1.1 The rise of Wearables and Big Data in sports ............... 3
  1.2 Endurance running - the early adopter of wearables ...... 5
  1.3 Objectives of the thesis ................................... 6
  1.4 Contributions ............................................... 7
  1.5 Overview of the thesis ................................... 10

II Fundamentals & state of the art .............................. 11
2 Fundamentals in sports science ............................... 13
  2.1 Spatio-temporal parameters and foot kinematics in endurance running ...................................... 13
    2.1.1 Stride phases and stride events ...................... 14
    2.1.2 Stride length, stride time and velocity .......... 15
    2.1.3 Foot joint angles .................................... 16
    2.1.4 Laboratory measurements .......................... 20
  2.2 Physiology of endurance running ......................... 21
    2.2.1 Basics of exercise physiology ..................... 21
    2.2.2 Laboratory measurements .......................... 24

3 State of the art .................................................. 27
  3.1 Wearables in endurance running .......................... 27
    3.1.1 Sensor systems ...................................... 27
    3.1.2 Applications of IMUs in endurance running ...... 29
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>Adenosine Diphosphate</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>AWS</td>
<td>Amazon Web Services</td>
</tr>
<tr>
<td>BSN</td>
<td>Body Sensor Network</td>
</tr>
<tr>
<td>DCNN</td>
<td>Deep Convolutional Neural Network</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>EKF</td>
<td>Extended Kalman Filter</td>
</tr>
<tr>
<td>FAU</td>
<td>Friedrich-Alexander-Universität Erlangen-Nürnberg</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HIT</td>
<td>High Intensity</td>
</tr>
<tr>
<td>HMM</td>
<td>Hidden Markov Model</td>
</tr>
<tr>
<td>IC</td>
<td>Initial Ground Contact</td>
</tr>
<tr>
<td>IMU</td>
<td>Inertial Measurement Unit</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile Range</td>
</tr>
<tr>
<td>LED</td>
<td>Light Emitting Diode</td>
</tr>
<tr>
<td>LIT</td>
<td>Low Intensity</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long Short-Term Memory</td>
</tr>
<tr>
<td>MEMS</td>
<td>Micro-electricalmechanical Systems</td>
</tr>
<tr>
<td>MS</td>
<td>Midstance</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PPG</td>
<td>Photoplethysmography</td>
</tr>
<tr>
<td>RNN</td>
<td>Recurrent Neuronal Network</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>TO</td>
<td>Toe Off</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
List of Figures

1. Research interest in the topics 'Wearable' and 'Big Data'. This plot was created using the Web of Science search engine. For 'Wearable' and 'Big Data' the number of results from a topic search were counted per year.

2. Visualization of the running gait cycle.

3. Visualization of decreasing ground contact time $t_{gc}$ with increasing stride velocity $v_{stride}$.

4. Visualization of (a) stride length $d_{stride}$ and (b) stride frequency $f_{stride}$ over four velocity bins from 2-6 m/s. The intervals include the left bin edge (2-3 means $2 \leq v < 3$). The boxes represent the Interquartile Range (IQR) and the whiskers extend up to 1.5⋅IQR. The plots indicate that velocity can be controlled either by increasing $d_{stride}$ or $f_{stride}$.

5. Visualization of the talocrural joint between the tibia and the talus and the subtalar joint between the talus and the calcaneus. Adapted from [38].

6. Visualization of rear-, mid- and forefoot strike type.

7. Visualization of pronation, neutral running style and supination. $\gamma$ indicates the rearfoot eversion angle, $\beta$ the achilles tendon angle, which are common surrogates to quantify the degree of pronation in runners.

8. Visualization of contribution of aerobic, lactic anaerobic and alactic anaerobic energy production to overall energy supply over the duration of exercise with maximum intensity. Adapted from [65].

9. Visualization of the aerobic system. Adapted from [68].

10. Visualization of the lactate concentration in blood with increasing workout intensity. Adapted from [70].

11. Visualization of (a) principle of GPS sensors and (b) a GPS-track on a map.

12. Overview of four steps which have to be considered when implementing algorithms capable of extracting spatio-temporal and kinematic parameters using foot-mounted IMUs.
List of Figures

13 Visualization of the effects handled by the sensor calibration. The real-world acceleration $\mathbf{a}$ is rotated by a rotation matrix $\mathbf{K}_a$, scaled by a scaling matrix $\mathbf{S}_a$ and translated by a bias vector $\mathbf{b}_a$ due to temperature and manufacturing inaccuracies, such that the IMU sensor measures the acceleration $\mathbf{a}_m$. .............................................. 37

14 Illustration of the four sensor-to-segment or sensor-to-shoe alignment techniques from literature. The blue squares illustrate the IMU attached to the shoe with its coordinate system in black. The orange coordinate system is the shoe coordinate system. The green arrows indicate the axis of the shoe coordinate system, which is aligned during the calibration process. The red square in the anatomical technique illustrates the IMU on the external devices which is used to determine the segment axis. .... 39

15 Acceleration and gyroscope signal of foot-mounted IMU (Cavity). The vertical lines represent the Initial Ground Contact (IC), MS and Toe Off (TO) events, which were set manually. The $x$-axis points in medial/lateral direction, the $y$-axis in anterior/posterior direction and the $z$-axis in superior/inferior direction. .............................................. 41

16 Visualization of difference between classification and regression task in machine learning. (a) For the two-class classification problem, $y$ can only be either 0 or 1. A decision boundary can be trained to differentiate the two classes. (b) For the regression, problem, $y$ is continuous. The plot depicts two polynomial functions of degree $p = 1$ and degree $p = 2$, which were fit to the data. ................. 46

17 Process graph for scientific Big Data projects. Adapted from [30]. .............................................. 49

18 Overview of commercial fitness platforms athletes can use for tracking their runs. .............................................. 52

19 (a) Shoe equipped with a miPod sensor and the marker setup. The IMU is located within the sole of the running shoe. The marker setup allowed for a computation of velocity and stride length. (b) Illustration of reference system setup. The subjects ran through the capture volume of the motion tracking system, created by 16 infrared cameras, and looped back around. ......... 64
Example for the stride segmentation. The plot shows the acceleration signal in the dorsoventral direction $a_y[n]$, the detected initial ground contact $n_{IC}$, and the beginning of the swing phase (zero crossing $n_{ZC}$) to confirm the stride candidate. The marked area depicts the integration area for the swing phase detection. ........................................ 66

Polynomial function of second degree (red line) that relates the velocity integration value $t$ to the reference velocity values $v_{\text{stride}}$ (grey dots). ......................... 70

The four steps of the algorithm for the trajectory reconstruction based on linear dedrifting. ......................... 70

Visualization of the dedrifting method that ensures that the velocity during the second midstance is zero. (a) Velocity signal before dedrifting. The grey dotted linear function is fit between the first and last point of the stride (midstance). (b) Velocity signal after dedrifting. ........... 72

Architecture of the convolutional neural network for stride length regression based on the raw 6D-IMU signal. For the first convolutional layer, we used $N_1 = 32$ filter kernels of kernel length $K_1 = 30$. The second convolutional layer consisted of $N_2 = 16$ filter kernels of kernel length $K_2 = 15$. The first fully-connected layer had $M_1 = 128$ outputs that served as input to the second fully-connected layer, which had only a $M_2 = 1$ output. This output represented the computed stride length. ......................... 73

Mean error of stride length of the four different algorithms for the different velocity ranges the subjects ran in the lab study. ................................................................. 77

Bland–Altman plots for stride length and velocity for the four algorithms. Each row contains the metrics for one algorithm. The individual samples are color coded depending on the velocity bin of the sample: 2–3 m/s blue, 3–4 m/s red, 4–5 m/s green, 5–6 m/s purple. The dotted-dashed horizontal lines depict the mean error and the dotted horizontal line the 95% confidence interval. . . . 79

Mean absolute error of the 3.2-km run for the four different algorithms and GPS. ................................. 80
28 Visualization of the correlation between the stride time $t_{\text{stride}}$ and the relative stride length $d_{\text{stride,rel}}$ for male subjects. The light gray dots depict the data obtained from the field study, whereas the red curve and the black dashed lines visualize the step function obtained from literature and implemented in the Stride time algorithm.

29 Violin plots of the error ($v_{\text{ref}} - v_{\text{stride}}$) in the velocity computation for forefoot strikers and rearfoot strikers. The vertical lines show the mean errors for the two strike type groups.

30 Visualization of the running gait cycle.

31 Visualization of sole angle and range of motion.

32 Visualization of sensor positions on the running shoes, the global coordinate system $(x_g, y_g, z_g)$, the shoe coordinate system $(x_s, y_s, z_s)$, and the individual sensor coordinate systems. When the foot is flat on the ground, the global and the shoe coordinate system are aligned.

33 Visualization of the functional calibration procedure. The first part of the functional calibration consisted of standing still with the foot flat on the ground in order to measure gravity. During the second part the subjects rotated their feet on a balance board to compute the medial/lateral axis using a principal component analysis.

34 Exemplary IMU data of one stride segmented from IC to IC for the four different sensor positions.

35 Visualization of the stride segmentation for the cavity sensor using the gyroscope signal in the sagittal plane $\omega_x[t]$. The fiducial points at swing phase $n_{SP}$ are local minima of the angular rate in the sagittal plane. The index $n_{IC}$ indicates the index of IC, which corresponds to the bias corrected local maximum after $n_{SP}$. The MS event $n_{MS}$ is at the minimum of the gyroscopic energy. The TO event at $n_{TO}$ is based on the second local maxima and a bias correction.

36 Acceleration $\ddot{a}[t]$ and angular rate $\dot{\omega}[t]$ raw data of the heel sensor segmented from IC to TO of the consecutive stride.
37 Visualization of the gravity removal in the acceleration signal for a sample stride of the heel sensor. The upper plot shows the raw acceleration $\vec{a}[t]$ segmented from MS to MS measured by the accelerometer. The lower plot shows the gravity corrected acceleration signal $\vec{a}_{gc}[t]$ after rotating the raw acceleration by the quaternion sequence $q[n]$ and removing gravity from the rotated signal. After the gravity removal, both the z-components of the acceleration at the first midstance ($t=0$ s) and the second midstance ($t=0.81$ s) have values close to zero.

38 Visualization of the dedrifting of the velocity after the first integration of the acceleration signal for a sample stride of the heel sensor. The upper plot shows the velocity $\vec{v}[t]$ before dedrifting. This signal displays that the velocity at the second midstance ($t=0.81$) is not zero. We enforce the velocity to be zero by dedrifing the velocity using a linear dedrifting function. The lower plot shows the velocity $\vec{v}_{dedrfted}[t]$ after dedrifting. Now, the velocity at the second MS is zero in all directions.

39 Visualization of the trajectory for a sample stride of the heel sensor. The upper plot shows the orientation $\vec{a}[t]$ obtained by the quaternion based forward integration after converting the quaternions back to their angle representation. The lower plot shows the translation $\vec{d}[t]$ obtained by dedrifted double integration of the gravity corrected acceleration $\vec{a}_{gc}[t]$.

40 Visualization of angle computation for a sample stride from the cavity sensor. The angles are depicted from $n_{IC}$ ($t = 0$ s) to $n_{TO}$ ($t = 0.32$ s). The sole angle is defined as the rotation in the sagittal plane between IC and MS. As the orientation is initialized with zero at MS, the sole angle is the angle at $n_{IC}$. The range of motion is defined as the difference between the maximum and minimum (red dots) of the angle in the frontal plane during ground contact.
Results of the evaluation of the Pearson’s correlation coefficients between the IMU raw signals. Each box visualizes the correlation coefficients between two sensors for all the strides in $x$, $y$, and $z$ direction. The box plots also display the median of the correlations (median line), the IQR (box), and the 5 and 95 percentiles (whiskers). The upper plot depicts the correlation of the full strides, the middle plot the correlations during ground contact phase, and the lower plot the correlations during swing phase. 

Visualization of the error for (a) stride length and (b) the acceleration at the zero-velocity update for the four different sensor positions in different speed ranges.

Exemplary visualization of (a) duration distribution curve $P_{tb_1}[X = \tilde{V}]$ and (b) cumulative duration distribution curve $F_{tb_1}[X = \tilde{V}]$ for the normalized velocity $\tilde{V}$ for training block 1. The red lines indicate the barriers for the intensity zones defined based on the marathon performance velocity $v_{mp}$. 

Visualization of the training response $\Delta v_{10}$ across ten velocity groups. Each dot represents one subject. The response categories are color coded. The vertical black lines are located at the percentile values of $v_{10,tb_1}$. The horizontal grey line indicates the zero line, where subjects showed neither a positive nor negative improvement. Due to the statistical approach in the response group definition, which assured equally sized groups, the low and moderate response group also included subjects with negative $\Delta v_{10}$.

Validation of data set. (a) Distribution of marathon performance time $T_{mp}$. (b) Distribution of maximum training heart rate $hr_{max}$. (c) Visualization of correlation between best 10 km velocity $v_{10}$ in the complete training period and marathon performance velocity $v_{mp}$. The blue dots indicate the individual subjects, the green line the linear regression function. (d) Adaptive potential for improvement of best 10 km velocity $\Delta v_{10}$. 

xx
Visualization of the difference between $tb_4$ and $tb_1$ of (a) the normalized mean velocity and (b) the normalized mean heart rate for all subjects in the 3 response groups. The velocity values were normalized by the marathon performance velocity $v_{mp}$ and the heart rate by the maximum training heart rate $hr_{max}$.

Visualization of the share of time spent in the intensity zones for (a) slow marathon performances, (b) medium marathon performances and (c) the fast marathon performances. For each marathon performance group, we provide three plots showing the share of time spent in the three zones for the low response, moderate response and high response group. The individual boxes in each plot visualize the IQR within the training blocks. The black horizontal lines within the boxes indicate the median. The whiskers extend to $1.5\cdot$IQR. We computed repeated measure Analysis of Variances (ANOVAs) for each response and marathon performance category for each intensity zone. The asterisks indicate the effect size of the results: * for $0.01 \leq \eta^2 < 0.05$, ** for $0.05 \leq \eta^2 < 0.12$, *** for $\eta^2 \geq 0.12$.

Visualization of (a) workout duration $T$, (b) workout distance $D$ and (c) number of workouts $W$ over the four training blocks.
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Definition of strike type [46]. Plantarflexion is indicated by a negative angle, dorsiflexion by a positive angle.</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>Overview of number of IMU sensors, sensor positions and type of endurance running related parameters based on IMU data with references from literature.</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>Data sets from fitness platforms used in science. Different subsets of the full Strava data set with 31190 runners were used in [145, 146, 147, 148].</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>Anthropometric data of subjects participating in the lab study.</td>
<td>61</td>
</tr>
<tr>
<td>5</td>
<td>Number of trials and recorded strides per velocity range in the lab study.</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>Definition of step functions for the Stride time algorithm for the relative stride length $d_{\text{stride,rel}}[t_{\text{stride}}]$ for (a) male and (b) female runners.</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>Mean error (ME) and standard deviations (Std), mean percentage error (MAPE), and mean absolute error (MAE) of average velocity per stride $v_{\text{stride}}$ and stride length $d_{\text{stride}}$ of the four algorithms for the lab study dataset.</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>Results of other publications related to stride length and velocity calculation.</td>
<td>81</td>
</tr>
<tr>
<td>9</td>
<td>Differences in the study setup and architecture presented by Hannink et al. [140] from our DCNN implementation. $N_2$ is the number of filters in the second layer, $M_1$ the number of output nodes in the first fully connected layer.</td>
<td>86</td>
</tr>
<tr>
<td>10</td>
<td>Naming of the sensor position and details on the mounting of the sensors.</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>Number of trials and recorded strides per velocity range. During the data acquisition, we controlled for speed and the subjects only changed the velocity range, if the required number of trials in the previous (slower) velocity range was reached.</td>
<td>98</td>
</tr>
</tbody>
</table>
List of Tables

12 Median error (Med.) and IQR of error for the parameters stride time $t_{\text{stride}}$, ground contact time $t_{\text{gc}}$, stride length $d_{\text{stride}}$, average stride velocity $v_{\text{stride}}$, sole angle $\alpha_{\text{sole}}$, and range of motion ROM compared to the motion capture system. .......................................................... 113

13 Median error (Med.) and IQR of the error for the parameters sole angle $\alpha_{\text{sole}}$ without the bias correction for IC. ... 118

14 Mean and standard deviation of the training parameters for the 16 week training process. The statistical values are reported for the different response groups and the different marathon time categories. $T$ indicates the total training duration, $D$ the total distance, $\bar{Hr}$ the relative mean heart rate $v$ the relative mean velocity, $v_{10}$ the best 10 km velocity, Low Intensity (LIT) the share of time in the LIT zone, threshold the share of time in the threshold zone, High Intensity (HIT) the share of time in the HIT zone, $W$ the number of workouts, $I_{T90}$ the number of workouts longer than 90 minutes and $I_{D15}$ the number of workouts with a higher than 15 km of the 16 week training process. For all parameters except $W$, $I_{T90}$ and $I_{D15}$, we report the effect sizes $\eta^2_{\text{resp}}$ (response groups) and $\eta^2_{\text{mp}}$ (marathon performance groups) of the main effect of the two-way ANOVA if the $p$-value was below a significance level $\alpha = 0.05$. The last row of the table lists the number of subjects in each group. .......................................................... 137
Part I

Introduction
1  Motivation

1.1  The rise of Wearables and Big Data in sports

The impact of wearable technology is constantly growing. Wearable devices are body-worn sensors, which are capable of monitoring human health and performance by continuously measuring biosignals, analytes, and biomechanical parameters [1]. Due to the huge potential of such devices, the research interest in wearables has been constantly growing in the last decade (Figure 1). Within the research field, there are different branches, which all contributed to the progress of wearable technology in recent years. One of those branches deals with the hardware of wearable devices. It ranges from the development of new sensor types (e.g. sensors integrated in garments) over new antenna designs to the constantly progressing miniaturization of sensors. Another branch deals with the development and evaluation of various applications. Researchers have developed algorithms to extract meaningful measures from the raw data in different areas like sport [2, 3, 4, 5], health [6, 7, 8, 9], safety of work [10, 11], and lifestyle [12, 13]. And last but not least, more and more researchers also focus on the usability of and interaction with wearable devices to create entertaining and effective user experiences [3, 14, 15].

![Figure 1: Research interest in the topics 'Wearable' and 'Big Data'. This plot was created using the Web of Science search engine. For 'Wearable' and 'Big Data' the number of results from a topic search were counted per year.](image)

1 on 2021/03/06
One of the early adopters of wearable technology was the sports domain. In individual as well as in team sports, many athletes track their progress by gathering data with wearable devices. Instead of writing training diaries by hand, the performance data is stored digitally. This not only simplifies the analysis of performance over time, but also the data sharing with coaches who support the athletes from remote locations [16]. Apart from the retrospective analysis of sports data, wearables are capable of providing real-time feedback to athletes about technique and performance [17, 18]. Especially for endurance athletes, who follow prescribed training plans, the direct feedback makes it easier to meet the requirements for a defined workout and thus helps to improve athletes’ response to training [19]. The data gathered from wearable devices also have the potential to prevent and detect imbalances or injuries during sports. By finding abnormal movement patterns and unexplainable drops in performance both on the fly and retrospectively, potential injuries can be anticipated and prevented by adjusting the training of an athlete [20, 21].

However, data of wearables not only offers benefits to athletes but enables other interesting use cases for third parties as well. Traditional sports science studies have been conducted in laboratories under controlled conditions. Body-worn sensors enable new kinds of studies, where subjects train and perform sports in real-world environments [22]. Besides, TV stations can make broadcasting sports events more interesting by providing insights into the performance of athletes equipped with body-worn sensors in real-time [23]. Also, insurance companies have picked up on data of wearables and started to offer benefits on insurance contracts if subjects track themselves during sports and consequently live a healthy lifestyle [24].

The steady technological progress, the constantly improved usability, and especially the applications caused the huge success of wearable devices in recent years. The high number of sold wearable devices [25] was accompanied by a large amount of data, which was acquired by the consumers of wearable devices. As in many other fields, this large amount of data can not be handled with traditional data analysis tools anymore, which turned wearable data into Big Data. Similar to wearables, the research field dealing with Big Data is constantly growing (Figure 1). Big Data also has different research branches, which were summarized in the work from De Mauro et al. [26]. They reviewed Big Data literature and came up with the following definition for Big Data: “Big Data is the Information asset characterized by such a High Volume, Velocity
and Variety to require specific Technology and Analytical Methods for its transformation into Value.” Apart from special database structures for the storage of data (e.g. Hadoop distributed file system [27]) and tailored algorithms for processing data (e.g. MapReduce [28]), large computational resources are necessary to enable the main analysis goal: creating value from Big Data [29]. To accomplish this goal, also the scientific methodology in data analysis has to be adapted to prevent the community from false conclusions. While in many domains, including the sports domain, traditional studies were supervised, the majority of Big data is acquired in an unsupervised manner. Thus, the quality and the validity of the data, which scientific conclusions are based on, have to be assured in Big Data studies [30].

1.2 Endurance running - the early adopter of wearables

Within the different sport disciplines, athletes in endurance running were one of the first to use wearable devices. With the rise of smartphones and their applications, more and more runners started monitoring their performance. Especially Global Positioning System (GPS) sensors, which were integrated into smartphones, enabled endurance runners to track their performance. Based on the GPS data, the distance, pace, and altitude profile of workouts could be analyzed [31]. Additionally, more and more people started tracking their heart rate using either chest straps or Photoplethysmography (PPG) sensors [32, 33]. Both GPS and heart rate sensors are mainly used to analyze the workouts from a physiological perspective by monitoring the training load. Other commonly utilized sensors in endurance running are small and lightweight Inertial Measurement Units (IMUs). Those sensors are capable of measuring various parameters and thus enable an in-field evaluation of both biomechanics and the performance of endurance athletes. IMUs are usually attached to specific body segments and analyze the segments’ movement by measuring its acceleration and angular rate in 3D space. Depending on the mounting position, different running-related parameters can be extracted. Wrist-worn IMUs (e.g. in smartwatches) are capable of monitoring the cadence during running [34], whereas IMUs integrated into chest straps can also monitor vertical oscillation, ground contact time [35] and provide a balance measure between left and right foot [36]. Another popular position is the foot. In addition to the already mentioned IMU
parameters, further kinematic parameters describing the orientation of the foot during running can be extracted [37]. The promising results from literature also motivated companies like Garmin Ltd., Stryd, or Polar Electro Oy to develop and start selling foot pods based on IMUs commercially.

The data of wearable devices are often uploaded to and evaluated on fitness platforms like adidas Runtastic, Strava, Garmin Connect or Polar Flow. These platforms not only store the data of workouts in form of a workout diary but also offer services like the estimation of the fitness status, prediction of race times, or the generation of free training plans. Due to the large number of users on these platforms, a huge amount of running data is available which can be used for Big Data analysis. This data has the potential to drive many interesting applications and provide scientific insights into various aspects of endurance running.

1.3 Objectives of the thesis

The data of fitness platforms seem to have a large potential to create new insights into endurance running and to enable new services for athletes. However, the fitness platforms currently mainly contain data from GPS and heart rate sensors, which can be used for providing physiological feedback to the users. Adding biomechanical data in form of spatio-temporal and kinematic parameters can further improve the feedback on running technique as well as enable a more holistic view on the endurance athlete. However, the extracted parameters have to be accurate to be accepted by consumers, to provide valuable feedback to runners, and to be eligible to draw robust conclusions in scientific analysis. Thus, the underlying algorithms which compute the parameters have to be evaluated for accuracy.

Nevertheless, the current data available on fitness platforms can already be used to answer interesting research questions in exercise physiology. Thus, the objectives of this thesis are as follows:

1. Evaluation of the accuracy of spatio-temporal and kinematic parameters based on foot-mounted IMU data in order to extend the parameter set of fitness platforms and to increase trust in the measures.
2. Showing the scientific potential of Big Data from wearables, which are already stored on fitness platforms.

1.4 Contributions

For the first objective, I present two contributions, in which I evaluated the accuracy of running-related spatio-temporal and kinematic parameters based on IMU data. These publications can lead the way to the extension of existing fitness platforms with additional data. For the second objective, I contributed one of the first studies, which used Big Data from wearable devices to evaluate the training process of athletes for a marathon. All three publications extended the state of the art and were published in ranked international journals. Hereinafter, the three major contributions are summarized:

- In literature, different methodologies to compute velocity and stride length using foot-mounted IMUs have been introduced and evaluated. However, the algorithms were developed using different, self-recorded data sets, which made an accuracy benchmark difficult. In the journal article “Comparison of Different Algorithms for Calculating Velocity and Stride Length in Running Using Inertial Measurement Units” [P1], which was published in the special issue on Sensors for Gait, Posture, and Health Monitoring of the MDPI sensors journal, I implemented four different algorithms presented in literature and compared them on two different data sets. I showed that a reconstruction of the foot trajectory during running utilizing a zero-velocity update yielded the best results. Also, a deep convolutional neural network regression showed promising results but needs more data for the training phase.

- Similar to the type of algorithm, there was no agreement in literature where to place inertial measurement units on a running shoe in order to compute foot kinematic parameters accurately. I identified four different sensor positions where researchers mounted sensors on running shoes. In the journal article “Does the position of foot-mounted IMU sensors influence the accuracy of spatio-temporal parameters in endurance running?” [P2], which was published in an MDPI sensors special issue on Technologies for Sports Engineering and Analytics, I compared the four sensor positions concerning the accuracy of foot kinematic parameters using the trajectory
algorithm which has proven to be best suitable to compute velocity and stride length in running. I showed that the accuracy of the parameters differed between the sensor positions. In my analysis, the data of sensors located in a cavity of the midsole of a running shoe under the arch of the foot were best suited to determine spatio-temporal and kinematic parameters.

- While training processes of professional athletes for marathons have been under scientific investigation for decades, the research on the training processes of recreational athletes is scarce. In “Retrospective analysis of training and response in marathon finishers using data from wearable sensors” [P3], I evaluated the training process of 6771 marathon finishers who tracked their training using wearable devices on the adidas Runtastic fitness platform. In the analysis presented in Frontiers in Physiology, I showed how to clean a data set to assure high data quality to draw robust scientific conclusions, I introduced a data inherent measure to evaluate response to training on an unsupervised data set and provided insights into the training of recreational marathon runners.

Apart from the three major contributions above, I further contributed to the research field of data science in sports. The works, which influenced the major contributions methodologically and extended the state of the art in the respective fields, are listed below. Papers, where I contributed as first or shared first author, are emphasized in italics:

- In “Human authentication implemented for mobile applications based on ECG-data acquired from sensorized garments” [P4], I developed an algorithm to classify the person who wears a sensorized garment using a feature-based nearest neighbor classifier. For this work, I developed the algorithm, and partly wrote the manuscript.

- In “Kinematic parameter evaluation for the purpose of a wearable running shoe recommendation” [P5], I presented and evaluated a signal processing algorithm to compute parameters describing the foot orientation during running using IMUs and discussed the usage of those parameters for a shoe recommendation. In this work, I conducted the study, implemented the algorithm, and wrote the paper.
1.4 Contributions

- In “Classification of Match Phases in Handball” [P6], an algorithm to determine match phase during handball matches based on positional data of players on the field was introduced. I contributed to this paper by reviewing both the processing pipeline during the development process as well as the final paper.

- In “Sleeve Based Knee Angle Calculation for Rehabilitation” [P7], which was a nominee for the best paper award at the IEEE Conference on Body Sensor Networks 2018, a smart knee sleeve with two integrated IMUs was introduced. The data was used to compute the knee flexion-extension angle. I contributed to this paper by guiding the algorithm implementation, planning and conducting the evaluation study, and reviewing the final manuscript.

- The poster “Development of Algorithms for Computing Knee Stability Parameters Using a Sensor Equipped Knee Sleeve” [P8] presented a new approach to monitor knee stabilizing exercises using the above-mentioned knee sleeve. Algorithms measuring specific parameters of the knee-stabilizing exercises were introduced and evaluated. The work won the award for the best scientific poster during the ORTHO-congress. I contributed by guiding the algorithm development as well as reviewing the poster and abstract.

- In the publication “Movement Speed Estimation Based on Foot Acceleration Patterns” [P9], a regression-based algorithm to determine speed for different movements based on acceleration data was proposed. I contributed to the paper during the data acquisition and by reviewing the final manuscript.

- In “Evaluation of foot kinematics during endurance running on different surfaces in real-world environments” [P10], I analyzed the effect of six different surfaces on rearfoot kinematic parameters. I showed, that the kinematic parameters are elevated on stiffer and more consistent surfaces. I conceptualized the study, was involved in the data acquisition, implemented the algorithms and wrote the paper.

- In “Data-driven optimization of sensor placement for pressure insoles using particle swarm optimization” [P11], I used a particle swarm optimizer to find the best placements for pressure sensors on pressure insoles for the reconstruction of the vertical ground reaction force. I showed that the sensor placement resulting from
the optimization outperforms current placements from literature. In this work, I conceptualized the study, acquired the data, implemented the optimization problem and wrote the workshop paper. By the time of submission of this thesis, the manuscript was accepted but not published, yet.

Besides, I was involved in organizing a workshop on wearables for sports at the Conference on Pervasive and Ubiquitous Computing in 2016, which connected researchers from over the world and fostered scientific exchange [P12].

1.5 Overview of the thesis

This thesis is structured as follows: In chapter 2, I introduce fundamentals of sports science which are relevant to understand the content of the three main contributions. In section 2.1, I introduce the spatio-temporal parameters and foot kinematics which can be measured using foot-mounted IMUs and show why biomechanists are interested in those parameters. Similarly, the fundamentals of exercise physiology are introduced (section 2.2). Chapter 3 introduces the technical state of the art relevant for this thesis. Both the literature concerning wearables and especially IMU-based running parameter computation (section 3.1) as well as literature with respect to Big Data analysis with wearable fitness data (section 3.2) are reviewed. In chapter 4-6, I present the main contributions of this work, which were already summarized above. The contributions are discussed in chapter 7, before I draw a conclusion for the thesis and provide an outlook in chapter 8.
Part II

Fundamentals & state of the art
Fundamentals in sports science

In this chapter, the endurance running-related fundamentals in sports science that are required to understand the content of this work will be introduced. I will provide insights into spatio-temporal and foot kinematic parameters that biomechanists use to characterize the movement of runners. Afterwards, I will introduce the fundamentals of exercise physiology, which are relevant to understand the context of the Big Data contribution (chapter 6). For both disciplines, I also briefly introduce the laboratory ground truth systems, which scientists in the respective field currently use to characterize runners.

2.1 Spatio-temporal parameters and foot kinematics in endurance running

As the name indicates, the field of biomechanics deals with all mechanical aspects of biological systems. Even if limited to endurance running, this field of research is huge, because it investigates the locomotion of the complete human musculoskeletal system. As a complete introduction to biomechanics in endurance running would be beyond the scope of this work, I will focus on the spatio-temporal and kinematic parameters of the feet during running within this section.

In general, kinematics describe the motion of objects, without considering the forces causing the motion. In endurance running, especially the motion of limbs and joints have been investigated in the last century [38, 39]. Both the position and orientation, as well as changes of position and orientation of limbs during single strides or bouts of strides have drawn the attention of researchers. In this context, the absolute pose and orientation of segments in general, the relative orientation of segments (joint angles) as well as the rate of changes (angular velocities) are referred to as kinematic parameters. Spatio-temporal parameters describe the translation of the body segments (e.g. the foot) in space and the duration of sub-phases of a stride. A stride is defined as the motion of the foot between two consecutive Initial Ground Contacts (ICs) of the same foot [38](Figure 2). Within a stride, several events, phases, and parameters can be determined, which describe a runner and his movement pattern during running. In the following sections, I will introduce those
and describe their relevance for runners. The plots visualizing certain biomechanical aspects all rely on motion capture data acquired during the study described in chapter 4 and 5.

![Running gait cycle diagram](image)

**Figure 2: Visualization of the running gait cycle.**

### 2.1.1 Stride phases and stride events

A stride as well as its subphases are defined by different stride events. The first stride event is the IC. IC is the first instance in time, where the foot touches the ground in a stride cycle. The moment in time, where the foot leaves the ground, is called Toe Off (TO) because the toe is the last part of the foot touching the ground during a healthy running gait cycle. Between IC and TO lies Midstance (MS). At MS the foot flattens on the ground and the foot bears the full weight of a runner. Please notice, that MS is not an event, but rather a phase. However, an integral part of the algorithms introduced in chapter 4 and 5 will use an event during MS, where the foot shows minimal motion during ground contact. Within this work, this event will be called MS.

IC and TO subdivide the stride cycle into ground contact phase and swing phase. As outlined by Dugan et al. [40], the duration of these phases is often used to distinguish walking gait from running gait. In walking gait, the ground contact time $t_{gc}$ accounts for more than 50% of the full stride time $t_{stride}$. This implicates, that walking gait presents a double support phase, where two feet are on the ground. In comparison to that, the ground contact time for running gait $t_{gc}$ accounts for less than 50% of the full stride time $t_{stride}$. Thus, runners exhibit a double float phase, where no foot touches the ground [39].
Apart from the fact, that the share of \( t_{gc} \) can be used to classify walking and running gait, the absolute \( t_{gc} \) correlates to running velocity. The faster subjects run, the shorter \( t_{gc} \) (Figure 3). This is due to the fact, that longer ground contacts decelerate runners to a higher extent, which causes lower running speeds [41].

![Figure 3: Visualization of decreasing ground contact time \( t_{gc} \) with increasing stride velocity \( v_{stride} \).](image)

Both ground contact and swing phase can be further segmented into sub-phases. The ground contact phase can be segmented into absorption and propulsion phase. These two phases can be separated by MS. During absorption phase, the lower limbs absorb the forces acting due to IC, whereas during propulsion phase, the muscles are generating forces for propelling the runner forward. Swing phase can also be further segmented into initial and terminal swing.

### 2.1.2 Stride length, stride time and velocity

After segmenting the strides by finding stride events, the parameters stride length \( d_{stride} \), stride time \( t_{stride} \) and average stride velocity \( v_{stride} \) can be computed. Stride length describes the translation of the foot between two consecutive stride events of the same foot and stride time the duration between those events. In practice, IC is often used as the stride event to determine these two parameters, because it can be detected with different gold standard systems like force plates. The stride time \( t_{stride} \) can be converted to the stride frequency \( f_{stride} \), which is the inverse of \( t_{stride} \). Many authors also report the step frequency \( f_{step} \), which is
also called cadence. The relationship between $f_{\text{step}}$ and $f_{\text{stride}}$ is often simplified by

$$f_{\text{step}} = 2 \cdot f_{\text{stride}} \quad .$$

However, this equation is only an approximation for $f_{\text{step}}$. It is valid if the stride times of the left and the right foot are equal. If runners have bilateral instabilities, the equation does not hold anymore.

Stride length $d_{\text{stride}}$ and stride time $t_{\text{stride}}$ can be used to determine the average stride velocity $v_{\text{stride}}$, which is a surrogate for the running velocity during the stride:

$$v_{\text{stride}} = \frac{d_{\text{stride}}}{t_{\text{stride}}} = d_{\text{stride}} \cdot f_{\text{stride}} \quad .$$

This formula shows, that runners can increase their velocity by increasing $d_{\text{stride}}$ and by increasing $f_{\text{stride}}$ (Figure 4). However, the actual control mechanism for steering running velocity by those two parameters is highly individual and dependent on different factors. One of those factors is anthropometry [42]. Taller people with longer legs can rely on a higher stride length, while smaller people with shorter legs more often control speed by stride frequency. Similarly, gender influences the control mechanism because the average body height of men is higher than the one of women [43]. Research has shown, that runners’ self-selected combination of stride length and stride frequency is close to their personal economical optimum [42]. However, there seems to be a range around the self-selected optimal balance between $d_{\text{stride}}$ and $t_{\text{stride}}$, that does not affect running economy [44] (for definition of running economy see page 24 in section 2.2). For example, endurance runners use this range to reduce $f_{\text{stride}}$ in a fatigued state [45].

2.1.3 Foot joint angles

Apart from the translational and temporal parameters introduced above, researchers evaluated the orientation of the foot during the running gait cycle. The orientation is mainly influenced by the ankle, which consists of two joints: the talocrural joint between the tibia and the talus and the subtalar joint between the talus and the calcaneus (Figure 5).

The talocrural joint allows for dorsiflexion and plantarflexion in the sagittal plane. The degree of rotation of this joint during IC determines the classification of runners into rearfoot, midfoot, and forefoot runners.
2.1 Spatio-temporal parameters and foot kinematics in endurance running

Figure 4: Visualization of (a) stride length $d_{\text{stride}}$ and (b) stride frequency $f_{\text{stride}}$ over four velocity bins from 2-6 m/s. The intervals include the left bin edge (2-3 means $2 \leq v < 3$). The boxes represent the IQR and the whiskers extend up to 1.5·IQR. The plots indicate that velocity can be controlled either by increasing $d_{\text{stride}}$ or $f_{\text{stride}}$.

Figure 5: Visualization of the talocrural joint between the tibia and the talus and the subtalar joint between the talus and the calcaneus. Adapted from [38].
The names of the three categories of runners describe the part of the foot, which touches the ground first during ground contact (Figure 6). According to Altman and Davis [46], the strike type can be classified based on the angle between the sole of the running shoe and the ground, which I will call sole angle $\alpha_{sole}$ within this work (Table 1).

Table 1: Definition of strike type [46]. Plantarflexion is indicated by a negative angle, dorsiflexion by a positive angle.

<table>
<thead>
<tr>
<th>Strike type</th>
<th>Sole angle range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forefoot</td>
<td>$\alpha_{sole} \leq -1.6$</td>
</tr>
<tr>
<td>Midfoot</td>
<td>$-1.6 &lt; \alpha_{sole} \leq 8.0$</td>
</tr>
<tr>
<td>Rearfoot</td>
<td>$8.0 \leq \alpha_{sole}$</td>
</tr>
</tbody>
</table>

More than 80% of endurance runners are rearfoot runners [39, 47]. However, the share of midfoot and forefoot runners increases for faster professional endurance runners [48]. Also, recreational runners adapt their running style from a rearfoot to a midfoot/forefoot strike type with increasing running velocity [49]. One reason for this might be the fact that midfoot and forefoot runners have a smaller ground contact time $t_{gc}$, which decelerates them less during a running gait cycle [48, 50]. Based on these findings, Stearne et al. [51] evaluated, whether retraining of the strike pattern at IC from rearfoot to midfoot/forefoot strike is useful, which they could not prove. That retraining for recreational athletes is also not useful from a running pace perspective was proven by Larson et. al [47], because they showed that there is no statistical difference between the foot strike type with respect to the race velocity in marathon races.

Figure 6: Visualization of rear-, mid- and forefoot strike type.
In contrast to the talocrural joint, which only allows for rotation in the sagittal plane, a rotation around the subtalar joint is a tri-planar motion with respect to the cardinal planes of the human body. This is because the axis of rotation of the subtalar joint is tilted around 23° medially and 42° upward [38]. An inward rotation around the subtalar axis is called pronation, whereas an outward rotation is called supination. In terms of cardinal planes, pronation is a combination of external rotation, dorsiflexion, and eversion. Supination is a combination of internal rotation, plantarflexion, and inversion.

Based on the movement around the subtalar joint, endurance runners can be classified into pronators, neutral runners, and supinators. An inward rotation of the foot is a natural process during the running gait cycle. Runners usually touch the ground with the lateral side of their foot and start to pronate with the maximum pronation happening at around 40-50 % of the ground contact phase [52]. While also for neutral runners a certain amount of pronation over the neutral position of the subtalar joint is normal, pronators show a high maximum pronation angle with characteristic eversion of the foot. Supinators often don’t even reach the neutral position during the natural inward rotation and stay on the lateral side of the foot during the full ground contact phase in an inverted position (see Figure 7).

However, a clear definition of pronators, neutral runners supinators by a defined absolute angle cannot be found in literature. Due to individual anatomical differences and the inaccessibility of the talus bone from the outside, the axis of the subtalar joint cannot be determined. Researchers derived surrogates to measure the complex pronation movement. The rearfoot eversion angle and the achilles tendon angle are two of the most common definitions (Figure 7), even though both of these techniques reduce the measurement of the complex pronation/supination movement to a measurement of the eversion/inversion movement [52]. Apart from these absolute angular measurements, also pronation velocity (i.e. velocity of rotation around subtalar joint) has been used to investigate differences between runners of different pronation types [38].

Excessive pronation has long been linked to a higher risk of overuse injuries in running by many athletes and coaches. It was argued, that the internal rotation of the tibia caused by pronation is one of the causes for running-related knee injuries [53, 54]. Due to this reason, motion control shoes [55] and shoe inserts [56] have been developed to control rearfoot
motion and decrease the amount of pronation during the ground contact phase. However, nowadays more and more studies contradict the traditional belief of runners and coaches by showing no correlation between the amount of pronation and running injuries [57, 58, 59, 60]. Due to this reason, Richards et al. [61] concluded, that the current prescription of motion control shoes is not evidence-based. This scientific debate shows that the timing of pronation, the maximum pronation angle as well as the maximum pronation velocity have to be further investigated [39]. Using wearable devices, the current limitation of only capturing snapshots in short-term laboratory assessments can be overcome. By unobtrusively measuring those parameters during full training processes new insights into the mechanics of pronation and related injuries might be possible.

### 2.1.4 Laboratory measurements

Many different tools to assess running-related spatio-temporal and kinematic parameters exist. One popular tool is the force plate. The main purpose of a force plate is to measure the ground reaction force during running. Even though ground reaction forces are not in the scope of this thesis, force plates can also be used to determine stride events and stride phases by applying threshold techniques on the measured vertical ground reaction force [62]. In practice, force plates are often used in combination with other tools. For example, treadmills can be instrumented with force plates. The combination of these two tools allows the evaluation of stride phase durations and thus stride frequencies at different velocities. If also orientation parameters like foot joint angles are to be investigated, different motion capture systems are additionally used.
Motion capture systems can be divided into marker- and camera-based. Marker-based systems utilize retro-reflective markers, which reflect infrared-light and are attached to anatomical landmarks of the runner. Multiple synchronized cameras with infrared sensors capture the reflected light and reconstruct the position of the markers by triangulation. If at least three markers per body segment are attached, the orientation of the body segment in space can be computed. By comparing the orientation of body segments, joint angles can be computed. Camera-based motion capture systems only rely on conventional cameras without infrared sensors. New deep learning-based models perform well in detecting anatomical landmarks from images/videos [63]. Based on those anatomical landmarks, joint angles can be computed.

2.2 Physiology of endurance running

2.2.1 Basics of exercise physiology

To run, the human body needs energy. How this energy is provided and how the body converts this energy into movement have been under scientific investigation over the last century. Archibald Vivian Hill (1886-1977) was one of the pioneers of the research field of exercise physiology [64]. With his work on oxygen consumption during exercise, he built the foundation for today's understanding of the process in the human body while running. In this section, the fundamentals of this research field will be introduced.

During running there are three ways, how the human body produces energy, which muscles use to cause forward motion: alactic anaerobic, lactic anaerobic, and aerobic energy production. The following remarks about the three ways of energy production are summarized from Gastin's work [65]. All three options have in common that they produce Adenosine Triphosphate (ATP), which is a compound muscle cells need for contraction. During the alactic anaerobic energy production, phosphocreatine (PCr), which is a phosphorylated creatine molecule stored in muscle cells, donates a phosphate group to Adenosine Diphosphate (ADP) to produce ATP. The lactic anaerobic process describes the breakdown of carbohydrates in form of muscle glycogen. In this process, glycogen is broken down to pyruvic acid, which is converted to lactic acid by glycolysis. Alactic and lactic anaerobic energy production are capable of producing ATP at high rates but are limited in the total amount of energy production.
during a single exercise. This limitation is caused by the finite amount of PCr stored in the muscle cells as well as the decreasing pH by lactic acid. The third way of ATP production for muscle cells is the combustion of carbohydrates and fats in the presence of oxygen. While the amount of ATP being produced can be larger for aerobic energy production, the rate at which it is produced is lower compared to anaerobic energy production. Figure 8 visualizes the share of alactic/lactic anaerobic and aerobic energy while performing an exercise at maximum intensity with increasing exercise duration. It depicts that during maximum efforts of short duration like sprinting, muscles use mainly ATP produced aerobically. The longer the maximum exercise gets, the more the muscles use ATP from the aerobic energy production.

![Figure 8: Visualization of contribution of aerobic, lactic anaerobic and alactic anaerobic energy production to overall energy supply over the duration of exercise with maximum intensity. Adapted from [65].](image)

Since the exercise distance also increases with exercise duration, it becomes clear that the aerobic metabolism has a high impact on the performance of long-distance endurance runners. One parameter used to measure the capacity of the aerobic metabolism is the maximal oxygen consumption \( \text{VO}_2 \text{max} \). It describes the maximum rate of oxygen consumption in ml/min/kg and is a measure, how much energy can be produced using aerobic metabolism. \( \text{VO}_2 \text{max} \) values range between 46.5 ml/min/kg for recreational marathon runners and 85 ml/min/kg for professional marathon runners [66, 67] and can be greatly influenced by training. To understand factors influencing \( \text{VO}_2 \text{max} \), Figure 9 visualizes key components of the aerobic system. By ventilation, runners inhale oxygen, which diffuses from the air to the blood at the pulmonary alveoli.
within the lung. Within the blood, oxygen binds to the hemoglobin in the red blood cells and circulates through the body to the muscle cells due to the beating heart. Lastly, the oxygen diffuses into muscle cells, where it produces both energy and heat in combination with fats and carbohydrates.

Along this way, different factors are influencing the amount of oxygen, which reaches the muscle cells. Prampero [69] reports that VO\textsubscript{2max} is influenced by maximal cardiac output, the hemoglobin concentration in the blood, the peripheral diffusion and perfusion as well as the mitochondrial capacity.

Apart from VO\textsubscript{2max}, threshold concepts have been derived to evaluate the fitness status of endurance athletes [70]. As described earlier, lactic acid or its degradation product lactate is produced during lactic aerobic energy production. In the case of heavy exercise lactate is not broken down fast enough and accumulates in the blood. Figure 10 depicts the concentration of lactate with increasing intensity. Based on this curve, two thresholds were defined: the aerobic threshold and the anaerobic threshold. The aerobic threshold is located at the intensity, where lactate starts accumulating above resting level. The anaerobic threshold is defined, where lactate starts to rise rapidly. However, one point of criticism in literature is that many different definitions of those thresholds exist, which makes reported values hard to compare [70]. Nevertheless, those thresholds are used to compare endurance athletes. Usually, those thresholds are reported as a percentage of the intensity at VO\textsubscript{2max} [71]. The interpretation of this value is that people, which are capable of running at higher intensities without accumulating lactate, can run longer at these high intensities and will ultimately be faster in endurance running races.

Apart from VO\textsubscript{2max} and the anaerobic threshold concept, running economy is used to describe the endurance performance of an athlete. This
2 Fundamentals in sports science

Parameter is the link between biomechanics and physiology in endurance running. Running economy is reported in ml/min/kg and is defined as steady-state VO2 at a constant running velocity [72]. Especially for professional endurance runners with high VO2 max, differences in speed at a steady-state VO2 could be observed. The best runners were faster at the same VO2 level, which indicates that those athletes run more economically. Research has shown, that kinetic, spatio-temporal, and kinematic running parameters influence the running economy [44, 72, 73]. Thus, running economy can be seen as the factor linking exercise physiology and biomechanics in endurance running.

2.2.2 Laboratory measurements

Even though there is scientific debate on the best way to determine the cardio-pulmonary fitness status of a runner [74], VO2 max, the aerobic/anaerobic threshold and the running economy are currently the most popular parameters to describe the fitness of endurance athletes and evaluate responses to training regimes [71]. Those parameters are usually determined inside laboratories during cardio-pulmonary exercise tests [75].

VO2 max and running economy are measured using pulmonary gas exchange during exercise. In the endurance sport context, these devices not only determine the respiration volume but also determine the concentration of inhaled and exhaled O2 and CO2. From these measurements, the
consumed oxygen \( VO_2 \) can be examined with gradually increasing exercise intensities. While different test protocols for those tests exist [76], intensities are usually increased every 2-3 minutes until either subjects indicate a stop of the test due to exhaustion or the \( VO_2 \) plateaus. The value of \( VO_2 \) at this plateau is the value of \( VO_2 \) max.

Even though methods to determine the aerobic and anaerobic thresholds from spirometry exist [77], the thresholds are usually determined by taking blood samples and determining the lactate concentration during the ramp test performed for the estimation of \( VO_2 \) max. From the resulting lactate concentration curve, one of the methodologies summarized by Faude et al. [70] can be used to determine the threshold values.

Running economy is also measured through pulmonary gas exchange by spirometry while running at a steady-state, sub-maximal intensity for 3-15 minutes. A reference velocity commonly used for professional athletes is 16 km/h, even though other reference values ranging from 12-21 km/h exist [73].
3 State of the art

Within this chapter, I want to introduce the state of the art both in endurance running-related wearables as well as in Big Data related research based on wearable sensor data. In the section on wearables, I will first briefly introduce the three most commonly used wearable sensors. Afterwards, I will focus on IMU-related research. Even though I also outline studies that used IMUs attached to various positions on the human body, I will mainly focus on the foot sensor position due to the position’s relevance for this work. In the Big Data section, I will outline the challenges of working with Big Data in science in comparison to traditional studies, introduce a process for assuring Big Data studies of good quality, list already used large-scale data sets in science, and show the applications researchers realized with these data sets.

3.1 Wearables in endurance running

3.1.1 Sensor systems

GPS sensors One wearable sensor which is already heavily used by many endurance runners is the GPS sensor. Those sensors determine the position of a runner over time by receiving signals from satellites circling the earth (Figure 11a). The satellites have very precise clocks, which are synchronized to each other. They continuously transmit a radio signal with their current timestamp and information regarding their position in their orbit. The GPS sensors receive those transmission signals and can compute the position of and the distance to the satellites due to the constant speed of the radio waves. Using triangulation, the position (latitude, longitude, altitude) of the sensor can be calculated [78].

Runners use this technology not only to visualize their running tracks (Figure 11b), but also to track the distance of their runs as well as their running velocity. GPS-sensors are usually embedded in smartphones or smartwatches that runners carry during their run. Especially by calculating velocity on the fly, those devices are also capable of giving real-time feedback regarding performance during a run.

Heart rate tracker To track the heart rate during workouts, runners use two different technologies. On the one hand, they use chest straps,
which measure electrical changes caused by the heart contraction on the skin via electrodes. During one cardiac cycle, the heart depolarizes and repolarizes to contract its muscles and to pump blood through the human body. Similar to an Electrocardiogram (ECG), these electrical changes are measured and digitized. By using algorithms tailored to ECG signals like the algorithms of Pan and Tompkins [79] or ElGendi [80], distinctive fiducial points of one heart cycle can be detected, which allows a measurement of the heart rate in real-time.

The other sensors used to track heart rate rely on PPG. Those sensors consist of two diodes: a Light Emitting Diode (LED) and a photodiode measuring reflected light. While measuring the heart rate, the LED constantly emits light on the skin. The amount of reflected light, which is measured by the photodiode, depends on the volume of blood flowing through the microvascular bed of tissue [81]. The changes in the volume of blood with every heartbeat result in a signal, which can be used to determine heart rate. Nowadays, many commercially available smartwatches have integrated PPG sensors, making the use of chest straps unnecessary. However, current studies on the accuracy of PPG sensors indicate higher errors for heart rate based on PPG sensors during activity [82, 83] and suggest the usage of chest straps for measuring heart rate during exercise [84].

Inertial measurement units IMUs are Micro-electricalmechanical Systems (MEMS), which are capable of measuring acceleration and angular velocity in space. The following section describing the basics of
3.1 Wearables in endurance running

Accelerometers and gyroscopes are an adapted version of Aminian and Najafi [85].

Accelerometers sense acceleration using a mass, which is connected to a frame by a beam. The beam can be seen as a damped spring, which deforms when the frame is accelerated. The deformation is transduced to changes in electrical impedance for resistive and capacitive accelerometers or charge generation for piezoelectric accelerometers. The main difference between resistive/capacitive accelerometers and piezoelectric accelerometers is that the first not only measure acceleration while motion, but also at rest. Those kinds of sensors constantly measure the gravitational acceleration, which enables the calculation of the sensor’s inclination at rest. During motion, however, the superposition of the gravitational acceleration over the motion acceleration complicates the computation of velocity and displacement of the sensor by integration of acceleration over time. In this case, algorithms to separate gravitational and motion components of the acceleration have to be applied. Additionally, the presence of drift in the velocity signal obtained from accelerometer data makes the computation of displacement even harder.

Gyroscopes measure angular velocities using the Coriolis force. In a MEMS gyroscope a mass is vibrated either by piezoelectric, electromagnetic, or electrostatic energy. If the vibrating mass is rotated, the Coriolis force causes a deviation of the movement path of the vibrating mass, which can be detected using a capacitive, resistive, or piezoelectric effect. Due to the fact, that the Coriolis force is proportional to the angular velocity \( \omega \) of the rotation, gyroscopes are capable of capturing angular velocity data. The angular velocity data of gyroscopes can be used to compute rotations of objects by integration over time.

3.1.2 Applications of IMUs in endurance running

Computing endurance running-related parameters using IMUs is a sub-field of IMU-based gait analysis. Apart from the sports field, this branch of research has many medical applications since gait disorders can be unobtrusively detected [86]. Many of the applications in endurance running were inspired and adapted from applications and algorithms originally developed for such medical use cases. As in the medical domain, the algorithms and applications in endurance running differ in the number of IMU sensors the data originates from, the position of the sensors, and
the types of parameters being estimated. Table 2 gives an overview of these variables.

Table 2: Overview of number of IMU sensors, sensor positions and type of endurance running related parameters based on IMU data with references from literature.

<table>
<thead>
<tr>
<th>Num. sensors</th>
<th>Position</th>
<th>Type of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>single IMU</td>
<td>foot</td>
<td>spatio-temporal [87, 88]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kinematic [22, 89, 90, 91, 92, 93]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>physiological [94, 95]</td>
</tr>
<tr>
<td></td>
<td>tibia</td>
<td>kinetic [96, 97]</td>
</tr>
<tr>
<td></td>
<td>lower back</td>
<td>kinematic [98, 99, 100]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kinematic [101]</td>
</tr>
<tr>
<td></td>
<td>chest</td>
<td>spatio-temporal [104, 105]</td>
</tr>
<tr>
<td></td>
<td>wrist</td>
<td>kinematic [106, 107, 108]</td>
</tr>
<tr>
<td></td>
<td>single joint</td>
<td>physiological [109, 110]</td>
</tr>
<tr>
<td>multiple IMUs</td>
<td>lower extrem.</td>
<td>kinetic, kinematic, physiological [112]</td>
</tr>
<tr>
<td></td>
<td>full body</td>
<td>kinematic [113]</td>
</tr>
</tbody>
</table>

Sensor position  Commercially, the most commonly used sensor positions in endurance running are the wrist within smartwatches, the chest within heart rate chest straps (e.g. Garmin HRM Pro, Polar H10, Wahoo Tickr X), and the foot within foot pods attached to the laces on the instep of the running shoe (e.g. Stryd, Garmin Foot Pod, Suunto Foot Pod). Besides, first smart running shoes are launched, which have IMUs integrated within the sole of the running shoe (e.g. UnderArmor HOVR Velociti 2). One of the advantages of those sensor positions is that the integrated IMUs can be combined with other sensors or devices (e.g. heart rate trackers, smartwatch), which consumers already accepted. This adds usability because users do not need multiple devices. By adding IMUs, the feature set of those devices can be increased.
In science, researchers did not limit themselves to only one sensor and also used multiple IMUs to compile Body Sensor Networks (BSNs), which enabled a more holistic tracking of the movement of the human body. The number of sensors in these cases is defined by the number of extremities/joints which shall be monitored. Thus, IMUs are also mounted to the upper leg, the upper and lower arm as well as to the head [113].

In the scope of this thesis, I evaluated data of IMUs attached to the foot due to a large amount of different endurance running parameters that can be estimated based on the foot movement. Researchers have attached IMUs externally to the running shoe on the heel cap [89, 91], the lateral side below the ankle [92] or the laces [22, 90, 93, 114]. The advantage of those positions from a usability perspective is, that users can easily mount the sensors to different pairs of running shoes and can easily take the sensors off for charging. Others have placed the IMUs internally within a cavity inside the sole of a shoe [87]. The advantage in terms of usability of this position is that the running-related parameters are estimated unobtrusively and that the user can not simply forget the sensors. However, the different sensor positions for foot-mounted IMUs have not been evaluated concerning the accuracy of spatio-temporal and kinematic parameters, yet.

**Type of parameters** IMUs were used to derive different types of parameters, which can be grouped into spatio-temporal, kinematic, kinetic, and physiological parameters.

*Spatio-temporal parameters* can be estimated with data from all sensor positions. Especially the stride time and the stride frequency are two parameters, which can be computed independently of the sensor position due to the periodic nature of the IMU signal during running. From an algorithmic point of view, either time-domain, frequency-domain or autocorrelation analysis can be applied [107] to determine those parameters. Running velocity and distance are often computed by reconstructing the translation of the respective body segment. For the computation of translational parameters, but also rotational parameters like the sole angle, different methods exist. A detailed review on how those parameters can be computed using a single foot-mounted IMU can be found in section 3.1.3.

To compute *kinematic joint angles* during the running gait cycle, two IMUs are required. In those use cases, the IMUs are fixed to the segments,
which the joint connects. By comparing the orientation of the IMUs to each other, the specific joint angles can be estimated. For a bigger overview of literature, I would like to refer to the review paper on the computation of IMU-based kinematic parameters of Weygers et al. [111], which also includes running use cases.

Other authors also introduced algorithms to compute running-related kinetic parameters using IMU data. Especially, parameters extracted from the vertical component of the ground reaction force were approximated. The traditional approach was to correlate the peak acceleration measured by IMUs to either the maximum vertical loading rate [97] or the peak vertical ground reaction force [101]. The sensor positions in those publications were either the tibia [97] or the sacrum [101]. The advantage of those sensor positions in comparison to the foot sensor position is that the acceleration signal is already dampened by the kinematic chain below the respective position. The maximum peak acceleration of a foot-mounted IMU reaches up to 70 g [115], which often causes problems with a saturation of the acceleration signal. However, a recent publication by Derie et al. [96] also introduced a machine learning approach based on multiple time-series features, which might be transferable to the foot-position in the future.

Besides, IMU data was used to estimate physiological parameters like the energy expenditure during running [109, 110]. To do so, algorithms were introduced which counted activity levels beyond defined thresholds and correlated these activity counts to energy expenditure. In recent years, this approach was extended by the introduction of running power. Running power is a measure that was introduced to analyze the metabolic workload during running in an instantaneous, reliable, and sensitive manner [116]. Different commercial enterprises have introduced ways to measure running power in real-time using either smartwatches (e.g. Garmin, Polar) or foot pods (Stryd). While running power derived by smartwatches relies on GPS measurements, it relies on IMU data for the foot pod technology. Nevertheless, both options require additional sensors like barometers, air resistance, or temperature sensors to increase the accuracy of the metric. In scientific studies, the IMU-based method has proven to determine running power reliably [94, 95] and more accurately then the GPS-based approach [116].

Last but not least, IMU networks can be used to track the position and the orientation of a subset of body segments or even all body segments.
This approach aims to be an alternative to marker- or video-based motion capture systems. To deal with the sensor drift of all IMUs attached to the body, biomechanical constraints limit the solution space for position and orientation of body segments. Among others, these constraints include the connection of certain body segments at all times, special type of joints between certain body segments (e.g. knee as hinge joint) or a maximum range of joint motion [113]. From an algorithmic point of view, the constraints can be incorporated differently into the models. Where Luinge et al. [117] computed the orientation of all body segments using an Extended Kalman Filter (EKF) and incorporated the biomechanical constraints via a least-squares filter, Kok et al. [113] used an optimization approach and included the constraints in the optimization function. Another approach to IMU-based motion tracking, which allows for the computation of spatio-temporal, kinetic, kinematic, and physiological parameters, was introduced by Dorschky et al. [112]. The underlying idea of their approach is to find a motion for a biomechanical model by minimizing the difference between measured IMU data and simulated IMU data of virtual sensors placed on the model in the objective function of an optimal control optimization. In comparison to the work of Kok et al. [113], the biomechanical model also includes muscles (musculoskeletal model), which allows the computation of an effort term and thus an optimization with respect to the physiological effort of running.

### 3.1.3 Data processing for spatio-temporal and kinematic parameters using foot-mounted IMU

In this section, I review the literature regarding the extraction of spatio-temporal and kinematic parameters based on data of a single foot-mounted IMU. Figure 12 gives an overview of steps that have to be considered when implementing algorithms capable of parameter extraction.

#### 3.1.3.1 Sensor settings

When designing applications with IMUs, the sensor used to measure the data and especially its underlying parameters have to be chosen depending on the use case. The sensor parameters which have to be considered are the sampling rate $f_s$, the resolution of the sensors as well as the operating ranges of the accelerometer and gyroscope. While $f_s$ defines the temporal resolution, the operating range of the sensors defines the
3 State of the art

4. Parameter computation

Different algorithmic approaches to compute parameters

3. Stride segmentation

Segmentation of strides by detecting gait relevant events

2. Sensor calibration

Scaling of sensor signals to correct unit and alignment of coordinate systems

1. Sensor settings

Definition of sampling rate as well as accelerometer and gyroscope range

Figure 12: Overview of four steps which have to be considered when implementing algorithms capable of extracting spatio-temporal and kinematic parameters using foot-mounted IMUs.

minimum and maximum acceleration/angular rate the accelerometer/gyroscope can measure. Lastly, the resolution of the sensor defines the number of different discrete sampling values between the minimum and maximum defined by the operating range of the sensor.

Sampling rate There is no clear indication of how to set the sampling rate $f_s$ for foot-mounted IMU running analysis. Bailey and Harley [115] evaluated the effect of reducing the sampling rate from 1 kHz to 500 Hz, 250 Hz, 125 Hz and 62.5 Hz by sub-sampling the IMU signal they acquired of subjects, who ran on a treadmill. They showed that the error for the position and running velocity they computed using an EKF did not increase until 250 Hz, but with sampling rates below that. Similar results were obtained by Potter et al. [114]. Mitschke et al. [88] also subsampled data acquired with a sampling rate of 1 kHz. They found that the sampling rate to determine kinetic parameters has to be 500 Hz to not miss acceleration peaks within the signal, but that a sampling rate of 200 Hz is sufficient to compute temporal as well as kinematic parameters. Lastly, an interesting approach concerning the sampling rate was proposed by Liu et al. [104], who used sparse adaptive sensing for an online running analysis. The idea of sparse adaptive sensing is to adaptively use different sampling rates for different parts of a signal. Thus, a high temporal resolution can be obtained when needed. The authors showed that they could reduce the energy consumption by 83.6 % while maintaining an accuracy of 97.7 %. While this approach is very interesting, the authors only evaluated it using a chest-mounted IMU. It has to be noted, that chest-mounted
IMUs show a narrower frequency band than foot-mounted IMUs and are thus easier to represent with a lower sampling rate.

**Operating range of accelerometer** With respect to the operating range of the accelerometer, Bailey and Harle [115] reported that the acceleration peaks around IC within IMU signals reach up to 70 g, while the rest of the signal has a lower acceleration range. They conducted a study with two IMUs attached above each other on the lateral side of the running shoe below the ankle. One of the IMUs had an operation range of ±16 g and the other one ±200 g. They compared, how much the errors of stride parameters were influenced by accelerometer saturation of the ±16 g sensor. They showed better position and velocity estimates with the higher operating range. Mitschke et al. [118] evaluated the influence of the acceleration range on running parameters by acquiring acceleration signals with a range of ±70 g and afterwards simulated signals with a lower range by cutting the signals at ±32, ±16 and ±8 g. They proposed to use an operating range of ±32 due to higher errors for the parameters stride length and velocity with lower operating ranges. Similar results were obtained by Potter et al. [114], who also simulated lower operating ranges by cutting an acceleration signal originally sampled with ±100 g. Based on their results, they suggest using an operating range of at least ±50 g to obtain the best possible results. Nevertheless, their results also indicate, that errors are also low for an operating range of ±16 g, if the running velocity does not exceed 6 m/s.

**Operating range of gyroscope** Potter et al. [114] also evaluated the operating range of the gyroscope similarly to the one of the accelerometer by cutting an angular rate signal originally sampled by ±2000 deg/s at ±1500, ±1000, ±750 and ±500 deg/s. To accurately compute stride length and velocity using a sensor fusion technique (strap-down integration, see chapter 3.1.3.4), they suggest a gyroscope range of at least 1000 deg/s.

To the best of my knowledge, the effect of the resolution of the IMU on stride parameter has not been evaluated, yet.

**3.1.3.2 Calibration of IMU sensors**

After choosing the sensor settings, the IMUs have to be calibrated for the computational steps to follow. Two different types of calibrations can be distinguished: sensor calibration and sensor-to-segment alignment. For
both types of calibration, different methods exist in literature, which will be briefly summarized in the following.

**Sensor calibration** The sensor calibration has three different purposes which are relevant especially if signal processing algorithms are to be applied later in the processing pipeline, which require the measured signals to be in physically interpretable units:

1. Correction of misalignment of sensor axes: During manufacturing the coordinate axes of the 3D-accelerometer and the 3D-gyroscope are not perfectly aligned [119]. In mathematical terms, this misalignment can be represented by a rotation matrix $K$.

2. Scaling of voltages measured by MEMS technology to physically interpretable units: The digitized outputs of MEMS sensors are voltages. These voltages can be scaled to the units of acceleration and angular rate. For accelerometer signals, the unit is $\text{m/s}^2$ or $g$ (earth acceleration). For gyroscope signals, the unit is $\text{deg/s}$ or $\text{rad/s}$. In mathematical terms, the scaling can be achieved by multiplication with a diagonal scaling matrix $S$.

3. Removal of sensor bias: The MEMS sensors often produce a non-zero output, even if the IMU is in a static position. This sensor bias causes an accumulated drift error of position and orientation estimates when the sensor signals are integrated. The bias can be mathematically modeled by a bias vector $\vec{b}$.

Figure 13 visualizes the effect of the described issues on an acceleration signal $\vec{a}$. $\vec{a}$ is misaligned by matrix $K_a$, scaled by matrix $S_a$ and translated by a bias vector $\vec{b}_a$ to produce the measured acceleration $\vec{a}_m$.

To deal with the described effects, a sensor model was formulated by different authors [119, 120, 121], which mathematically describes how the actual acceleration $\vec{a}$ and angular rate $\vec{\omega}$ can be computed from the measured acceleration $\vec{a}_m$ and $\vec{\omega}_m$.

\[
\vec{a} = K_a^{-1} S_a^{-1} (\vec{a}_m - \vec{b}_a) \\
\vec{\omega} = K_\omega^{-1} S_\omega^{-1} (\vec{\omega}_m - \vec{b}_\omega)
\]

(3)

In this equation, the indices $a$ and $\omega$ indicate that the alignment matrix $K$, the scaling matrix $S$ and the bias vector $\vec{b}$ are different for acceleration and angular rate. In some publications, the equation above is extended.
3.1 Wearables in endurance running

by a noise term, which also has to be removed by subtraction from the measured values before rotating and scaling the signal [119, 121].

The goal of the sensor calibration procedure is to compute $K$, $S$ and $\vec{b}$ for both the accelerometer and the gyroscope, such that $\vec{a}$ and $\vec{\omega}$ can be reconstructed from the measured signals using Equation 3. The general idea for all different calibration procedures is to compare multiple sets of measured signals to known accelerations or angular rates. The acceleration signal is usually compared to gravity when the sensor is in a static position [119, 120, 121]. For the gyroscope, different options exist. A very precise option is to use a mechanical turntable, which is capable of producing rotations with adjustable angular rates [122]. In this case, a set of measurements with rotations of defined angular velocity can be produced. However, such turntables are very costly and an in-field calibration is not possible. Due to these reasons, other calibration routines were developed. Ferraris et al. [120] introduced a methodology, where the sensor has to be rotated by a defined angle around each axis. By integrating the angular rate signal, the resulting angle can be compared to the defined angle of rotation. This approach was extended by Tedaldi et al. [121]. The method of these authors only requires the sensor to be placed in at least nine static positions. They exploit the fact that the rotation between two static positions can be measured by angular rate integration. By comparing the orientation of the measured acceleration to gravity in the static positions, a second measure for the orientation in space (without heading) can be computed. By acquiring nine of those orientation pairs, the calibration parameters can be estimated from the acquired data.
Sensor-to-segment alignment  After the sensor calibration and before data acquisition, often a sensor-to-segment alignment procedure is performed. This technique is required if the orientation of a specific body segment, in the scope of this thesis the foot, has to be estimated. For this purpose, the coordinate system of the sensor and the coordinate system of the segment have to be aligned, which corresponds to a rotation of the measured signals by a rotation matrix $\mathbf{R}_{align}$. Similar to the sensor calibration, different methods were introduced to determine $\mathbf{R}_{align}$, which were summarized by Pacher et al. [123] in a review on sensor-to-segment alignment methodologies for lower-body kinematic analysis. The authors identified four different alignment methodologies in literature, which help to find the correspondence between one axis in the sensor frame with one axis in the segment frame:

1. Manual alignment: The axis in the sensor frame is manually aligned with the axis in the segment frame by attaching the IMU accordingly.

2. Static alignment: After attaching the sensor with a random orientation to the segment of a subject, the subject adopts a defined posture. In this posture, one of the axes of the segment has to be aligned with gravity. In this case, the misalignment between the IMU sensor axis and gravity can be estimated.

3. Functional alignment: The IMU sensor axis is aligned to a segment axis by a rotation movement around one of the segment’s axis. From the gyroscope data, the rotation axis is extracted using a Principal Component Analysis (PCA). For foot-mounted IMUs, the data can originate from a controlled movement before the data acquisition [P2] or from the movement during running as proposed by Falbriard et al. [90]. The later procedure assumes, that the principle component of rotation during the running gait cycle is around the medial-lateral axis.

4. Anatomical alignment: This alignment procedure is based on anatomical landmarks. The basic idea is to align the sensor axis to an axis defined by two anatomical landmarks. To find the later axis orientation in space, an external device is used. The external device connects two anatomical landmarks linearly and has an IMU attached, which is capable of measuring the orientation of the line connecting the anatomical landmarks. By comparing the measurement of the IMU on the segment to the one on the external
device, the sensor-to-segment alignment can be performed. More information about the anatomical calibration procedure can be found in the work of Picerno et al. [124].

1. Manual
2. Static
3. Functional
4. Anatomical

Figure 14: Illustration of the four sensor-to-segment or sensor-to-shoe alignment techniques from literature. The blue squares illustrate the IMU attached to the shoe with its coordinate system in black. The orange coordinate system is the shoe coordinate system. The green arrows indicate the axis of the shoe coordinate system, which is aligned during the calibration process. The red square in the anatomical technique illustrates the IMU on the external devices which is used to determine the segment axis.

The four methodologies, which are illustrated in Figure 14, are capable of determining one vector pair in the IMU sensor coordinate system and the segment coordinate system for one single axis. To fully align the two coordinate systems, at least two vector pairs need to be determined. The third orthogonal axis in both coordinate systems can then be computed using the cross-product. After finding the corresponding axes, the rotation matrix $R_{align}$, which rotates the sensor to the segment coordinate system, can be computed as a solution to Wahba’s problem [125], which is to find the rotation matrix between two coordinate systems given a set of observations in both coordinate systems. The problem can be solved in different ways [126]. Within this work, I used an implementation of the approach presented by Markley [127] using a Singular Value Decomposition (SVD).

The methods used to determine the two vector pairs in the two coordinate systems can differ. For example, I used a combination of a static
(superior/inferior axis) and a functional alignment (medial/lateral axis) to determine the vector pairs (see chapter 5).

3.1.3.3 Stride event detection

For most of the algorithms used in running analysis, the first step in the processing pipeline is to detect stride events within the IMU signal. The three most common events, which have to be detected are IC, MS and TO. To better understand the task of event detection, Figure 15 depicts the acceleration and gyroscope signal of one running stride cycle from a foot-mounted IMU sensor placed in a cavity within the sole of a running shoe. The vertical lines indicate the stride events and the shaded areas visualize the ground contact and the swing phase. The IC is well visible due to the high acceleration peaks during or shortly after the initial ground contact. During MS, the sensor is exposed to the least motion during the gait cycle, which causes a minimum of the gyroscope signal in the sagittal plane (medial/lateral axis, blue line). Besides, the majority of the acceleration measured during MS is caused by the gravitational acceleration, which is in this case aligned with the $z$-axis (acceleration in superior/inferior direction, green line). Due to the dynamic nature of running, the TO event is hard to detect in the IMU signal [128, 129]. The TO event in walking gait is indicated by a zero-crossing of the gyroscope in the sagittal plane [6]. However, this zero crossing is not apparent in running gait due to the higher and more dynamic foot lift. While for some subjects, a small discontinuity in the gyroscope signal is apparent during TO, there is no clear overall pattern that can easily be detected algorithmically.

In IMU-based running analysis the majority of researchers used threshold- and peak detection-based signal processing techniques to find signal characteristics that correlate to events. Especially the IC with its distinct peaks in both the acceleration and the angular rate signal was often determined using those methods [22, 92, 129, 130]. However, the type of signal (acceleration, gyroscope), as well as the sensor axis, differed between publications. Due to this fact, Falbriard et al. [128] evaluated which signal characteristic was most suitable to determine IC. They not only included peaks in the acceleration and angular rate signal in the evaluation, but also peaks in the derivative of both signals. For each signal characteristic, they computed the error against a synchronized force plate. They found the maximum of the gyroscope signal in the sagittal plane (medial/lateral
3.1 Wearables in endurance running

Figure 15: Acceleration and gyroscope signal of foot-mounted IMU (Cavity). The vertical lines represent the IC, MS and TO events, which were set manually. The $x$-axis points in medial/lateral direction, the $y$-axis in anterior/posterior direction and the $z$-axis in superior/inferior direction.
axis) to be the most reliable feature to detect IC even though this feature has a certain bias (11 ms).

They also evaluated the accuracy of signal characteristics for detecting TO. Similarly, they found the maximum of the angular rate signal in the sagittal plane after IC to be the most accurate measure for TO, because it was the most reliable feature with the lowest standard deviation for the error. In Figure 15, this feature is the peak in the medial/lateral angular rate signal (\( \alpha \), blue color) before the marked TO. However, they also report, that this feature has quite a large bias (-24 ms). A similar bias was reported by Mo and Chow [129], who used a simple acceleration threshold after stance phase for TO detection. These biases underline that detecting TO using foot-mounted IMUs is a difficult task.

As already mentioned in section 2.1.1, MS is rather a phase than an event. However, finding the time instance during the gait cycle with the least motion is an integral part of the strap-down integration algorithm introduced in section 3.1.3.4 (page 45). MS detectors generally operate in a defined window after IC. The duration of the search window is defined by the running velocities and the corresponding ground contact time \( t_{gc} \). MS happens within the first 50\% of ground contact phase [40]. Thus, the duration of the search window has to be at least 50\% of the maximum ground contact time to assure that MS is within the window. Due to the correlation of \( t_{gc} \) to running velocity, the maximum ground contact time and consequently the duration of the search window is determined by the lowest velocity MS has to be detected for. From Figure 3 a search window duration of 200 ms can be deduced due to the maximum ground contact for running velocities down to 2 m/s was 400 ms.

To find MS within the search window, three different methods exist which were summarized and compared by Skog et al. [131]. The first two methods rely on the acceleration signal. The acceleration moving variance detector computes the variance of the acceleration signal within \( N \) samples of the search window over time. The \( N \) samples with the smallest variance contain MS. The acceleration magnitude detector computes the magnitude of the measured acceleration and compares it to the magnitude of gravity (9.81 m/s\(^2\)). The index of the acceleration sample with the minimum difference of its magnitude to gravity is MS. The last method relies on the gyroscope signal. The angular rate energy detector decides for MS to lie within \( N \) samples of the search window when the energy of the \( N \) angular rate measures is minimal. The index of MS is then often
chosen in the middle of the \( N \) samples. The comparisons of the three methods by Skog et al. [131] found the angular rate energy detector to find MS most accurately.

In the domain of running analysis using foot-mounted IMUs mainly signal processing techniques were used to find individual events or segment strides. However, I want to briefly mention other stride segmentation techniques, that were either applied to running IMU analysis using another sensor position or in walking gait analysis for health applications. Barth et al. [132] introduced an approach called subsequence dynamic time warping. While this method relies on template matching, it allows the signal to be segmented to differ in duration, amplitude and form of the original template. Besides, Hidden Markov Models (HMMs) have been used to segment strides [133, 134]. Within the technique, the stride phases are modeled as states in a HMM. After training the HMM the individual samples of the IMU signal are mapped to the states of the model (stride phases) and the transitions between the states are labeled as the gait events. Similar to this approach, Robberechts et al. [135] introduced a deep learning approach, which works similarly. In their work, they combined a Recurrent Neural Network (RNN) with a structured perceptron model. While the RNN learned how to extract representative features from the IMU signal, the structured learning perceptron on top of the RNN mapped the features to a series of states with the gait events being the transitions between states. Lastly, generic cyclicity estimators can be used to segment strides. Generic cyclicity estimators are designed to find periodic signals within a data stream. Sprager et al. [136] introduced such an algorithm, which relies on autocorrelation and is capable of detecting different kinds of repetitive actions within IMU signals.

3.1.3.4 Parameter computation

In literature different ways how to compute spatio-temporal and kinematic stride parameters are described. They can be grouped into biomechanical, digital signal processing, and machine learning approaches.

Biomechanical The easiest, yet most inaccurate way to compute stride parameters from IMU data is to simply use biomechanical correlations. Many stride parameters correlate to duration features like stride time or ground contact time [38], which can be computed based on detected events and sampling frequency \( f_s \). If anthropometric data of subjects like gender and body height are taken into account, the accuracy of stride
parameters based on such correlations can be further improved. I used such an approach in chapter 4 to evaluate the accuracy of stride length and average stride velocity based on stride time obtained from a foot-mounted IMU and anthropometric data. However, these biomechanical approaches are limited to known correlations of biomechanical features detectable with IMUs.

**Signal Processing** The majority of the research community used digital signal processing techniques to compute stride parameters using IMU data. After calibrating the IMU to SI-units, the acceleration \( \vec{a}[n] \) and angular rate data \( \vec{\omega}[n] \) can be used to determine displacement \( \vec{d}[n] \) and orientation \( \vec{\alpha}[n] \) by numerical double integration or integration, respectively.

\[
\begin{align*}
\vec{d}[n] &= \frac{1}{f_s} \sum_{k=0}^{n} \left( \frac{1}{f_s} \sum_{m=0}^{k} \vec{a}[m] + \vec{v}_0 \right) + \vec{d}_0 \\
\vec{\alpha}[n] &= \frac{1}{f_s} \sum_{k=0}^{n} \vec{\omega}[k] + \vec{\alpha}_0
\end{align*}
\]

In these formulae \( \vec{v}_0 \) is the initial velocity of the IMU sensor, \( \vec{d}_0 \) the initial displacement and \( \vec{\alpha}_0 \) the initial orientation. There are two problems associated with this formula, which need to be solved:

1. \( \vec{v}_0, \vec{d}_0 \) and \( \vec{\alpha}_0 \) are unknown.
2. Without gravity removal, the displacement can not be computed using numerical double integration (see section 3.1.1).

These problems could be solved by applying a zero-velocity update during each gait cycle. The zero-velocity update assumes, that the foot is flat on the ground and has zero velocity during MS. This assumption has been used by different researchers [6, 87, 88], because it solves the described problems by reinitializing orientation and displacement within each gait cycle. Thus, \( \vec{v}_0 \) can be set to zero for each gait cycle. Also \( \vec{d}_0 \) and \( \vec{\alpha}_0 \) can be set to zero, if no absolute position or orientation (heading) is of interest, but just changes within one stride, which is usually the cases when determining stride parameters. Thus, equation 4 can be simplified to
3.1 Wearables in endurance running

\[ \ddot{d}[n] = \frac{1}{f_s^2} \sum_{k=0}^{n} \sum_{m=0}^{k} \ddot{d}[m] \]

\[ \ddot{a}[n] = \frac{1}{f_s} \sum_{k=0}^{n} \ddot{\omega}[k] . \]  

Besides, the gravitational acceleration can be removed from the acceleration signal using the zero-velocity assumption. If the foot is expected to be flat on the ground during MS, the accelerometer measures only gravity. By computing the change in orientation using the gyroscope data, the orientation of gravity can be computed for every time instance during one gait cycle and thus be removed from the acceleration signal. Consequently, the resulting acceleration signal can be integrated twice to get the displacement of the foot. This approach is commonly known as strap-down integration. From the resulting displacement and orientation sequence, both foot kinematic parameters and spatio-temporal parameters can be computed. More details on the implementation of the algorithm as well as the final parameter computation can be found in chapter 5.

**Machine learning** The last type of methods to determine spatio-temporal and kinematic parameters found in literature are machine learning approaches. In machine learning, algorithms are trained to determine a target variable \( y \) using a set of \( d \) features \( \vec{x} = (x_1, x_2 \ldots x_d)^T \). In general, two different kinds of problems can be solved using machine learning: classification and regression (Figure 16). These two problems differ in the target variable \( y \). While in a classification problem, \( y \) is a discrete output from a predefined set of classes (\( y \in y_1, y_2 \ldots y_N \) for \( N \) classes), \( y \) is a continuous, real-valued number in a regression problem (\( y \in \mathbb{R} \)).

Both of these techniques have been applied to IMU-based running analysis. Classifiers, which were trained with features from IMU data, were for example used to determine the terrain subjects are running in [137, 138] or whether runners are forefoot, midfoot, or rearfoot strikers [139].

However, the computation of spatio-temporal or kinematic parameters from IMU data is a regression task. For this purpose different algorithms
Figure 16: Visualization of difference between classification and regression task in machine learning. (a) For the two-class classification problem, $y$ can only be either 0 or 1. A decision boundary can be trained to differentiate the two classes. (b) For the regression, problem, $y$ is continuous. The plot depicts two polynomial functions of degree $p = 1$ and degree $p = 2$, which were fit to the data.
exist. One way to solve a regression problem is to fit a polynomial function of degree $p$ to the training data consisting of a set of observations (see Figure 16b). The features $\mathbf{x}$ are usually handcrafted from the IMU signal. Gradl et al. [P9] used this kind of methodology to determine running velocity from accelerometer data. They defined a single acceleration feature during swing phase and trained a polynomial regression function of degree $p = 2$ to compute running velocity.

Another way of realizing a regression function is to train a neural network with one output neuron, which represents the target value. Neural networks are currently very popular due to the success of deep learning especially in image and speech recognition in recent years. In comparison to traditional approaches, no feature handcrafting is necessary to train the regression. The IMU data can be fed without any preprocessing into the network. During the training process, the networks learn the best feature representation of the IMU data on their own. Such a methodology was used by Hannink et al. [140, 141], who computed both kinematic as well as spatio-temporal parameters in walking gait. Additionally, Derie et al. [96] showed, that kinetic parameters can be computed using deep learning regression.

### 3.2 Big Data analysis using data of wearables

In comparison to IMU-based endurance running analysis, the research branch of Big Data analysis using data of wearables is rather new. Obviously, the wearables first had to be developed, before large databases could be compiled which allow for Big Data analysis. Within this section, I want to introduce the literature about the process of Big Data studies, give an overview of commercial and scientifically used databases as well as summarize applications, which are driven by Big Data.

#### 3.2.1 Overview of Big Data process

Working with large-scale data sets, which were recorded using wearable and portable sensors, is challenging. Hicks et al. [30] summarized the challenges and gave suggestions on best practices when working with such data. Even though their work mainly targeted the health research community, their suggestions can easily be transferred to the sports
State of the art in general and the analysis of large-scale endurance running data sets in particular. In the following, I will summarize the challenges they outlined and the process they introduced to guarantee high-quality Big Data research. Within this section I will transfer the examples they provided from the Health domain to the endurance running domain.

Traditionally, studies in sports science were controlled and supervised. The study design clearly described what data had to be captured to answer a specific research question. Furthermore, gold standard systems were used to acquire ground truth data. In comparison to these traditional approaches, the majority of data acquired using wearable devices is recorded for non-scientific purposes in an unsupervised manner. Typically, the motives of consumers of commercial wearables are different than the objectives of research. They use the devices to monitor their training process or share their progress with friends rather than to participate in a study or acquire data for answering research questions. This uncontrolled data acquisition causes some challenges for usage in a scientific context. First of all, the ground truth for evaluating certain research questions is missing. While in exercise physiology studies, ground truth parameters like VO$_2$ max are used to monitor training progress, such parameters are not available for data sets from wearables. Another challenge is that the data acquired is often unstructured. Some consumers use the wearables intentionally for different purposes by monitoring unrelated activities, while others mistakenly tag activities incorrectly (bike activity as running activity). Also, missing data can be a problem. If the training process of a runner shall be evaluated, it would be beneficial to have recordings of all workouts during a defined training period. If a runner forgets to use the wearable device during single workouts, the data on fitness platforms might not describe the full training process. Lastly, dealing with sensor inaccuracies is challenging as well. Many platforms receive data from different wearable devices, which use different sensor settings and preprocessing steps before sending the data to fitness platforms.

To deal with those challenges, Hicks et al. [30] suggested a process for scientific Big Data projects, which addresses the described challenges along the way. This process is depicted in Figure 17.

**Identification of research question and data set** As a first step, a research question has to be identified, which can be answered using a large-scale wearables data set. In this regard, access to data sets, which are often from commercial fitness platforms (see section 3.2.2), is a big
challenge. If access is granted, it makes sense to ask a specific research question rather than an open research question. The advantage of a specific research question is that the authors can also assess, whether the data of wearables is suitable and sufficient to draw conclusions. In open research questions, missing data might become a problem at a later point in time.

While analyzing the suitability of the data, the ground truth is an important aspect to consider. While hand-labeling these large amounts of data is often not an option, there are ways to create ground truth. The first and easiest for the researcher is that the data contains ground truth. Different fitness platforms ask users to tag specific workouts (e.g. race, performance assessment run) or enter context information (e.g. Likert scale for exhaustion). Besides, some platforms automatically add weather information. The second way to label data is to combine the data set with an external data source. For example, Althoff et al. [142] combined a step count data set of more than 700k people worldwide with obesity information of the World Health Organization (WHO). By aggregating the step counts on a country level, they could correlate the inequality in step counts per day with the country level obesity data of the WHO. Lastly, ground truth can also be created from the data itself. In this case,
specific metrics, which can be computed from the data and were validated in literature before, can act as ground truth.

**Preparation of data**  As described before, the data sets are often unstructured. Due to this reason, data preparation and improvement of data quality is an important step to draw scientifically valid conclusions. As a first step, Hicks et al. [30] suggest addressing sensor quality. The sensors included in the study shall be characterized with respect to data accuracy and - if possible - users with inaccurate devices shall be excluded. Apart from sensor accuracy, other inclusion criterion should be defined as well. Evaluating the training process of runners only makes sense, if a minimum number of workouts is available to evaluate the process. By defining a minimum number of workouts as an inclusion criteria, only subjects that regularly upload data are included in the study. Lastly, the data has to be cleaned by removing outliers. On the one hand, outliers can be removed by empirically defining thresholds. For example, the 10 km performance of recreational athletes is unlikely to be faster than the current 10 km world record. Thus, maximum average running velocities for different workout distances can be defined to remove unreasonable data. Besides, it makes sense to use IQR or similar filters to also clean data with respect to its underlying parameter distributions.

To prove that the data preparation was successful, a final validity check should be performed. By reproducing known values and trends using the data of wearables, trust in the data and consequently also in the conclusions can be increased. As an example, Althoff et al. [142] reproduced relationships between age, gender, weight status, and activity on a country level before analyzing the data.

**Data Analysis**  The final data analysis also changes for Big Data projects. The processing of large amounts of data with just one single notebook or desktop PC becomes unfeasible due to a lack of memory and processing power. However, cloud computing providers like Amazon Web Services (AWS) or Microsoft Azure with large storage capacities and high processing power can be utilized to speed up processing times. Once the data is processed, the conclusions have to be checked for robustness. In this context, Hicks et al. [30] suggest to rather analyze effect sizes of the statistical tests than the p-values for statistical significance. The large sample sizes in Big Data projects cause small p-values and indicate statistical significance, even though the effects might be really small.
Sharing of Results Once the robustness of the conclusions is validated, the results can be published. While the general procedure is similar to traditional publication processes, differences regarding ethics approval exist. Typically, ethics approval is required for studies involving human subjects. The studies have to be in accordance with the Declaration of Helsinki [143], which requires informed consent of all subjects participating in the study. Due to the fact, that sports science studies involve human subjects, many journals in this field require ethics approval. However, it is not feasible to get the consent of all subjects who used wearable devices and uploaded the data to a fitness platform. Due to this reason, the Declaration of Helsinki [143] states the following (point 32):

"There may be exceptional situations where consent would be impossible or impracticable to obtain for such research. In such situations the research may be done only after consideration and approval of a research ethics committee."

Thus, especially retrospective studies using data from fitness platforms can be conducted after approval by an ethics committee [144]. In the case of the study presented in chapter 6, the local ethics committee raised no objections to the study, because the data was anonymized before I received it for processing.

3.2.2 Databases

Nowadays, many commercial fitness platforms exist that runners can use to upload their workout data to (Figure 18). Most of these platforms provide similar functionalities, with training diaries being the most common use case. Those training diaries are capable of storing and visualizing GPS and heart rate data for the consumers. Some platforms also store additional information like heart rate variability data (HRV4training) or running power (Stryd). Besides, data can be shared between platforms. Especially the profiles of users from companies selling wearables (e.g. Polar, Garmin, Fitbit) can be linked to other platforms such that the data is automatically synchronized between the platforms after data upload.

While all the platforms contain large amounts of interesting data, access for researchers is limited. Companies often keep the data for themselves. On the one hand, they are afraid of competitors being able to use results based on their data. On the other hand, training platforms often contain personal data of users, which may not be shared due to data privacy
Figure 18: Overview of commercial fitness platforms athletes can use for tracking their runs.

Despite these risks, some studies with data from fitness platforms have recently been published, which shows that the companies are becoming more open for research collaborations. Table 3 lists four data sets, which were used for scientific analysis including the adidas Runtastic data set I used for the analysis in chapter 6.

Table 3: Data sets from fitness platforms used in science. Different subsets of the full Strava data set with 31190 runners were used in [145, 146, 147, 148].

<table>
<thead>
<tr>
<th>Platform</th>
<th>Number of users</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strava</td>
<td>31190</td>
<td>[145, 146, 147, 148]</td>
</tr>
<tr>
<td>Polar</td>
<td>14000</td>
<td>[149]</td>
</tr>
<tr>
<td>HRV4Training</td>
<td>2113</td>
<td>[150]</td>
</tr>
<tr>
<td>adidas Runtastic</td>
<td>14773</td>
<td>[P3]</td>
</tr>
</tbody>
</table>

To foster more such collaborations, Hicks et al. [30] published elements of their data-sharing agreement, which helps them approaching companies. These elements define data ownership, the scope of data usage, data access, user anonymity, potential publications, and the licensing of results. By addressing these elements, trust between the partners can be established, which increases the chances of getting access to such data under a special research license.
3.2.3 Applications with data of wearables from fitness platforms

The data sets summarized in Table 3 have been used for realizing different applications. This section summarizes the already realized applications and ideas from literature.

**Race time predictions:** One of the first attempts to predict race time was from Peter Riegel [151], who introduced a formula that can be used to predict the race time based on the performance in former, shorter races. By incorporating additional data not only from former races but also from training, more accurate models can be built using the data from fitness platforms. Altini et al. [150] built a multiple linear regression model to predict the 10 km performance of runners using the data from the HRV4Training platform. In the model, the authors included anthropometric data, heart rate variability at rest, training volume, and speed per training of past workouts as well as the best 10 km performance in the past 3 months. Feely et al. [145] used data from Strava to build an explainable race-time prediction. They used data from 16 weeks of marathon training to build the predictor. They aggregated training statistics on a weekly basis and used a nearest neighbor approach to find athletes with similar training characteristics. The prediction for the race time was finally computed by a similarity-weighted mean of the $k$ nearest neighbors.

**Personalized training plans:** Similarly, Feely et al. [145] also used the Strava data to build an explainable recommender system for personalized training plans. Many fitness platforms offer training plans for recreational athletes. However, these training plans for running are static and usually built to reach a certain target time for a determined distance. They are not able to adapt to the individual responses of runners to training. To solve this problem, Feely et al. [145] extended their race time prediction approach to design a training plan for the next week. As in the race time prediction scenario, they aggregated data on a weekly basis and found the $k$ nearest neighbor with respect to the training characteristics. After predicting the marathon performance, they recommended the training for the next week based on the training of one of the $k$ neighbors with a better marathon performance.

**Pacing recommendations:** Especially recreational runners often face the phenomenon of *Hitting the Wall* during a marathon race [152, 153].
This means that athletes run too fast at the beginning of a marathon, which results in a significant drop in performance at the end of the race. Recommendations on a personal pacing strategy can prevent runners from experiencing this phenomenon. Smyth et al. [146] introduced an indirect way for such a recommendation. Based on Strava data, they computed critical speed values for each runner based on the training records of marathon runners. Critical speed is defined to be the highest intensity in which physiological steady state can be reached [154]. Knowing the critical speed of a runner, the time a runner needs to reach a target distance can be predicted. Smyth et al. [146] showed that recreational runners, who started a marathon race above 94% of their personal critical speed were at higher risk to significantly slow down in the second half of the race. Thus, knowing the critical speed value can help runners to set up a pacing strategy. Besides, Berndsen et al. [148] developed a recommendation system, which is theoretically capable of giving recommendations during the race. Based on a data set of 7931 marathon recordings, they trained several XGBoost regression models for each 500 m mark within the race, which predicted the amount of decrease in speed for the remaining race distance based on features extracted from the race recording up to that point. If the XGBoost model predicted a significant slowdown, they searched for runners with similar race features up to that point in the race and recommended an averaged pacing strategy of runners, which did not slow down for the rest of the race.

Evaluation of endurance running related models: Previously defined endurance running related models can be evaluated using large amounts of wearable running data. Both Smyth et al. [146] as well as Emig and Pelton [149] used either endurance running-related models from literature or self-developed models and evaluated them on large-scale marathon data sets based on wearable data, which also included training records. They both followed the same approach in the evaluation. Based on their models, they were able to predict the marathon performance. By proving an accurate prediction, they argued that the model itself seems to describe the endurance of an athlete well. Smyth et al. [146] evaluated the previously mentioned concept of critical speed [154] using this approach. Emig and Pelton [149] evaluated a universal performance model, which was introduced by the same research group around Thorsten Emig [155]. Both research groups could prove that the underlying assumptions of their models were valid and that the models generalize well to a large cohort of endurance runners.
Part III

Contributions
4 Paper 1 - Evaluation of algorithms for IMU data processing

The following chapter is a replicate of the following publication:


Only a few phrases have been adapted to fit the wording within this thesis.

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

4.1 Introduction

Distance running is a very popular sport. Two main reasons for this popularity are simplicity and the health benefit, as running can be done in small and restricted time-frames and does not require a specific location. Apart from sports gear, no equipment is needed. Moreover, running improves health. Studies have shown that aerobic endurance training like running can reduce blood pressure [156] and that moderate running twice a week (>51 min or six miles) reduces overall mortality risk and the occurrence of cardiovascular diseases [157]. However, overtraining can also lead to a higher risk of injury of the lower extremities for distance runners [158].

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

4.1.1 Literature Review

The predominant approach to tracking average velocity and total distance during running is the Global Positioning System (GPS). Smartphones or even smartwatches comprise a GPS chip, which allows a satellite-based localization of a runner. By tracking the runner’s absolute position over a complete run and using a solution to the second geodetic problem [159], the distance of a run can be measured. By incorporating the sampling frequency of the GPS module, a continuous time series of velocity values for the run can be computed. Thus, GPS delivers a time series of velocity, the cumulative distance, and the localization of the running track. From these data, the average velocity and the total distance can be extracted. The drawbacks of GPS are the additional gear (smartwatch,
4.1 Introduction

smartphone), the high energy demand, and the restriction to outdoor use.

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

In biomechanics, the relationships between stride frequency, stride length, running velocity, and body height was investigated [38]. The results indicated that with increasing running velocity, stride frequency and stride length increase. Thus, increasing running velocity is an interaction of increasing stride length and stride frequency [38]. Stride length itself depends on body height and can be expressed as a relative stride length. From these relationships, a generic model relating running velocity and stride length on the basis of the stride frequency can be deduced. The general idea behind this approach is the inverse correlation between velocity and stride time (the higher the velocity, the shorter the stride time). Thus, in order to estimate the stride length, only the stride time has to be distinguished by segmenting the data into single strides. An average velocity of the stride can then be calculated using the stride length and the measured stride duration.

Recently, Gradl et al. [P9] proposed an algorithm that uses quadratic regression to compute the velocity of movements. The velocity was evaluated during running, as well as other movements and showed a relative error of $6.9 \pm 5.5\%$. The proposed algorithm is solely based on foot acceleration. Single strides are segmented from the data stream. Afterwards, the acceleration signal of all axes is integrated prior to the
initial ground contact. Finally, the resulting integral value is converted to a velocity value using a quadratic regression model.

Another method to compute velocity and stride length values from IMU signals is to reconstruct the trajectory of the sensor in the course of a stride. This method is heavily used for gait analysis for geriatric patients [6, 160, 161] or in inertial navigation scenarios [162, 163]. For trajectory reconstruction, sensor fusion techniques must be applied to both the accelerometer and the gyroscope. Several fusion algorithms to cope with this task exist. Bailey et al. [92] and Foxlin et al. [164] used extended Kalman filters to compute the trajectory from the acceleration and angular rate signals, while Rampp et al. [6] applied a linear dedrifting technique. Both algorithms rely on a zero-velocity update during the stance phase for the initialization of the orientation. The literature shows that this approach works well while analyzing walking [6], but it was not evaluated for free running. Bailey et al. [92] applied their approach to treadmill running and showed a good accuracy of $0.03 \pm 0.2$ m/s. However, they evaluated neither the velocity nor the stride length in a free running scenario.

Deep learning techniques also show good results in IMU-based classification and regression tasks [165, 166]. Hannink et al. [140] showed that deep convolutional neural network regression outperforms traditional stride length estimation in geriatric gait analysis. They trained a network with two convolutional layers, which was fed with the 6D IMU raw data of a stride. The output layer had a single node and provided an estimate of stride length.

### 4.1.2 Contribution

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

4.2 Methods

4.2.1 Data Collection

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

4.2.1.1 Lab Study

In the lab study, data from 27 recreational runners (21 male, 6 female) were recorded. The dataset included runners with different strike types. Six of the subjects were forefoot/midfoot runners, and 21 subjects were rearfoot runners. The classification of the strike type was based on the definitions of Altman et al. [46]. Further anthropometric data can be found in Table 4. Before data acquisition, all subjects were informed about the related risks and gave written consent to participate in the study and for the collected data to be published.

Table 4: Anthropometric data of subjects participating in the lab study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ± Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>24.9 ± 2.4</td>
</tr>
<tr>
<td>Shoe size (U.S.)</td>
<td>9.3 ± 1.4</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>178.6 ± 8.0</td>
</tr>
</tbody>
</table>

The subjects were equipped with running shoes in matching sizes (Response Cushion 21, Adidas AG, Herzogenaurach, Germany), as depicted in Figure 19a. This model had a cavity in the right shoe midsole for the placement of a sensor. We cut another cavity of the same size at the same
location into the left shoe midsole to be able to acquire data from both the left and the right shoe in order to record more data for the training and evaluation of the algorithms. The specific IMU we used was the miPod sensor [167]. The accelerometer of the sensor was configured with a range of $\pm 16 \, \text{g}$ and the gyroscope with a range of $\pm 2000 \, \text{g/s}$, and data were sampled with a frequency of $f_s = 200 \, \text{Hz}$ and a resolution of 16 bit. Before each data acquisition, the IMUs were calibrated using the calibration procedure introduced by Ferraris et al. [120]. Figure 19a depicts the orientation of the sensor in the sole of the running shoe: $x$ points in the lateral direction, $y$ in the dorsoventral direction, and $z$ in the craniocaudal direction.

As the gold standard for velocity and stride length, we used a motion capture system (Vicon Motion Systems Inc., Oxford, U.K.) with 16 infrared cameras and recorded data with a sampling rate of $f_s = 200 \, \text{Hz}$. A submodel of the marker setup introduced by Michel et al. [168] containing six markers on each shoe (see Figure 19a) was used. The marker on the heel (for rearfoot runners) and the lateral sided toe marker (for forefoot/midfoot runners) were used to extract strides. Depending on the strike type, minima in the trajectory of the corresponding markers were used to label initial ground contacts [169]. The IMUs and the motion tracking system were synchronized using a wireless trigger [170], which was connected to light barriers (S40 Series, Datalogic, Bologna, Italy). The light barriers triggered the start and the end of the recording for each trial in both systems. Using the described synchronization technique, we were able to match strides in the motion capture gold standard data to strides in the IMU signal.

The subjects were asked to run various trials with different velocities in the range of 2–6 m/s. We defined these velocity ranges to cover a wide range of relevant running velocities. As the capture volume was restricted to 6 m, and the stride length varied depending on the running velocity, different numbers of strides were recorded for the different running velocities. We recorded five additional trials for the two high velocity ranges to increase the number of captured strides. The velocity ranges and number of trials recorded can be found in Table 5. The subjects were asked to accelerate before and keep the pace within the capture volume. We measured the velocity at the beginning of the motion capture system volume using the above-mentioned light barriers used for synchronization. The velocity measured by the light barriers was used to ensure that a sufficient number of trials were recorded within each velocity range, for each subject. If
necessary, the subjects were instructed to run faster or slower in order to ensure the defined number of trials in each velocity range. The ground truth value for each stride’s velocity \( v_{\text{ref}} \) was computed from the motion capture reference as:

\[
v_{\text{ref}} = \frac{d_{\text{ref}}}{t_{\text{ref}}} = \frac{d_{\text{ref}} \cdot f_s}{N_{\text{stride}}} \quad (6)
\]

where \( d_{\text{ref}} \) is the stride length obtained by the difference of the positional data obtained by the motion capture system between two consecutive initial ground contacts, \( t_{\text{ref}} \) the corresponding reference stride time, \( N_{\text{stride}} \) the number of samples in between two consecutive initial ground contacts, and \( f_s \) the sampling rate. Figure 19b illustrates the setup and running path of the subjects during the lab data recording. Overall, 2377 strides were recorded during the lab study for the evaluation of the algorithms.

Table 5: Number of trials and recorded strides per velocity range in the lab study.

<table>
<thead>
<tr>
<th>Velocity Range</th>
<th># of Trials</th>
<th># of Strides</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–3 m/s</td>
<td>10</td>
<td>921</td>
</tr>
<tr>
<td>3–4 m/s</td>
<td>10</td>
<td>558</td>
</tr>
<tr>
<td>4–5 m/s</td>
<td>15</td>
<td>544</td>
</tr>
<tr>
<td>5–6 m/s</td>
<td>15</td>
<td>354</td>
</tr>
</tbody>
</table>

### 4.2.1.2 Field Study

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

The equipment (IMUs, shoes) and settings were the same as described in the lab study (Section 4.2.1.1) to enable a direct comparison of the results. The subjects participating in the field study were not part of the lab study.
Additionally, we recorded GPS data using a smartphone (Galaxy S8, Samsung Inc., Seoul, South Korea) and the fitness application Strava (Strava Inc., San Francisco, CA, USA), which is, with 136 million uploaded runs in 2017, one of the most popular fitness apps worldwide [171]. It also has the capability to export the GPS track in the GPX-format [172] allowing for a computation of the distance of the running track. The accuracy of the implemented algorithms can be compared to the GPS data for the total distance of the run. We used the great circle distance to compute the total distance of the GPS measurements [173]. Our computed total distance from the exported GPX file matched the distance that Strava provided via its services. Thus, we could compare the accuracy of the different algorithms to state-of-the-art running platform distance measurements.

4.2.2 Algorithms

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

4.2.2.1 Stride Segmentation

The first step in the IMU signal processing for velocity and distance calculation was the stride segmentation. In this step, single strides were extracted from the continuous IMU data stream with a threshold-based algorithm. Common algorithms use the distinct peaks in the acceleration signal in the dorsoventral direction $a_y[n]$ during initial ground contact to mark the beginning of a stride [22]. We enhanced this idea and used the beginning of the distinctive peak to mark the beginning of a stride. This procedure is valid, as the ground already exerts a force to the IMU at the time instance of the peak in the acceleration signal. Using the peak itself would mean to mark a point in time that is part of the ground contact. To find the sample before the acceleration peak, we first differentiated the acceleration signal in the dorsoventral direction $a_y[n]$ and consecutively squared the resulting value to obtain a signal $H[n]$ with amplified peak values.

$$H[n] = (a_y[n] - a_y[n-1])^2$$

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

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

In this equation, $n_{ZC}$ corresponds to the index of the zero crossing marking the start of the deceleration. If the integral value $S[n_{IC}]$ exceeds an empirically-set threshold (empirical threshold: $S[n_{IC}] < -3 \left( \frac{m}{s} \right)$), a swing phase is detected, and thus, the index of the first non-zero value
before the acceleration peak is labeled as an initial ground contact. The described stride segmentation is depicted in Figure 20.

![Figure 20: Example for the stride segmentation. The plot shows the acceleration signal in the dorsoventral direction \( a_y[n] \), the detected initial ground contact \( n_{IC} \), and the beginning of the swing phase (zero crossing \( n_{ZC} \)) to confirm the stride candidate. The marked area depicts the integration area for the swing phase detection.](image)

### 4.2.2.2 Stride Time

Cavanagh et al. described the relationship between running velocity, stride length, and stride frequency [38]. Stride frequency is an inverse measure of the stride time and describes the number of strides per minute. They showed that runners can increase their running velocity either by increasing their stride length or by increasing their stride frequency, thus decreasing the stride time. For lower velocities, runners tend to increase the stride length, while for higher velocities, they tend to increase the stride frequency. Thus, both the stride time and stride length have no linear dependency on running velocity. Furthermore, it has to be noted that runners control their velocity individually. The stride length and therefore the velocity also depend on other parameters like the gender and the height of the runner. Male runners show greater stride lengths compared to female runners. The stride length increases with the body height [38].

We used these biomechanical relations to build an algorithm that estimates stride length and velocity. Cavanagh et al. [38] provided averaged values for the non-linear correlation between the stride time \( t_{stride} \) and a relative stride length \( d_{stride,rel} \), which is calculated by dividing the absolute stride length \( d_{stride} \) by the runner’s height \( h \). We looked for further publications describing this relationship and came up with two step functions for males and females that discretized the underlying non-linear relationship between stride time and stride length for each
4.2 Methods

The definition of the functions for both male and female runners can be found in Table 6.

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

$$t_{\text{stride}} = \frac{N_{\text{stride}}}{f_s} = \frac{(n+1)IC - nIC}{f_s} \tag{9}$$

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

$$d_{\text{stride}} = h \cdot d_{\text{stride,rel}} \tag{10}$$

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

$$v_{\text{stride}} = \frac{d_{\text{stride}}}{t_{\text{stride}}} \tag{11}$$

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

4.2.2.3 Acceleration

The Acceleration method introduced in [P9] uses only acceleration data for step segmentation and the computation of stride length and stride velocity. The method correlates the velocity of the foot (and thus, the subject) with the acceleration during the swing phase of the foot. It consists of three different algorithmic steps: (1) a continuous calculation of an integration value with a strong correlation to the movement velocity, (2) a stride segmentation based on initial ground contacts to determine the swing phase of the foot, and (3) a regression model to translate the continuous integration value to the velocity value.

We used the step segmentation algorithm from the cited publication due its applicability to movements other than running [P9]. The inputs to
Table 6: Definition of step functions for the Stride time algorithm for the relative stride length $d_{\text{stride,rel}}[t_{\text{stride}}]$ for (a) male and (b) female runners.

<table>
<thead>
<tr>
<th>$t_{\text{stride}}$ (s)</th>
<th>$d_{\text{stride,rel}}$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.800 &lt; t_{\text{stride}}$</td>
<td>0.830</td>
<td>[38, 174]</td>
</tr>
<tr>
<td>$0.748 &lt; t_{\text{stride}} \leq 0.800$</td>
<td>1.080</td>
<td>[174, 175]</td>
</tr>
<tr>
<td>$0.720 &lt; t_{\text{stride}} \leq 0.748$</td>
<td>1.260</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.713 &lt; t_{\text{stride}} \leq 0.720$</td>
<td>1.330</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.706 &lt; t_{\text{stride}} \leq 0.713$</td>
<td>1.410</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.698 &lt; t_{\text{stride}} \leq 0.706$</td>
<td>1.490</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.694 &lt; t_{\text{stride}} \leq 0.698$</td>
<td>1.590</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.687 &lt; t_{\text{stride}} \leq 0.694$</td>
<td>1.740</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.678 &lt; t_{\text{stride}} \leq 0.687$</td>
<td>1.880</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.664 &lt; t_{\text{stride}} \leq 0.678$</td>
<td>1.960</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.649 &lt; t_{\text{stride}} \leq 0.664$</td>
<td>2.015</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.500 &lt; t_{\text{stride}} \leq 0.649$</td>
<td>2.060</td>
<td>[38]</td>
</tr>
<tr>
<td>$t_{\text{stride}} \leq 0.500$</td>
<td>2.170</td>
<td>[176]</td>
</tr>
</tbody>
</table>

(a) Male

<table>
<thead>
<tr>
<th>$t_{\text{stride}}$ (s)</th>
<th>$d_{\text{stride,rel}}$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.800 &lt; t_{\text{stride}}$</td>
<td>0.826</td>
<td>[174, 177]</td>
</tr>
<tr>
<td>$0.735 &lt; t_{\text{stride}} \leq 0.800$</td>
<td>1.110</td>
<td>[38, 174, 175]</td>
</tr>
<tr>
<td>$0.720 &lt; t_{\text{stride}} \leq 0.735$</td>
<td>1.260</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.704 &lt; t_{\text{stride}} \leq 0.720$</td>
<td>1.400</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.667 &lt; t_{\text{stride}} \leq 0.704$</td>
<td>1.500</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.607 &lt; t_{\text{stride}} \leq 0.667$</td>
<td>1.720</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.578 &lt; t_{\text{stride}} \leq 0.607$</td>
<td>1.920</td>
<td>[38]</td>
</tr>
<tr>
<td>$0.500 &lt; t_{\text{stride}} \leq 0.578$</td>
<td>2.080</td>
<td>[38]</td>
</tr>
<tr>
<td>$t_{\text{stride}} \leq 0.500$</td>
<td>2.170</td>
<td>[176]</td>
</tr>
</tbody>
</table>

(b) Female
the processing pipeline were the sampled triaxial acceleration signals from the foot sensor $a_x$, $a_y$, and $a_z$. After smoothing the input signals using a sliding window filter, the integration value $\iota$ was calculated as a multi-step absolute averaging across all directional components with:

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

(12)

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

Individual strides were determined in the smoothed dorsoventral acceleration signal $s_y$ using a peak detection process in combination with two knowledge-based parameter thresholds. The goal was to detect and isolate the high impact response of initial ground contact in the smoothed signal [P9]. Each time a valid stride was detected by the stride segmentation algorithm, the average velocity per stride $v_{\text{stride}}$ was determined based on a second degree polynomial regression function:

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

(13)

where the constants $A, B,$ and $C$ are derived during a regression model training phase where known reference velocity observations are matched to velocity integration values using parametric regression analysis. A trained regression model can be observed in Figure 21.

4.2.2.4 Trajectory

Based on the foot trajectory, the stride length and stride velocity can be deduced. The trajectory of the sensor during running can be computed using an extended Kalman filter approach or using dedrifting techniques. We applied dedrifting techniques due to two reasons: Firstly, Bailey et al. [92] showed that the results for the mean step velocity of the two techniques did not differ significantly with respect to accuracy (extended Kalman filter: $0.03 \pm 0.02$ m/s, linear dedrifting: $0.0 \pm 0.03$ m/s). Secondly, the same authors showed in a different article that a sampling rate of more than 250 Hz is required for an extended Kalman filter approach [115]. For embedded use cases (e.g., a smart shoe scenario), low sampling rates are beneficial from an energy perspective. In gait analysis,
4 Evaluation of algorithms for IMU data processing

![Figure 21: Polynomial function of second degree (red line) that relates the velocity integration value \( \iota \) to the reference velocity values \( v_{\text{stride}} \) (grey dots).]

the linear dedrifting technique showed promising results for a lower sampling rate of 200 Hz [6].

Trajectory reconstruction algorithms based on linear dedrifting consist of four steps, as depicted in Figure 22, and have both the triaxial accelerometer and the triaxial gyroscope measurements as an input. In the following paragraphs, the four algorithmic steps will be explained in detail. Orientation is computed by integrating the gyroscope measurements, and the position is obtained by integrating the accelerometer measurements.

![Figure 22: The four steps of the algorithm for the trajectory reconstruction based on linear dedrifting.]

**Midstance detection**: A common problem with computing the trajectory from IMU measurements is the drift of the sensors introduced by noise in the acceleration and the angular rate measurements. This drift is limited by using zero velocity updates [131]. The idea behind these updates is to reinitialize the position and the orientation of the sensor for every stride. By applying that technique, absolute position in space is lost; however, the individual stride parameters can be computed more accurately. The reason for the higher accuracy lies in the integration of
shorter durations and thus a smaller accumulated error. The point in
time for the reinitialization of the stride values originates from gait analysis
and is the midstance phase during a stride cycle. At this point in time,
the foot has its lowest velocity, and the orientation of the foot is known,
because during midstance in gait, the foot is expected to be flat on the
ground. Thus, it can be assumed that the orientation of the sensor can
be computed statically using the acceleration measurement. This allows
the initialization of the position and velocity to zero and the orientation
with respect to gravity. To find midstance, we computed the L2-norm of
the gyroscope signal after initial ground contact [131] in a 250-ms time
interval. The duration of this time interval is the average time of the
stance phase for velocities up to 6 m/s [178]. Hereafter, the trajectory
reconstruction will be performed on strides segmented from midstance
to midstance.

**Orientation estimation**: After initializing the orientation based on
the accelerometer measurement during midstance, the orientation of
the sensor was computed using the gyroscope measurements. This step
is necessary to calculate the orientation of the sensor so that gravity
can be removed, which is an essential step for the computation of the
position in space from the acceleration signal. For the orientation comu-
putation, we used the same quaternion integration approach as described
by Rampp et al. [6].

**Gravity removal**: After the orientation estimation, gravity was removed.
Without this removal, the gravitational acceleration of 9.81 m/s² would
be integrated additionally into the acceleration caused by running, which
would lead to a large error over the duration of a stride. To remove gravity,
we used the orientation of the sensor obtained by the gyroscope integra-
tion to rotate the acceleration measured in the sensor coordinate system
to the world coordinate system. In the world coordinate system, we sub-
tracted the gravitational acceleration from the measured acceleration.

**Dedrified integration**: The last step to come up with the full trajectory
of the stride was to compute the position of the sensor by a double integra-
tion of the gravity removed acceleration. The first integration computed
the velocity of the sensor over time, followed by the second integration,
which resulted in the position of the sensor over time. Despite the gravity
removal, there was still noise in the acceleration signal, causing drift
in the results. This drift was reduced by dedrifting the velocity signal
obtained after the first integration. The core idea behind dedrifting is the
fact that we assume the velocity to be zero during midstance. For every stride, we fit a linear function in the velocity signal for all three directions, which was determined by the first and last velocity value of the stride. To dedrift the velocity signal, we subtracted the linear function from the integrated velocity signal, which enforced the velocity to be zero for both the first and the second midstance. This process is depicted in Figure 23.

![Velocity and dedrifted velocity graphs](image)

Figure 23: Visualization of the dedrifting method that ensures that the velocity during the second midstance is zero. (a) Velocity signal before dedrifting. The grey dotted linear function is fit between the first and last point of the stride (midstance). (b) Velocity signal after dedrifting.

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

### 4.2.2.5 Deep Learning

After outperforming conventional methods in various other fields like speech recognition, visual object recognition, and object detection [179], the methodology of deep learning started to become more and more popular for IMU data processing. Hannink et al. [140] introduced a deep convolutional regression network for calculating the stride length from raw IMU data in geriatric patients. The network learned a model for stride length regression based on raw IMU data without any domain knowledge. In this work, we used an adapted architecture for the stride
length computation in running gait, which is depicted in Figure 24. It consisted of two convolutional layers, two max pooling layers, one flattening layer, and two fully-connected layers. For the implementation of the architecture, we used Keras [180] with a TensorFlow backend [181].

**Figure 24:** Architecture of the convolutional neural network for stride length regression based on the raw 6D-IMU signal. For the first convolutional layer, we used $N_1 = 32$ filter kernels of kernel length $K_1 = 30$. The second convolutional layer consisted of $N_2 = 16$ filter kernels of kernel length $K_2 = 15$. The first fully-connected layer had $M_1 = 128$ outputs that served as input to the second fully-connected layer, which had only a $M_2 = 1$ output. This output represented the computed stride length.

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

$$O(j) = \text{ReLU}\left(\sum_{c=0}^{6} (\Phi_c(j) \times x_c + b_c(j))\right)$$  \hspace{1cm} (14)$$

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

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

\[ F_j = \text{ReLU} \left( \sum_{k=0}^{N_f-1} (w_{k,j} \cdot \varphi_k + b_{k,j}) \right) \]  

(15)

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

To prevent overfitting, we also added a dropout layer to our network [182]. The dropout layer was stacked between the two fully-connected layers and dropped 30% of the neurons. During training, we fed the data into the network in five epochs with a batch size of 16. We trained the network both for the stride length and for the velocity. The network with the stride length as the output outperformed the velocity approach and was therefore used for the evaluation in this publication. Thus, the velocity $v_{\text{stride}}$ for the Deep Learning approach was computed by dividing the stride length $d_{\text{stride}}$ obtained from the neural network by the stride time $t_{\text{stride}}$ obtained from the stride segmentation.
4.2.3 Evaluation

4.2.3.1 Lab Study

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

For the evaluation of the Acceleration and Deep Learning algorithms, we used leave-one-subject-out cross-validation to prevent overfitted results. We also show Bland–Altman plots [183] to visualize the results.

4.2.3.2 Field Study

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

4.3 Results

4.3.1 Lab Study

Table 7 depicts the mean errors and standard deviations for both stride velocity and stride length of the four different algorithms for the lab study dataset. The results were averaged over all strides in the lab study dataset. The results show that the Trajectory algorithm performed best considering both the ME ± Std and the MAE.

Figure 25 shows the results of the stride length for the different velocity ranges. The MEs of the Deep Learning algorithm increased with higher velocities. The Trajectory showed lower MEs for the three slower velocity ranges than for the highest velocity range. The Acceleration algorithm
Table 7: Mean error (ME) and standard deviations (Std), mean percentage error (MAPE), and mean absolute error (MAE) of average velocity per stride $v_{\text{stride}}$ and stride length $d_{\text{stride}}$ of the four algorithms for the lab study dataset.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Error measure</th>
<th>$v_{\text{stride}}$ [m/s]</th>
<th>$d_{\text{stride}}$ [cm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stride time</td>
<td>ME ± Std</td>
<td>0.209 ± 0.782</td>
<td>17.7 ± 57.3</td>
</tr>
<tr>
<td></td>
<td>MAPE (%)</td>
<td>17.2</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>0.622</td>
<td>45.2</td>
</tr>
<tr>
<td>Acceleration</td>
<td>ME ± Std</td>
<td>0.005 ± 0.350</td>
<td>−0.5 ± 25.6</td>
</tr>
<tr>
<td></td>
<td>MAPE (%)</td>
<td>7.7</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>0.272</td>
<td>19.9</td>
</tr>
<tr>
<td>Trajectory</td>
<td>ME ± Std</td>
<td>0.028 ± 0.252</td>
<td>2.00 ± 14.1</td>
</tr>
<tr>
<td></td>
<td>MAPE (%)</td>
<td>3.5</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>0.133</td>
<td>7.6</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>ME ± Std</td>
<td>0.055 ± 0.285</td>
<td>2.5 ± 20.1</td>
</tr>
<tr>
<td></td>
<td>MAPE (%)</td>
<td>5.9</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>0.216</td>
<td>15.3</td>
</tr>
</tbody>
</table>
showed small errors from 3–5 m/s. Its performance dropped for the outer velocity ranges from 2–3 m/s and from 5–6 m/s. The *Stride time* algorithm worked well for the velocity range of 2–3 m/s and 5–6 m/s; however, it showed large errors of more than 40 cm for the other velocity ranges.

Figure 25: Mean error of stride length of the four different algorithms for the different velocity ranges the subjects ran in the lab study.

Figure 26 shows the Bland–Altman plots for both the stride length and the average velocity per stride for the lab study dataset. The results are color coded into the velocity ranges presented in Table 5. The *Trajectory* algorithm performed well for velocities up to 5 m/s. For the high velocity range, larger errors could be observed. The *Stride time* algorithm performed worst and showed a linear error distribution in the Bland–Altman plots. In the *Acceleration*, *Trajectory*, and *Deep Learning* plots for stride length, we see the samples of the different velocity ranges overlapping. This overlap is not visible in the velocity plots.

### 4.3.2 Field Study

Figure 27 shows the MAE of the total running distance for the field study dataset, both for the algorithms and GPS-based estimation. The figure indicates that the *Trajectory* algorithm performed best with a MAE of
Evaluation of algorithms for IMU data processing

(a) Stride time; stride length

(b) Stride time; velocity

(c) Acceleration; stride length

(d) Acceleration; velocity
Figure 26: Bland–Altman plots for stride length and velocity for the four algorithms. Each row contains the metrics for one algorithm. The individual samples are color coded depending on the velocity bin of the sample: 2–3 m/s blue, 3–4 m/s red, 4–5 m/s green, 5–6 m/s purple. The dotted-dashed horizontal lines depict the mean error and the dotted horizontal line the 95% confidence interval.
94.0 m. This error was comparable to the error of the GPS-based estimate (82.1 m).

4.4 Discussion

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

4.4.1 Comparison to Existing Literature

Different papers already evaluated the stride length or the velocity of single strides. Three of these papers are listed in Table 8. These three publications used similar approaches to ours: Bailey et al. [92] used a trajectory approach using a linear de-drifting technique; Gradl et al. [P9] used the described acceleration approach; and Hannink et al. [140] a DCNN approach.

With respect to the standard deviation, the results of the trajectory implementation of Bailey et al. [92] are better than our results (Table 7). They also evaluated 1800 strides; however, these strides only originated from
4.4 Discussion

Table 8: Results of other publications related to stride length and velocity calculation.

<table>
<thead>
<tr>
<th></th>
<th>[92]</th>
<th>[P9]</th>
<th>[140]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gait type</td>
<td>Running</td>
<td>Running</td>
<td>Walking</td>
</tr>
<tr>
<td># Subjects</td>
<td>5</td>
<td>9</td>
<td>101</td>
</tr>
<tr>
<td># Strides</td>
<td>1800</td>
<td>795</td>
<td>~1392</td>
</tr>
<tr>
<td>Parameter</td>
<td>Velocity</td>
<td>Velocity</td>
<td>Stride length</td>
</tr>
<tr>
<td>Error Measure</td>
<td>ME</td>
<td>MAPE</td>
<td>ME</td>
</tr>
<tr>
<td>Results</td>
<td>0.04 ± 0.03 m/s</td>
<td>6.9 ± 5.5%</td>
<td>0.01 ± 5.37 cm</td>
</tr>
</tbody>
</table>

running velocities ranging from 2.3–3.4 m/s. We also evaluated our results for this velocity range and obtained an error of 0.004 ± 0.107 m/s. We observe that our standard deviation is still higher than the standard deviation reported from Bailey et al. One reason for that might be the higher number of different runners with different running styles who participated in our study. Furthermore, their study was conducted on a treadmill. On a treadmill, the variability of different strides at a given velocity is lower and does not reproduce overground running kinematics [184].

The errors reported by Gradl et al. [P9] were obtained on a smaller database than the one presented in this paper. Thus, our worse results are due to the higher variability in our dataset, which the second degree polynomial could not appropriately approximate.

The results of Hannink et al. [140] were evaluated for gait in geriatric patients. Hence, there is a general difference in the stride patterns, causing differences in the results. Further differences between the setup of our network architecture and study population are listed and discussed in the following section.

4.4.2 Lab Study

In this section, we will discuss the results of the lab study for each algorithm in detail with respect to their advantages/disadvantages and the number of sensors needed for their implementation.
**Stride time**: The *Stride time* algorithm leads to the lowest accuracy for the lab study dataset. Even though stride time and stride length relative to the subject’s height correlate non-linearly, the correlation does not seem to be high enough to compute velocity and stride length accurately. The low correlation is also visible in Figure 28. The gray dots are the relative stride length values obtained from the lab study dataset, and the red line is the step function for male subjects defined in Table 6a. We see that the step function does not approximate the underlying data accurately. The standard deviation of the relative stride length within a certain stride time range (e.g., $0.748 < t_{\text{stride}} \leq 0.800$) of the step function is high. This is due to the fact that velocity is controlled by stride frequency and stride length. The *Stride time* algorithm cannot handle that fact, as it only depends on stride frequency.

![Figure 28](image.png)

Figure 28: Visualization of the correlation between the stride time $t_{\text{stride}}$ and the relative stride length $d_{\text{rel, stride}}$ for male subjects. The light gray dots depict the data obtained from the field study, whereas the red curve and the black dashed lines visualize the step function obtained from literature and implemented in the *Stride time* algorithm.

In the Bland–Altman plots for the stride length metric (Figure 26), the other three algorithms showed overlapping sample clouds. This indicates that people increased their velocity both by increasing their stride length and by decreasing their stride time in higher velocities. The other algorithms are capable of dealing with this effect due to the fact that the sample clouds are separated in the Bland–Altman plots of the velocity metric. This is not observable in the plots for the *Stride time*
algorithm. Thus, the other algorithms can deal better with the velocity control via stride frequency and stride length.

Furthermore, we want to discuss the shape of the Stride time algorithm’s Bland–Altman plots briefly. The long diagonal lines in the plots (Figure 26b) originate from the steps in the step function introduced in Table 6. One line belongs to one stride time range. The small deviations within the diagonals originate from the different body heights. We observed that for some stride time ranges, the gold standard velocity ranged from 2–6 m/s (color coded within one diagonal), showing that the stride time ranges of the step function obtained from the literature do not generalize well. Furthermore, the relative stride lengths presented in Table 6 are averaged over specific study populations. Even if a subject controls its stride frequency in the exact same manner as encoded by the stride time ranges of the step function, the resulting stride length could be incorrect due to an incorrect relative stride length.

Despite the algorithm’s low accuracy, an advantage of the stride time algorithm is that it can be implemented very energy efficiently. In the case of an IMU scenario, only a stride segmentation is necessary to compute the stride time. The stride segmentation presented in this paper only needs the sampling of the acceleration in the dorsoventral direction; thus, a 1D-accelerometer would be sufficient. In fact, strides could be segmented without an IMU using sensors such as piezo-electric switches to detect the ground contact [185].

**Acceleration:** The plot with the ME for the different velocity ranges in Figure 25 shows that the Acceleration algorithm works better for the two velocity ranges from 3–4 m/s and 4–5 m/s. In addition, the Bland–Altman plots in Figure 26 show outliers especially for the highest velocity range for both the stride length and the average velocity. The reason for that can be observed in Figure 21, where we see that the second degree polynomial used to map the velocity integration value $\pi$ to the velocity value approximates the reference data better for the velocity range from 3–5 m/s and especially not well for the highest velocity range. This can be explained by the spread of the underlying data being too large to be represented by the polynomial.

However, the Acceleration algorithm outperforms the Stride time algorithm and shows comparable performance to the Deep Learning algorithm.
for the velocity range of 3–4 m/s. The advantage of the Acceleration algorithm over the better performing Trajectory and slightly better performing Deep Learning algorithm is its energy efficiency. For the computation of the stride length and the velocity, only a triaxial accelerometer needs to be sampled. Sampling only an accelerometer consumes less energy than sampling the gyroscope or sampling both sensors. For example, for the MPU9250 from InvenSense, the supply current needed for sampling only the accelerometer is less than 15% of the current needed for sampling both the accelerometer and the gyroscope [186]. Furthermore, the sampling rate can be further reduced for the Acceleration algorithm [P9].

We also tested the reduction of the sampling rate for the lab study dataset and observed that a reduction to 60 Hz does not affect the accuracy of the algorithm. With such a low sampling rate, the energy consumption can be further reduced. Another advantage of the algorithm is its generalizability and its applicability to other movements like side stepping [P9].

**Foot trajectory:** The Trajectory algorithm performs best for the lab study dataset. Especially for velocities up to 5 m/s, the algorithm achieves a ME of less than 0.012 m for the stride length and 0.014 m/s for the average velocity. For velocities higher than 5 m/s, the accuracy drops. In the Bland–Altman plots (Figure 26f,e), outliers for this velocity range are visible. The zero-velocity update based on the detection of the minimum energy in the gyroscope signal is error prone for such high velocities. The foot has no real zero-velocity phase and is always in motion. Thus, the underlying zero-velocity assumption does not hold. One way how to improve this algorithm is to propose a better solution for the initial condition when applying it to higher running velocities. Future work could evaluate whether a regression model based on the velocity during the swing phase would be a better initial condition.

For the Trajectory algorithm, we were also interested in the applicability of the zero velocity update for the different strike types due to the foot never being flat on the ground for forefoot runners. Hence, we also evaluated the accuracy of the Trajectory algorithm for the different strike types. The violin plots of forefoot and rearfoot runners for the error in the velocity estimation are depicted in Figure 29. The plots show that the MEs do not differ significantly for the two strike types. However, the standard deviation is higher for forefoot runners. The low ME both for the forefoot and the rearfoot strike type can be explained by the fact that we align the foot during the zero velocity phase with gravity. The
higher standard deviation originates in the more dynamic nature of the forefoot running style. Thus, the zero velocity phase cannot be detected accurately, which results in higher errors.

An advantage of the Trajectory algorithm is that it provides more information about the stride than the velocity and the stride length. During the computation of these parameters, the orientation of the shoe in space is also calculated, which allows for a determination of other parameters like the sole angle, which defines the strike pattern or the range of motion in the frontal plane that is associated with pronation [P5]. Furthermore, the algorithm uses solely signal processing and has no training phase, which makes it well applicable to unseen data. This holds for lower velocities and the transition to walking.

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

**Deep learning:** The Deep Learning algorithm produced an ME of less than 0.095 m/s for the velocity and 0.104 m for the stride length for all velocity ranges in the lab study dataset. Compared to Hannink et al. [140],
we reduced both the number of filters in the second convolutional layer and the number of outputs in the fully-connected layer, because the results using the identical structure yielded worse results for our use case. The differences in the architecture are listed in Table 9. Generally, the performance of the DCNN network is worse compared to the results reported in [140].

Table 9: Differences in the study setup and architecture presented by Hannink et al. [140] from our DCNN implementation. $N_2$ is the number of filters in the second layer, $M_1$ the number of output nodes in the first fully connected layer.

<table>
<thead>
<tr>
<th></th>
<th>[140]</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_2$</td>
<td>64</td>
<td>16</td>
</tr>
<tr>
<td>$M_1$</td>
<td>1024</td>
<td>128</td>
</tr>
<tr>
<td># Parameters trained</td>
<td>2,332,385</td>
<td>85,425</td>
</tr>
<tr>
<td>ME ± Std</td>
<td>0.01 ± 5.37 cm</td>
<td>1.3 ± 19.4 cm</td>
</tr>
<tr>
<td>Range stride length</td>
<td>0.14–1.30 m</td>
<td>1.22–4.84 m</td>
</tr>
<tr>
<td>Training samples</td>
<td>~1392</td>
<td>2377</td>
</tr>
</tbody>
</table>

We see that our approach needs less parameters due to the reduction of filters in the second convolutional layer and the smaller output number of the fully-connected layer. However, our results show a larger error. The reason for that might be a larger variation in our training data and the different strike pattern in running. The range of the target parameter of stride length is 3.62 m in the lab study of this work and 1.16 m in the dataset for geriatric patients of Hannink et al. [140]. The strike patterns in running differ significantly for forefoot and rearfoot runners, which also introduces more variation in the input data.

Besides, we observed that during training, the training errors and validation errors still varied after the five training epochs, even though we had more training samples than Hannink et al. [140]. Increasing the number of epochs or batches did not change the varying validation errors. This indicates that the DCNN does not generalize well. Thus, the results might be further improved by incorporating more data samples in the training process of the network.
The embedded implementation of the presented method is a challenge as the DCNN model comprises 85,425 parameters. However, it is still in a range where it can be implemented on a microcontroller. For this method, the acceleration and the gyroscope have to be sampled. This further increases the energy demand compared to the Acceleration approach. Taking computational effort and performance into account, the Acceleration method would be a better trade-off for an embedded implementation.

4.4.3 Field Study

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

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

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

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

In future work, we want to address further parameters that can be computed using inertial measurement units and other sensors located in the sole of a running shoe. Using data acquired by sensors on both feet, it is possible to perform bi-lateral analysis by combining the information of both sensors. Thus, the contribution of the individual lower limbs to the running movement can be further evaluated. Using only data from IMUs within the sole of a running shoe and the Trajectory algorithm, analysis regarding imbalances in stride length, stride time or orientation of the two feet can be conducted. Furthermore, other temporal parameters like flight time or stance time could be computed by adding a toe-off detection. Due to inaccuracies with the toe-off detection in running using only one IMU per foot [129], we plan to also incorporate pressure sensors for toe-off detection into the soles of a running shoe.
4.6 Author contributions

M.Z. conceived of and designed the experiments; M.Z. and M.U. performed the experiments; M.Z., S.G., and U.J. analyzed the data; B.M.E. contributed reagents/materials/analysis tools; M.Z. wrote the paper.
5 Paper 2 - Evaluation of IMU sensor position

The following chapter is a replicate of the following publication:


Only a few phrases have been adapted and one paragraph has been added to fit the wording and scope of this thesis.

Abstract  Wearable sensor technology already has a great impact on the endurance running community. Smartwatches and heart rate monitors are heavily used to evaluate runners’ performance and monitor their training progress. Additionally, foot-mounted inertial measurement units (IMUs) have drawn the attention of sport scientists due to the possibility to monitor biomechanically relevant spatio-temporal parameters outside the laboratory in real-world environments. Researchers developed and investigated algorithms to extract various features using IMU data of different sensor positions on the foot. In this work, we evaluate whether the sensor position of IMUs mounted to running shoes has an impact on the accuracy of different spatio-temporal parameters. We compare both the raw data of the IMUs at different sensor positions as well as the accuracy of six endurance running-related parameters. We contribute a study with 29 subjects wearing running shoes equipped with four IMUs on both the left and the right shoe and a motion capture system as ground truth. The results show that the IMUs measure different raw data depending on their position on the foot and that the accuracy of the spatio-temporal parameters depends on the sensor position. We recommend to integrate IMU sensors in a cavity in the sole of a running shoe under the foot’s arch, because the raw data of this sensor position is best suitable for the reconstruction of the foot trajectory during a stride.
5 Evaluation of IMU sensor position

5.1 Introduction

Wearables have become increasingly important in many fields of our everyday life. Among applications in medicine, workplaces, and many others, the sports domain was one of the early adopters of wearable technology. The reasons for the quick spread of small body-worn sensors in sports was due to the manifold advantages of the technology for athletes, researchers, and the sports industry. Using wearables, athletes can utilize low-cost sensor technologies in order to enhance their performance, prevent injuries, and improve their motivation [19]. Sports research benefits from the fact that wearables allow in field data acquisitions, whereas many studies in the sports domain were traditionally laboratory bound [22]. Furthermore, the sports industry can, on the one hand, offer innovative and more attractive sports products with integrated sensor technology, and, on the other hand, gather consumer data which they can use to improve their products.

One sport where wearable technology already has a great impact is endurance running. Both professional and recreational runners track themselves and use online platforms like adidas Runtastic (Runtastic GmBH, Austria, Pasching) or Strava (Strava, California, San Francisco) to monitor their training progress and performance. Three heavily used sensor technologies in endurance running are GPS trackers, heart rate monitors, and IMUs. While GPS trackers like smartwatches or smartphones are utilized for visualizing the running track and providing real-time feedback on pace and distance [187], heart rate monitors are used to evaluate the physical effort of runs and provide real-time feedback on exercise intensity and training effect [188]. However, these two sensor modalities are not capable of revealing insights into the biomechanics of runners. IMUs are low-cost sensors which consist of 3D-accelerometers measuring linear acceleration as well as 3D-gyroscopes measuring angular velocity. By attaching those sensors to different parts of the human body, various endurance running-related biomechanical parameters can be computed and evaluated.

In this context, IMUs can be used in various ways. Researchers developed body sensor networks with multiple sensors in order to reconstruct the movement of different extremities in a synchronized manner [189]. The advantage of those sensor networks is the holistic evaluation of runners’ movements. However, attaching all the sensors requires a lot of time which recreational runners are often not willing to spend for everyday
runs. That is why researchers also investigated single sensors at specific positions on the human body which can easily and quickly be attached. Apart from placing sensors on the lower back [102, 103], the tibia [98, 99], or the ankle [100], a popular sensor position used in literature is the foot or the running shoe [22, 37, P1, 89, 90, 91]. One reason for the popularity of this sensor position is the amount of different spatio-temporal parameters that can be computed. Falbriard et al. [128] showed that foot-mounted IMUs can be used to accurately segment running strides into their sub-phases (ground contact phase and swing phase) and thus allow computing stride time and ground contact time. Apart from that, researchers developed algorithms to reconstruct the trajectory, that is, orientation and translation, of the foot during a stride. A popular approach for the computation of the trajectory is strap-down integration using a zero-velocity assumption during midstance [92]. From the resulting foot translation, stride length and average stride velocity was calculated with a mean error of 2 cm and 0.03 m/s, respectively [P1]. Additionally, the orientation can be used to compute angular foot kinematic parameters like the sole angle in the sagittal plane or the range of motion in the frontal plane [90, 91].

However, publications using foot-mounted IMUs differ not only in the computed spatio-temporal parameters, but also in the position of the IMU sensors on the running shoes. Shiang et al. [37], Falbriard et al. [90], and Strohrmann et al. [22] mounted the IMU sensors on the instep of the foot on top of the shoelaces, whereas Lederer et al. [89] and Koska et al. [91] mounted the sensors on the heel. Other sensor positions presented in literature are on the lateral side of the running shoe below the ankle [92] and inside the sole of the running shoe [P9, 87].

To our knowledge, the effect of the IMU sensor position on the foot with respect to raw data quality and accuracy of spatio-temporal parameters has not been evaluated yet. Peruzzi et al. [190] evaluated the best possible IMU sensor position for algorithms using zero-velocity updates and found a lateral mounting below the ankle to be the best sensor location. They evaluated the quality of the sensor position using a motion capture system by investigating the motion of the retroreflective markers at different locations. However, they did not include IMU raw data or spatio-temporal parameters to evaluate the sensor position.
We contribute an in-depth analysis by comparing different sensor positions with respect to the raw signals as well as their suitability for the computation of spatio-temporal parameters. We base our evaluation on a study with 29 subjects, which wore running shoes with IMUs attached to the instep, the heel, the lateral side of the foot, and within the cavity of the running shoe. We compare the similarity of IMU raw data at different sensor positions by computing Pearson’s correlation coefficients between the individual sensor positions’ raw data. Besides, we compute and evaluate temporal stride features based on state-of-the-art event detection algorithms and spatial features using a zero-velocity-based strap-down integration algorithm.

5.2 Methods

5.2.1 Definition of spatio-temporal parameters

The temporal parameters we evaluated were stride time $t_{\text{stride}}$ and ground contact time $t_{\text{gc}}$ (Figure 30). One stride was defined by two consecutive initial ground contacts (ICs) of the same foot. The duration between those time instances is the stride time. One stride could be further segmented into ground contact phase and swing phase by finding the toe off (TO) event where the foot leaves the ground. The duration of the ground contact phase is called ground contact time. One further important phase during ground contact time is midstance (MS). It subdivides the ground contact phase into absorption and propulsion phase. We used the MS for the zero-velocity update in the strap-down integration algorithm.

![Figure 30: Visualization of the running gait cycle.](image)
5.2 Methods

The spatial parameters based on the translation of the foot were stride length $d_{\text{stride}}$ and the average stride velocity $v_{\text{stride}}$. The stride length is defined as the translation of the foot during one stride. The average stride velocity can be computed by dividing the stride length by stride time.

We used the sole angle and the range of motion as spatial parameters based on the orientation to evaluate the sensor positions (Figure 31). The sole angle is defined as the angle between the sole of the running shoe and the ground in the sagittal plane at IC. The range of motion describes the eversion movement of the foot during ground contact. Runners land on the lateral side of their foot and rotate inwards after IC. The angle describing the amount of inward rotation is the range of motion in the frontal plane.

![Figure 31: Visualization of sole angle and range of motion.](image)

5.2.2 Data set

We collected data of 29 recreational runners (23 male; 6 female) with a mean age of 24.9 ± 2.4 years. All subjects were informed about related risks and gave written consent to participate in the study and for the collected data to be published. The data was acquired in a laboratory with a motion capture system (Vicon Motion Systems Inc., Oxford, UK) as reference. All subjects wore the same kind of running shoes (adidas Response Cushion 21, adidas AG, Herzogenaurach, Germany). Both the left and the right shoe were equipped with four IMU sensors. The sensors were located in a cavity in the sole of the running shoe, laterally under the ankle, at the heel, and on the instep (Table 10; Figure 32).
Table 10: Naming of the sensor position and details on the mounting of the sensors.

<table>
<thead>
<tr>
<th>Name</th>
<th>Mounting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cavity</td>
<td>Cavity cut in the sole of the shoe under the arch</td>
</tr>
<tr>
<td>Instep</td>
<td>Mounted with suiting clip to laces of the shoe</td>
</tr>
<tr>
<td>Lateral</td>
<td>Mounted with tape laterally under ankle</td>
</tr>
<tr>
<td>Heel</td>
<td>Mounted with tape on heel cap</td>
</tr>
</tbody>
</table>

Figure 32: Visualization of sensor positions on the running shoes, the global coordinate system \((x_g,y_g,z_g)\), the shoe coordinate system \((x_s,y_s,z_s)\), and the individual sensor coordinate systems. When the foot is flat on the ground, the global and the shoe coordinate system are aligned.

For the study, we used miPod IMU sensors [191]. The accelerations \(\ddot{a}[n]\) and angular rates \(\dot{\omega}[n]\) at sample \(n\) measured with those sensors will be denoted as

\[
\ddot{a}[n] = \begin{pmatrix} a_x[n] \\ a_y[n] \\ a_z[n] \end{pmatrix} \quad \text{and} \quad \dot{\omega}[n] = \begin{pmatrix} \omega_x[n] \\ \omega_y[n] \\ \omega_z[n] \end{pmatrix}, \quad (16)
\]

where indices \(x\), \(y\), and \(z\) denote the vector components along the axis of the respective sensor (Figure 32). The sensors sampled accelerations and angular rates with a frequency of \(f_s = 200\ \text{Hz}\) and a resolution of 16 bit. According to Potter et al. [114], we set the range of the accelerometer to \(\pm 16\ \text{g}\) and the range of the gyroscope to \(\pm 2000\ \text{deg/s}\).

Prior to the data acquisition, the miPod sensors were calibrated using the calibration routine introduced by Ferraris et al. [120]. Additionally, a functional calibration routine was performed to align the individual sensors’ coordinate systems with the shoe coordinate system \((x_s,y_s,z_s)\).
The functional calibration routine (Figure 33) was performed for each subject and generated two vector pairs for each sensor. Each vector pair consisted of one vector in the sensor frame and one vector in the shoe frame:

1. Vector pair superior/inferior direction: The subjects were asked to stand still with both feet on the ground. Thus, the accelerometer of all sensors measured the gravitational acceleration in the sensor frame. The $z_s$-axis was defined as the corresponding vector in the shoe frame.

2. Vector pair medial/lateral direction: The subjects rotated their feet on a balance board, which only allowed for a rotation in the shoe frame’s sagittal plane. A gyroscope in the shoe frame measures the angular rate of the rotation on the medial/lateral axis. The medial/lateral axis of the shoe frame corresponds to the principle component of the angular rate data during rotation in the sensor frame. The $x_s$-axis was defined as the medial/lateral axis in the shoe frame.

Using these vector pairs, we computed a subject dependent rotation matrix for each sensor location, which rotated the IMU-data in the sensor frame into the shoe frame using an adapted version of the Whaba algorithm [125, 127]. After applying this rotation matrix to the sensor data, all sensor frames were aligned with the shoe frame, which makes the raw IMU data comparable on each axis. Besides, the functional calibration offers the possibility to run the same algorithms on each sensor due to the same alignment of the sensors and thus enable a fair comparison between the sensor positions. For the simplicity of the notation in this work, we keep the convention for acceleration $\mathbf{a}[n]$ and gyroscope $\mathbf{\omega}[n]$ defined in Equation 16, but use $x$, $y$, and $z$ as the axis in the shoe coordinate system from now on.

The motion capture ground truth system consisted of 16 infrared cameras and sampled the positional data of the retroreflective markers with a sampling rate of $f_s = 200$ Hz. The running shoes were equipped with a subset of the marker setup described by Michel et al. [168]. For our study, we only used the six markers attached to each foot. Using these markers and the marker-based stride segmentation method for IC and TO using motion capture data introduced by Maiwald et al. [169], the reference values for the spatio-temporal stride features could be estimated.
Evaluation of IMU sensor position

Figure 33: Visualization of the functional calibration procedure. The first part of the functional calibration consisted of standing still with the foot flat on the ground in order to measure gravity. During the second part the subjects rotated their feet on a balance board to compute the medial/lateral axis using a principal component analysis.

The sensors and the motion capture system were synchronized using an adapted version of the wireless trigger introduced by Kugler et al. [170]. Due to small differences in the IMUs sampling rates, this procedure only allowed for a stride-to-stride synchronization, not a sample-to-sample synchronization.

In the described set-up, each subject was asked to run 50 times through the motion capture volume. We controlled for speed by capturing different number of trials in different velocity ranges between 2–6 m/s (Table 11). Using the described study set-up, we were able to collect data of 2426 strides.

Table 11: Number of trials and recorded strides per velocity range. During the data acquisition, we controlled for speed and the subjects only changed the velocity range, if the required number of trials in the previous (slower) velocity range was reached.

<table>
<thead>
<tr>
<th>Velocity range</th>
<th>Number of trials</th>
<th>Number of strides</th>
</tr>
</thead>
<tbody>
<tr>
<td>2–3 m/s</td>
<td>10</td>
<td>962</td>
</tr>
<tr>
<td>3–4 m/s</td>
<td>10</td>
<td>558</td>
</tr>
<tr>
<td>4–5 m/s</td>
<td>15</td>
<td>544</td>
</tr>
<tr>
<td>5–6 m/s</td>
<td>15</td>
<td>362</td>
</tr>
</tbody>
</table>

An example stride for the four IMU sensors, which was segmented from IC to IC, is depicted in Figure 34.
5.2 Methods

Figure 34: Exemplary IMU data of one stride segmented from IC to IC for the four different sensor positions.
5.2.3 Algorithm

5.2.3.1 Stride segmentation

We used a combination of different existing stride event detection algorithms to find IC, MS, and TO in the IMU signal. IC and TO were computed by finding maxima in the angular rate data of the sagittal plane. According to Falbriard et al. [128], these two maxima are reliable features to detect IC and TO even though they have a certain bias from the actual IC and TO.

We found a two stage approach to be most reliable for IC detection. Firstly, we used a cyclicity estimator of Šprager et al. [136] to find the swing phase index \( n_{SP,i} \) before the \( i \)-th ground contact at the local minimum of the gyroscope in the sagittal plane (Figure 35). This instant in time corresponds to the forward swing of the foot during swing phase. From this fiducial point, we searched for the next local maximum of the gyroscope signal in the sagittal plane, which corresponds to the point of maximum rotation in the sagittal plane after \( n_{SP,i} \). Falbriard et al. [128] reported a bias of 11 ms for this fiducial point to the actual IC event. Due to this fact and the chosen sampling rate of 200 Hz, we corrected the detected maximum by two samples to find the index of the \( i \)-th IC \( n_{IC,i} \).

From this time instance, the index of the \( i \)-th MS \( n_{MS,i} \) was determined by finding the minimum in the gyroscope L2-norm after \( n_{IC,i} \) [131] in a time range of 250 ms (50 sample), which is the average time of a stance phase while running with speeds up to 6 m/s [178]:

\[
n_{MS,i} = \arg\min_n \sqrt{\omega_x^2[n] + \omega_y^2[n] + \omega_z^2[n]} , n \in [n_{IC,i}; n_{IC,i} + 50] \quad (17)
\]

For the detection of TO, we used the first maximum in the angular rate data after the maximum related to IC in a time range of 400 ms [178]. We added 150 ms to the average stance time of 250 ms for speeds up to 6 m/s to account for running styles with longer ground contact phase. According to Falbriard et al. [128], this maximum in the gyroscope signal can be detected reliably even though it has a speed-dependent bias from the actual TO. In their work, they provided speed-dependent biases for different fixed running speeds. Because the subjects in our study ran in wide speed bins of 1 m/s, we could not use the provided speed dependent
biases. However, we used the authors’ overall 24 ms bias from the second maximum in the angular rate signal of the sagittal plane to find the index of the $i$-th TO event $n_{TO,i}$.

![Figure 35: Visualization of the stride segmentation for the cavity sensor using the gyroscope signal in the sagittal plane $\omega_x[t]$. The fiducial points at swing phase $n_{SP}$ are local minima of the angular rate in the sagittal plane. The index $n_{IC}$ indicates the index of IC, which corresponds to the bias corrected local maximum after $n_{SP}$. The MS event $n_{MS}$ is at the minimum of the gyroscopic energy. The TO event at $n_{TO}$ is based on the second local maxima and a bias correction.]

5.2.3.2 Computation of foot trajectory

For the computation of the spatial parameters we implemented a strap-down integration algorithm to reconstruct the trajectory of the foot using a zero-velocity update to reduce IMU drift. The trajectory algorithm is an adapted version of the sensor fusion algorithm introduced by Rampp et al. [6]. It consists of a quaternion-based orientation estimation using the gyroscope data followed by a gravity corrected and dedrifted integration using the acceleration data.

The orientation estimation is based on the zero-velocity assumption. During MS the foot is assumed to be flat on the ground with a velocity of 0 m/s. Therefore, we can reset the orientation and the velocity for each stride during MS. For running, this assumption is often violated due to the dynamic nature of running for high running speeds. Besides, in the case of forefoot running no flat-foot phase exists. However, we could show in one of our prior publications [P1] that the results for stride length
and average stride velocity are still accurate using this assumption. The following computational steps were applied to each running stride. Using the zero-velocity assumption, we initialized the orientation $\vec{a}[n_{MS,i}]$ and the translation $\vec{d}[n_{MS,i}]$ of the foot at the $i$-th MS index $n_{MS,i}$ with zero:

$$\vec{a}[n_{MS,i}] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \vec{d}[n_{MS,i}] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$ (18)

After the initialization, we computed the orientation of the sensor during one stride using a quaternion-based forward integration. For this work, we use a vector representation for quaternions:

$$\mathbf{q} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}$$ (19)

The quaternion $\mathbf{q}[n+1]$ at sample $n+1$ was computed from the previous quaternion $\mathbf{q}[n]$ for all samples between two consecutive MSs ($n \in [n_{MS,i}; n_{MS,i+1}]$) using the following formulas [192, 193].

$$\mathbf{q}[n+1] = \mathbf{q}[n] \otimes \mathbf{d}\mathbf{q}\bar{\omega}[n] = \mathbf{q}[n] \otimes \exp\left(\frac{1}{2f_S} \mathbf{W}[n]\right)$$ (20)

$$\text{exp}(\mathbf{q}) = \exp(q_0) \begin{pmatrix} \cos \left(\frac{q_1}{\sqrt{q_1^2 + q_2^2 + q_3^2}}\right) \\ \frac{q_2}{\sqrt{q_1^2 + q_2^2 + q_3^2}} \sin \left(\frac{q_1}{\sqrt{q_1^2 + q_2^2 + q_3^2}}\right) \\ \frac{q_3}{\sqrt{q_1^2 + q_2^2 + q_3^2}} \sin \left(\frac{q_1}{\sqrt{q_1^2 + q_2^2 + q_3^2}}\right) \\ \frac{1}{\sqrt{q_1^2 + q_2^2 + q_3^2}} \sin \left(\frac{q_1}{\sqrt{q_1^2 + q_2^2 + q_3^2}}\right) \end{pmatrix}$$ (21)
In Equation 20, the quaternion \( q[n] \) describes the rotation of the sensor from the initial position during MS \( n_{MS} \) to the position at sample \( n \in ]n_{MS,i}; n_{MS,i+1} \). The quaternion \( q[n+1] \) is defined by rotating the quaternion \( q[n] \) by the differential quaternion \( dq_\omega[n] \) which describes the rotation during a time interval of duration \( T = \frac{1}{f_s} \). This differential quaternion can be computed using the angular rate data \( \dot{\omega}[n] \).

The quaternion sequence was used in two ways. First, the orientation of the sensor was computed in Euler angle representation by converting the quaternions to roll, pitch, and yaw:

\[
\hat{\alpha}[n] = \begin{pmatrix}
\text{atan}^2\left(\frac{q_0[n]q_1[n]+q_2[n]q_3[n]}{1-2(q_1[n]^2+q_2[n]^2)}\right) \\
\text{asin}\left(\frac{2(q_0[n] \cdot q_2[n])}{1-2(q_2[n]^2+q_3[n]^2)}\right) \\
\text{atan}^2\left(\frac{q_0[n]q_3[n]+q_1[n]q_2[n]}{1-2(q_3[n]^2+q_2[n]^2)}\right)
\end{pmatrix}
\]  

(22)

Second, the quaternion sequence was used for the gravity removal of the acceleration signal. The accelerometer constantly measures not only the acceleration of the foot movement, but also the gravitational acceleration. Using the determined quaternion sequence, the acceleration data \( \hat{a}[n] \) were rotated from the shoe frame \((x_s, y_s, z_s)\) into the global frame \((x_g, y_g, z_g)\). The reason for this is that the sensor frame and the global frame coincide during MS and each quaternion \( q[n] \) describes the rotation from the sample at position \( n \) to the initial position at MS. In the global frame, the magnitude and direction of gravity was known, and thus we could remove it from the movement acceleration:

\[
\begin{pmatrix}
0 \\
\hat{a}_{x,gc}[n] \\
\hat{a}_{y,gc}[n] \\
\hat{a}_{z,gc}[n]
\end{pmatrix} = q[n] \otimes \begin{pmatrix}
0 \\
a_x[n] \\
a_y[n] \\
a_z[n]
\end{pmatrix} \otimes q[n]^{-1} - \begin{pmatrix}
0 \\
0 \\
0 \\
-9.81
\end{pmatrix}
\]  

(23)

As a last step, the translation of the sensor was computed by a dedrifted double integration of the gravity corrected acceleration signal \( \hat{a}_{gc}[n] \). After the first integration of the acceleration signal over time, a linear dedrifting function was used to remove the velocity drift \( \dot{\delta}[n] \) introduced by the integration of the acceleration. From the dedrifted velocity
\( \vec{v}_{\text{dedriffed}}[n] \), the translation \( \vec{d}[n] \) of the sensor was computed by another integration over time:

\[
\begin{align*}
\vec{v}[n] &= \sum_{m=0}^{n} \frac{1}{f_s} \vec{a}_{gc}[m] \\
\vec{\delta}[n] &= \frac{\vec{v}[n_{MS,i+1}] - \vec{v}[n_{MS,i}]}{n_{MS,i+1} - n_{MS,i}} (n - n_{MS,i}) \\
\vec{v}_{\text{dedriffed}}[n] &= \vec{v}[n] - \vec{\delta}[n] \\
\vec{d}[n] &= \sum_{m=0}^{n} \frac{1}{f_s} \vec{v}_{\text{dedriffed}}[m]
\end{align*}
\]

The individual steps of the trajectory computation are visualized for one sample stride in Figures 36, 37, 38, and 39.

### 5.2.3.3 Parameter computation

Based on the segmentation indices, the parameters stride time \( t_{\text{stride}} \) and ground contact time \( t_{gc} \) were computed as follows,

\[
\begin{align*}
t_{\text{stride}} &= \frac{n_{IC,i+1} - n_{IC,i}}{f_s} \\
t_{gc} &= \frac{n_{TO,i} - n_{IC,i}}{f_s}
\end{align*}
\]

Stride length \( d_{\text{stride}} \) and average stride velocity \( v_{\text{stride}} \) were based on the translation of the foot obtained from the trajectory estimation:

\[
\begin{align*}
d_{\text{stride}} &= \sqrt{d_x[n_{MS,i+1}]^2 + d_y[n_{MS,i+1}]^2} \\
v_{\text{stride}} &= \frac{d_{\text{stride}}}{t_{\text{stride}}}
\end{align*}
\]

Please note that we assumed level running and thus only use the \( x \) and \( y \) component of the translation due to the fact that the running path
Figure 36: Acceleration $\vec{a}[t]$ and angular rate $\vec{\omega}[t]$ raw data of the heel sensor segmented from IC to TO of the consecutive stride.
Evaluation of IMU sensor position

Figure 37: Visualization of the gravity removal in the acceleration signal for a sample stride of the heel sensor. The upper plot shows the raw acceleration $\vec{a}[t]$ segmented from MS to MS measured by the accelerometer. The lower plot shows the gravity corrected acceleration signal $\vec{a}_{gc}[t]$ after rotating the raw acceleration by the quaternion sequence $q[n]$ and removing gravity from the rotated signal. After the gravity removal, both the $z$-components of the acceleration at the first midstance ($t=0$ s) and the second midstance ($t=0.81$ s) have values close to zero.
Figure 38: Visualization of the dedrifting of the velocity after the first integration of the acceleration signal for a sample stride of the heel sensor. The upper plot shows the velocity $\vec{v}[t]$ before dedrifting. This signal displays that the velocity at the second midstance $(t=0.81)$ is not zero. We enforce the velocity to be zero by dedrifing the velocity using a linear dedrifting function. The lower plot shows the velocity $\vec{v}_{dedrified}[t]$ after dedrifting. Now, the velocity at the second MS is zero in all directions.
Figure 39: Visualization of the trajectory for a sample stride of the heel sensor. The upper plot shows the orientation $\vec{\alpha}[t]$ obtained by the quaternion based forward integration after converting the quaternions back to their angle representation. The lower plot shows the translation $\vec{d}[t]$ obtained by dedrifted double integration of the gravity corrected acceleration $\vec{a}_{gc}[t]$. 
through the ground truth system was also flat. Thus, an error of the translation in $z$-direction was neglected.

In order to compute the angle parameters, one integration step was missing. We defined our stride from the $i$-th MS at $n_{MS,i}$ to the $(i + 1)$-th MS at $n_{MS,i+1}$. For the computation of the sole angle of the $i$-th ground contact, the orientation in the sagittal between the $i$-th IC $n_{IC,i}$ and MS $n_{MS,i}$ was not computed. For the computation of the range of motion of the $(i + 1)$-th ground contact phase, the last part of the eversion movement in the frontal plane happens after the $(i + 1)$-th MS at index $n_{MS,i+1}$. Thus, we could neither compute both sole angle and range of motion for the $i$-th ground contact nor for the $(i + 1)$-th ground contact. We decided to compute sole angle and range of motion for the first ground contact by adding a quaternion-based backward integration for the samples $n \in [n_{IC,i}; n_{MS,i}]$.

We used Equation 20 for the computation of the backward integration. For this, the measured angular rate data was inverted by multiplying it by minus one. After that, the gyroscope values were integrated backwards from $n_{MS,i}$ to $n_{IC,i}$ by applying Equation 20. Finally, we converted the obtained quaternion sequence to the Euler angle representation using Equation 22 and concatenated it with the Euler angle orientation sequence obtained from $n_{MS,i}$ to $n_{MS,i+1}$.

In this orientation sequence, the sole angle is the angle obtained at the first IC $n_{IC,i}$ in the sagittal plane, because this angle describes the rotation of the shoe from IC to MS, where we assume the foot to be flat on the ground. Please note that a negative sole angle indicates a rearfoot runner, whereas a positive angle indicates a forefoot runner. The range of motion in the frontal plane is defined as the difference of the maximum and the minimum of the angle in the frontal plane between IC and TO (Figure 40).

5.2.4 Evaluation

To evaluate the effect of the sensor positions, we compared the raw IMU signals of the individual sensor positions with each other as well as the errors for the IMU-based spatio-temporal parameters.
5 Evaluation of IMU sensor position

Figure 40: Visualization of angle computation for a sample stride from the cavity sensor. The angles are depicted from $n_{IC}$ ($t = 0$ s) to $n_{TO}$ ($t = 0.32$ s). The sole angle is defined as the rotation in the sagittal plane between IC and MS. As the orientation is initialized with zero at MS, the sole angle is the angle at $n_{IC}$. The range of motion is defined as the difference between the maximum and minimum (red dots) of the angle in the frontal plane during ground contact.

5.2.4.1 Evaluation of raw data similarity

For the comparison of the raw signals, we used Pearson’s correlation coefficients [194]. We computed the correlation coefficients individually for each raw data axes between all sensor positions for the full stride ($n \in [n_{IC}; n_{IC+1}]$), the ground contact phase ($n \in [n_{IC}; n_{TO}]$), and the swing phase ($n \in [n_{TO}; n_{IC+1}]$). We combined the correlation coefficients of the three acceleration and the three gyroscope axes for each sensor position pair and plotted the distribution of the correlation coefficients using boxplots. For the segmentation in ground contact phase and swing phase, we used the labels of the stride segmentation algorithm. In case the segmented signals had different length due to errors in the event detection, we cut the duration of all sensors signals to the shortest duration.

5.2.4.2 Evaluation of spatio-temporal parameters

To evaluate the spatio-temporal parameters with respect to the sensor positions, we computed the error of individual parameters for all sensor positions. For this work, we defined the error $E_{param}$ as follows,

$$E_{param} = P_{sensor} - P_{gold}$$ (27)
In this formula, $P_{\text{sensor}}$ is the value of the parameter computed by the IMU sensor and $P_{\text{gold}}$ the value of the parameter determined by the ground truth. This formula indicates, that a positive error indicates an overestimation of the parameter and a negative error an underestimation of the parameter.

In order to understand the impact of running speed on the spatial parameters, we evaluated stride length $d_{\text{stride}}$ and the quality of the zero-velocity update for the different speed ranges. To evaluate the zero-velocity update, we computed the L2-norm of the difference of the acceleration signal at MS and the gravity vector in the global frame:

$$E_{\vec{a}}[n_{MS,i}] = \left\| \begin{pmatrix} a_x[n_{MS,i}] \\ a_y[n_{MS,i}] \\ a_z[n_{MS,i}] \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 9.81 \end{pmatrix} \right\|_2$$

(28)

The idea of the error measure $E_{\vec{a}}[n_{MS,i}]$ is that no other acceleration except gravity should be measured by the accelerometer during MS $n_{MS,i}$. In case we also measure other accelerations, the zero-velocity assumption is violated and the error measure $E_{\vec{a}}[n_{MS}]$ increases.

5.3 Results

5.3.1 Results of raw data similarity

Figure 41 visualizes the Pearson’s correlation coefficients as box plots. Each box represents the correlation coefficients of either the accelerometer or the gyroscope raw data in all three spatial directions between two sensor positions for all strides recorded during the data acquisition. We observe, that the sensors at different sensor position measure different signals, especially for the accelerometer signals. This means, that the sensor position has an impact on the IMU raw data. We can also see that the correlation coefficients for the cavity, heel, and lateral sensor position always yield the lowest values in combination with the instep sensor. Generally, the correlation coefficients of the raw data are higher for the gyroscope values as for the accelerometer values and the correlation coefficients are higher during swing phase than during ground contact phase.
Figure 41: Results of the evaluation of the Pearson’s correlation coefficients between the IMU raw signals. Each box visualizes the correlation coefficients between two sensors for all the strides in x, y, and z direction. The box plots also display the median of the correlations (median line), the IQR (box), and the 5 and 95 percentiles (whiskers). The upper plot depicts the correlation of the full strides, the middle plot the correlations during ground contact phase, and the lower plot the correlations during swing phase.
5.3.2 Results of spatio-temporal parameters

Table 12 lists the median errors and IQRs of the IMU-based computation of stride time, ground contact time, stride length, average stride velocity, sole angle, and range of motion for the different sensor positions. We see that we can accurately measure stride time with all sensors. For ground contact time, we observe large IQRs of the errors for all sensor positions and higher median errors for the cavity and the instep sensor. For the sole angle, we observe higher median errors for the heel and the lateral sensor. For the range of motion in the frontal plane we find smaller errors than for the sole angle. For the parameters stride length and average stride velocity which are based on the computed translation, we see that the cavity sensor outperforms the other sensor positions with respect to both the median and the IQR.

Table 12: Median error (Med.) and IQR of error for the parameters stride time $t_{\text{stride}}$, ground contact time $t_{gc}$, stride length $d_{\text{stride}}$, average stride velocity $v_{\text{stride}}$, sole angle $\alpha_{\text{sole}}$, and range of motion ROM compared to the motion capture system.

<table>
<thead>
<tr>
<th></th>
<th>Cavity</th>
<th>Heel</th>
<th>Instep</th>
<th>Lateral</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{\text{stride}}$ [ms]</td>
<td>-0.5</td>
<td>0.0</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>$t_{gc}$ [ms]</td>
<td>-11.0</td>
<td>-1.3</td>
<td>-22.6</td>
<td>-1.7</td>
</tr>
<tr>
<td>$\alpha_{\text{sole}}$ [deg]</td>
<td>1.6</td>
<td>-6.1</td>
<td>2.1</td>
<td>-5.9</td>
</tr>
<tr>
<td>ROM [deg]</td>
<td>0.0</td>
<td>1.2</td>
<td>2.3</td>
<td>1.4</td>
</tr>
<tr>
<td>$d_{\text{stride}}$ [cm]</td>
<td>0.3</td>
<td>-8.3</td>
<td>-5.6</td>
<td>-3.3</td>
</tr>
<tr>
<td>$v_{\text{stride}}$ [m/s]</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Figure 42 depicts the error of the stride length $E_{d_{\text{stride}}}$ (Figure 42a) as well as the error of the acceleration at the zero-velocity updated $E_{\vec{a}}[n\text{ms}]$ (Figure 42b) for the different sensor positions in the four speed ranges defined in our study (Table 11). We see that both stride length as well as the error of the acceleration show larger errors for higher running velocities. Besides, the results are congruent with the median errors and show that the cavity sensor position outperforms the other sensor positions for the parameter stride length.
Figure 42: Visualization of the error for (a) stride length and (b) the acceleration at the zero-velocity update for the four different sensor positions in different speed ranges.
5.4 Discussion

Overall, no sensor position clearly outperforms all the other sensor positions even though the cavity sensor provides the best results for the translational parameters based on the reconstructed foot trajectory. In the following paragraphs, we will discuss the raw data comparison, the temporal parameters as well as the spatial parameters individually.

5.4.1 Differences in raw data

The correlation coefficients are higher for the gyroscope signals than for the acceleration signals (Figure 41). The reasons for this observation are twofold. On the one hand, the angular rate is less sensitive to movement artifacts than the accelerometer. Especially during the ground contact phase, the impact during IC introduces high frequency vibrations in the accelerometer signal, which differ for the individual sensor positions and cannot be correctly measured with a sampling frequency of $f_s = 200$ Hz. On the other hand, the accelerometer measures different centripetal accelerations depending on the IMU position. The four main joints that cause the rotations in running are the hip, the knee, the ankle, and the metatarsal joint. Whereas the distance to the hip joint and the knee joint is similar for all sensors and thus neglectable with respect to differences in the centripetal accelerations, the raw signals differ with respect to the distance to the ankle and metatarsal joint. Thus, the accelerations measured by the IMUs will be different due to the distance dependent centripetal accelerations caused by the rotations around those joints. The rate of rotation is independent of the distance to the rotation center, if we assume the foot to be a solid object.

Additionally, the correlation coefficients during swing phase are higher than during ground contact phase (Figure 41). In particular, the gyroscope raw data of sensors at different positions can reach correlation coefficients close to one. During swing phase, the foot can be seen as a rigid segment, because there is no rotation around the metatarsal joint or any other distortion which might cause the sensors to rotate differently. The high correlation coefficients also indicate that the functional calibration procedure applied before the data acquisitions is capable of aligning the coordinate systems of the sensors at the different positions of the foot. If this was not the case, such high correlations would not be possible because rotations would be captured around different sensor axes. During ground contact phase however, the shoe upper is deformed.
due to deformation of the foot and the movement of the foot within the shoe. This causes movements at different positions on the shoe, which results in the measurement of different signals and lower correlation coefficients.

The correlation coefficients for the accelerations of the full strides show that the correlations to the instep sensors always yields the lowest values. We argue that the instep sensor is exposed to the highest motion by the deformation of the foot. During ground contact the arch of the foot and the forefoot flatten, which causes the upper of the running shoe to deform. This deformation of the upper also causes the laces and the tongue of the shoe to move and, consequently, the instep sensor as well.

More detailed analysis of the correlations reveals that the closer two sensors are located to each other, the higher their pairwise correlation. This makes sense, as we already discussed that the accelerations for the sensor positions differ depending on the distance to the joints causing the rotations. Thus, spatially close sensors have similar distances to those centers of rotation and and might be similarly effected by deformations of the shoe.

One limitation of our approach to comparing the raw signals is that the Pearson’s correlation coefficients might be influenced by differences in the detected events. Depending on the sensor positions, the events might be detected with a slight time shift. However, the high correlation coefficients during swing phase show, that the signals temporally match well and are not heavily influenced by time shifts.

In summary, the correlation coefficients show that the signals of the sensors at the four positions of the running shoe vary which causes different results for the spatio-temporal parameters. Especially the larger differences during ground contact have an effect on the accuracy of the parameters.

### 5.4.2 Temporal parameters

The results for stride time $t_{\text{stride}}$ indicate that the bias corrected maximum of the angular rate signal in the sagittal plane is a reliable fiducial point which allows accurate estimations of the stride time for all sensor positions. With median errors of less than 0.5 ms and IQR of less than 8.6 ms (< 2 samples at a sampling rate of 200 Hz) no sensor position clearly performs best. These accuracies for stride time are similar to the
results presented by Falbriard et al. [128] who reported a stride time error $0 \pm 3$ ms for their evaluation of stride time with an instep sensors. We explain the higher IQRs with the higher variability in speed in our data acquisition. Due to the fact that a sample-to-sample synchronization was not possible using our study setup, we could not evaluate the actual accuracy of the IC event for the different sensor positions. Thus, our results for stride time show only that we can reliably detect the fiducial point, but not that we can accurately detect IC with all sensor positions.

Generally, both the median errors and the IQRs for ground contact time $t_{gc}$ at all sensor positions are worse than for stride time $t_{stride}$. One of the reason for that is the inaccuracy of the TO detection. In running gait the IMU signal exhibits no clear feature at TO, whereas in walking gait, TO is indicated by a zero crossing of the angular rate data in the sagittal plane [6]. The dynamic nature of the running gait does not exhibit such a feature [129]. Moreover, Falbriard et al. [128] showed that the bias of the local maximum in the sagittal plane used to detect TO is speed dependent. We neglected this speed dependency, because we had no information for the bias except for the fixed speeds that Falbriard et al. used in their study. We used the overall bias they reported to correct the index of the maximum in the sagittal plane which explains the high IQRs due to the high variance of speed in our study. The reasons for the higher median errors for the cavity and the instep sensor are not clear. It is possible that these results indicate an actual shift of the gyroscope maximum within the gait cycle based on the sensor position. However, due to the lack of sample-to-sample synchronization, we were not able to further investigate this. In the future, a study with sample-to-sample synchronization could help to evaluate this assumption.

5.4.3 Spatial parameters

For the orientation parameters, the cavity sensor positions shows the smallest median errors for both the sole angle and the range of motion. While the differences in accuracy between the sensor positions is smaller for the median error of the range of motion (2.3° cavity/instep), the differences between the sensor positions is larger for the sole angle (7.7° cavity/heel). The cavity and the instep sensors outperform the other two sensors with respect to the median error of the sole angle. We found the reason for these large differences in the bias we used to correct the local maximum in pitch angular velocity. When we removed the bias
correction and used the local maximum as the IC event, we obtained different results (Table 13).

Table 13: Median error (Med.) and IQR of the error for the parameters sole angle $\alpha_{sole}$ without the bias correction for IC.

<table>
<thead>
<tr>
<th></th>
<th>Cavity</th>
<th>Heel</th>
<th>Instep</th>
<th>Lateral</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{sole}$ [deg]</td>
<td>Med.</td>
<td>IQR</td>
<td>Med.</td>
<td>IQR</td>
</tr>
<tr>
<td></td>
<td>6.8</td>
<td>10.2</td>
<td>-2.9</td>
<td>6.8</td>
</tr>
</tbody>
</table>

If no bias correction is applied during the event detection of IC the median errors are better for the heel and the lateral sensor position. Falbriard et al. [90] explained the high standard deviation for the sole angle evaluation in their work with the fact that the accuracy of the sole angle is heavily dependent on the accuracy of the IC detection algorithm. Due to the high angular velocities around IC, the integrated angle values are sensitive to the timing of the IC event. Based on the improved accuracy for a different bias for the heel and lateral sensor position we assume that the bias for the IC event is dependent on the sensor position itself. This underlines the need for a closer investigation of those biases for the different sensor positions. Nevertheless, we can conclude for the angle parameters, that the differences in the accuracy do not originate from the underlying raw data, but rather the stride segmentation algorithms.

For the stride length parameter the cavity position performs best, followed by the lateral position. The heel and the instep sensor position perform worst. These observations also hold for the different speed ranges (Figure 42a). We can see for all speed ranges that both the median errors and the IQRs are smaller for the cavity and the instep sensor, even though the accuracy drops with higher speed for those positions as well. One possible explanation for this fact is that the zero-velocity assumption during MS becomes less valid with higher speeds. The results for the error of the acceleration during zero-velocity phase are congruent with the errors in stride length (Figure 42b). The error $E_{\ddot{a}}[n_{ms}]$ is smallest for the cavity sensor, followed by the lateral sensor, whereas the heel and instep sensor yield larger errors. Besides, the error increases with speed for all sensor positions. This shows that the zero velocity assumption is violated with higher speeds, but most valid for the cavity sensor. We argue that
one of the reasons for this observation is the attachment of the sensors. The cavity inside the sole of the running shoes was manufactured to fit the shape of the miPod sensors and thus prevent the sensors from moving. The sole itself might act as a physical low pass filter by damping high frequency noise in the acceleration caused by impacts like IC by its elastic nature. Besides, the location of the sensor in the sole under the arch was chosen to be least affected by the deformation of the insole during IC and the bending in the forefoot region during the propulsion phase. Thus, the cavity sensor position faces least perturbations during ground contact and can compute the zero-velocity-based trajectory with the highest accuracy. The heel and lateral sensor were taped to the heel cap and laterally under the ankle. While the attachment itself is also firm, those sensor positions are more affected by the movement of the upper of the shoe. Especially the heel sensor faces motion due to the deformation of the back of the running shoe caused by the heavy impact during IC for rearfoot runners. The instep sensor was mounted to the laces of the running shoe using a clip. Even though we tested the firmness of the attachment for each subject before the data acquisition, this sensor position is affected the most by additional movements. This is on the one hand due to the less firm attachment in comparison to the other sensor positions and on the other hand due to the highest amount of deformation of the upper of the running shoe, which we already discussed in the raw data section.

These movements and consequently the errors during MS harm the validity of the zero-velocity assumption which is the basis for the strap-down integration algorithm. Due to the fact that the sensor positions seem to be affected by additional movements or noise sources during MS to a different extent, the performance of the zero-velocity-based strap-down integration algorithm for the trajectory computation is different for the evaluated sensor positions.

The errors of the average stride velocity for the different sensor positions correlate with the ones for stride length. The reason for this correlation is that the average stride velocity is computed by dividing stride length by stride time. As the stride time computation works well for all sensor positions, the main error source for average stride velocity is the computed stride length. However, we can see from the small median errors and IQRs that average stride velocity can accurately be computed with foot worn IMUs.
5 Evaluation of IMU sensor position

5.4.4 General aspects

Our results indicate that the raw signals and thus the results for the different spatio-temporal parameters differ for the evaluated sensor positions. This implicates that when comparing studies of IMU-based spatio-temporal parameters with each other, the sensor position should be considered as a source of difference in future review studies. Besides, adaptations to the algorithms like changing the detection algorithm of IC for the computation of the sole angle can result in performance boosts for individual sensor positions. Thus, porting algorithms to other sensor positions can be possible, but algorithms might have to be adapted.

Nevertheless, we recommend the cavity sensor position due to the unobtrusive and firm attachment inside the running shoe sole. It is least affected by any additional movement than the actual foot movement and thus has the highest raw signal quality. However, we want to note at this point that this evaluation is purely based on the signal quality and the accuracy of the resulting spatio-temporal parameters and that we did not consider usability aspects. While the unobtrusiveness of the cavity sensor position is good for a smart shoe application with fully integrated sensors, it might not be a good solution for a sensor system which should be usable with different pairs of shoes. In this scenario, all the shoes would need a cavity and the sensors would need to be put under the sockliner before each run.

Last but not least, we only investigated the zero-velocity-based strapdown integration algorithm as a method to determine the spatio-temporal parameters. Another method which showed promising results in one of our prior publications for a cavity sensor was a deep learning based algorithm [P1]. The advantage of this algorithm was that no additional assumptions (zero-velocity-phase) or signal processing techniques (gravity correction) are necessary to determine the parameters. Thus, problems with the zero-velocity updates due to the movement of the sensor at different positions on the running shoe are not expected with this type of methodology. Nevertheless, the Deep Convolutional Neural Network (DCNN) model trained in [P1] still relied on stride segmentation and zero padding of the IMU data. Thus, the problem with the differences in the stride segmentation algorithm and the resulting differences in the angular parameters might remain. Using another network structure like Long Short-Term Memories (LSTMs) might solve this problem. This
type of recurrent neural networks can learn to also incorporate informa-
tion from prior parts of the IMU signal. However, another study setup to
create ground truth data would be needed for such an approach. Using
our study setup, we could only create ground truth for single strides in a
limited capture volume of the motion capture system. In order to train a
LSTM continuous ground truth data over a longer period of time would
be required. Nevertheless, one of the reasons why the deep learning
approach yielded worse results in [P1] was that the training data was not
enough to learn a good representation in feature space for the regression
of spatio-temporal parameters. With more data from the other sensor
positions, future work could investigate whether this additional data can
improve the accuracy and generalization capability of the network by
training one model with the data of all sensor positions. Thus, the model
would learn from more data during training and could generalize to other
sensor positions on the running shoe.

5.5 Conclusion

We presented an evaluation of the effects of four different IMU sensor
positions on the accuracy of IMU-based endurance running parameters.
We conducted a study with 29 subjects which were equipped with four
IMUs placed inside a cavity in the sole of the running shoe, on the heel,
the lateral side, and the instep of a running shoe. We compared the
raw data of the individual sensor positions and implemented algorithms
for stride segmentation and the computation of the trajectory of the
foot, respectively, the running shoe. Using the data acquired during the
study, we could show that the raw signals of the IMUs differ for the sensor
positions, especially for the acceleration during the ground contact phase.
We showed that all the sensor positions could accurately measure stride
time, but not ground contact time due to the large speed dependency of
the fiducial point we used to detect toe off. The angle parameters—range
of motion and sole angle—were hard to compare for the different sensor
positions, as they were affected by the accuracy of the IC event detection
algorithm. Finally, we showed that the cavity sensor outperforms the
other sensor positions for the computation of stride length, because this
sensor positions seems not to be exposed to movement artifacts of the
upper of the shoe.

Thus, we can conclude that the sensor position has an effect on the accu-
racy of different IMU-based running parameters due to the differences
in the acquired raw signals. From a data processing perspective, we recommend to use IMU sensors inside the cavity of a running shoe even though it only outperforms the other sensor positions for the reconstruction of the trajectory using a zero-velocity-based strap-down integration algorithm.

In the future, the event detection algorithm for the different sensor positions should be investigated further. Because a sample-to-sample synchronization was not possible in our study set-up, we could not evaluate the accuracy of the actual IC/TO events but only the parameters deduced from those events. Our results indicate that, both for the IC and TO event, the biases from the maxima in the pitch angular velocity to the actual events are different for the different sensor positions. Improving the detection of the IC and TO event would not only increase the accuracy of ground contact time but also the stride angle parameters. Further, the larger errors for stride length with higher speeds should be further investigated. Especially for professional athletes, who run with speeds higher than 5 m/s, the errors in stride length are very high.

5.6 Author contributions

6 Paper 3 - Big Data Analysis on Fitness platform data

The following chapter is a replicate of the following publication:


Only a few phrases have been adapted to fit the wording within this thesis.

Abstract Finishing a marathon requires to prepare for a 42.2 km run. Current literature describes which training characteristics are related to marathon performance. However, which training is most effective in terms of a performance improvement remains unclear. We conducted a retrospective analysis of training responses during a 16 weeks training period prior to an absolved marathon. The analysis was performed on unsupervised fitness platform data (Runtastic) from 6771 marathon finishers. Differences in training volume and intensity between 3 response and 3 marathon performance groups were analyzed. Training response was quantified by the improvement of the velocity of 10 km runs $\Delta v_{10}$ between the first and last 4 weeks of the training period. Response and marathon performance groups were classified by the 33.3rd and 66.6th percentile of $\Delta v_{10}$ and the marathon performance time, respectively. Subjects allocated in the faster marathon performance group showed systematically higher training volume and higher shares of training at low intensities. Only subjects in the moderate and high response group increased their training velocity continuously along the 16 weeks of training. We demonstrate that a combination of maximized training volumes at low intensities, a continuous increase in average running speed up to the aimed marathon velocity and high intensity runs $\leq 5\%$ of the overall training volume was accompanied by an improved 10 km performance which likely benefited the marathon performance as well. The study at hand proves that unsupervised workouts stored on fitness platforms can be a valuable data source for future sport science studies.
6 Big Data Analysis on Fitness platform data

6.1 Introduction

Finishing a marathon is a fascinating goal, especially for recreational runners. More and more people follow this dream in recent years, which is indicated by the rising number of marathon participants [195, 196]. The motivations for people to take on this huge effort are manifold. They can be of personal (goal achievement), social (respect of peers), physical (lose weight) and psychological (becoming less anxious) manner [197]. Independent of the motives behind the decision to participate in a marathon, all those runners are united in the task to prepare well by bringing their bodies in shape to run 42.2 km.

Marathon preparation techniques have been under scientific investigation for decades. Many researchers have evaluated long distance runners’ training load by analyzing their training strategies with respect to volume and intensity. A high training volume has been proven to positively influence marathon performance [66, 198]. Especially recreational runners with lower training volumes can potentially increase their performance by increasing the amount of training. This was underlined by the results of Roecker et al. [199] and Tanda [200], who found training volume to be one of the key predictors for marathon performance in recreational runners.

In regard to the training intensity, various overviews outline advantages when training intensity is distributed in a polarized, i.e. non-uniform manner [71, 201, 202, 203]. Such concepts suggest spending certain proportions of the total training time within a LIT zone, a HIT zone and optionally a threshold zone. In practice, training zones are either defined from a cardiopulmonary exercises test at defined percentages of the maximal oxygen uptake (V’O2max), at intensities related to ventilatory or lactate thresholds [204] or alternatively at percentages of maximum heart rate [201] as well as at percentages of target marathon velocity [205, 206].

Overall, a significant body of research provides evidence that certain physiological factors and training characteristics are systematically related to marathon performance. However, it has yet to be shown which training characteristics are the most effective in terms of an actual fitness improvement to positively influence an individual’s marathon performance. In order to demonstrate whether certain training characteristics lead to higher fitness improvements, the natural variability of the individuals’ responses to training has to be considered [207, 208].
6.1 Introduction

Current findings mainly result from studies with defined, recruited and instructed cohorts. Such supervised investigations suffer from low participant numbers. In contrast, longitudinal and unsupervised activity data from large populations recorded in a natural habitat might enable sport scientists to derive more generalizable conclusions. Nowadays, millions of runners with different fitness levels track their training progress by uploading recorded data from wearable sensors onto platforms like Garmin, Strava, Runtastic, etc. The challenge in working with this kind of data lies in its unsupervised nature. The data are unlabeled, which means that values of ground truth and contextual subject information for specific research questions are missing. Besides, the accuracy of the wearable sensors used to acquire the data is unknown. Due to this reasons, Hicks et al. [30] postulated that a plausibility check of the data from wearable sensors is an integral part prior to its analysis. Different publications have already shown the potential of wearable sensor data from fitness platforms to further improve performance prediction [147, 149, 150], to accurately determine the critical speed of runners and to set up pacing strategies [146] and also to individualize training plans for marathon preparation [145].

Longitudinal investigations of physical activities before a marathon appear to be a promising approach to further improve the applicability, impact and efficiency of marathon training plans. To the best of our knowledge, there is no research which evaluated systematic differences in marathon training characteristics in relation to its response based on longitudinal data from a large unsupervised study cohort. Thus, we contribute to the state of the art in the following way:

1. We retrospectively analyze the response to training using data from wearable sensors. We assess response by comparing runs of the same distance with comparable heart rates as proposed by Boullosa et al. [209].

2. Based on the quantity of response, we define different response groups and analyze corresponding differences in total training volume and training intensity distribution within a 16 weeks training period prior to a performed marathon.

3. Respectively for each response group, we further analyze corresponding differences in total training volume and training intensity distribution between different marathon performance groups within a 16 weeks training period prior to a performed marathon.
6 Big Data Analysis on Fitness platform data

6.2 Methods

6.2.1 Data set

After extensive filtering (explained below) we evaluated the marathon training of 6771 runners. We used data recorded by wearable sensors such as smartphones, smartwatches or heart rate chest straps from anonymized users of the Runtastic fitness platform for the evaluation. The subjects were chosen based on the following criteria:

- one workout between 2017 and 2019 with a total distance between 41 and 43 km
- at least 16 workouts in 16 weeks leading up to the marathon
- GPS and heart rate data for each workout

We defined a range around the exact marathon distance of 42.2 km in order to include marathons of slightly different distance and inaccuracies of GPS devices used to track the marathon. Apart from distance, no additional requirements like profile or location were set for the marathon workout. The threshold of 16 workouts in the 16 weeks leading up to the marathon was empirically chosen to assure a minimum amount of data for evaluation. The data set included 5288 male subjects (78.1%), 1250 female subjects (18.5%) and 233 subjects of unknown sex (3.4%). The subjects’ mean age was 38.5±9.7 years. Body weight and height were not taken into account, because they were not available for all subjects. The GPS data (latitude, longitude) and heart rate data were sampled with different sampling rates. However, data streams of each workout were synchronized by global timestamps (UTC). GPS data was anonymized by adding a random offset to the data stream. The study is in accordance with the Declaration of Helsinki, because the local ethics committee raised no objection to its conduction due to the anonymized nature of the data.

6.2.2 Data processing

Extracting overall subject features: For normalization purposes in later processing stages, we extracted the average marathon performance velocity $v_{mp}$ and the maximum training heart rate $hr_{max}$ for each subject. The average marathon performance velocity $v_{mp}$ was determined by the duration of the marathon performance time $T_{mp}$ for the distance between 41 km and 43 km. The maximum training heart rate $hr_{max}$ was
determined to be the median of the five highest recorded heart rates over the whole training process. We decided for this approach to cope with short term outliers in the heart rate recordings.

**Feature extraction of individual workouts:** We computed a set of features for each of the $W$ workouts leading up to a subject’s marathon. The first feature obtained from the $i$-th workout ($i \in \{1, 2, 3, \ldots, W\}$) was the training duration $T_i$. $T_i$ was computed by subtracting the first from the last UTC timestamps of the GPS data. If the workout duration was longer than 90 minutes, we saved an indicator $I_{T90,i}$, which was used further on to evaluate how many long workouts were performed:

$$I_{T90,i} = \begin{cases} 0 & \text{if } T_i < 90 \text{ minutes} \\ 1 & \text{if } T_i \geq 90 \text{ minutes} \end{cases} \quad (29)$$

For all other GPS-based features, we computed the distance and velocity over time from the GPS data. We used the great circle distance implementation of the Python package GeoPy [210] to compute the distance between two consecutive GPS recordings. This resulted in a data stream of distances between two consecutive GPS-samples over time $d_i[n]$. This data stream was used to compute the total distance of the $i$-th workout $D_i$ by computing the sum over all samples. Similar to $I_{T90,i}$ (Equation 29), we computed an indicator $I_{D15,i}$ for workouts with distances longer than 15 km.

In order to assess training progress, we extracted the best velocity $v_{10,i}$ for a 10 km segment within each workout (if $D_i \geq 10$ km). For the respective 10 km segment, we also computed the average heart rate $\overline{h}_r_{10,i}$ during the time interval.

After dividing $d_i[n]$ by the corresponding duration between two consecutive GPS timestamps $\Delta t_i^{gps}[n]$, we obtained a velocity data stream $v_i[n]$. We used this data stream to compute a distribution $T_i[\tilde{V}]$ which describes the duration a subject spent in a defined velocity bin $\tilde{V}$ during the $i$-th workout. To be able to define comparable velocity bins across all subjects, we normalized the velocity data stream $v_i[n]$ by the subject’s marathon performance velocity $v_{mp}$:

$$\tilde{v}_i[n] = \frac{v_i[n]}{v_{mp}} \quad (30)$$
The velocity bins for the distribution $T_i[\tilde{V}]$ were defined from $0.54 \cdot v_{mp}$ to $1.8 \cdot v_{mp}$ with a bin width of 0.02. Thus, we computed the duration distribution function in the following manner:

$$T_i[\tilde{V}_x] = \sum_{n \in \tilde{V}_x} \Delta t_i^{gps}[n]$$

with $n$:

- $\in \tilde{V}_0$ if $\tilde{v}_i[n] \leq 0.54$
- $\in \tilde{V}_1$ if $0.54 < \tilde{v}_i[n] \leq 0.56$
- ... 
- $\in \tilde{V}_{65}$ if $1.80 < \tilde{v}_i[n]$

For simplicity of notation, we remove the bin indicator $x$ from the relative velocity bin and denote the duration distribution for different velocity bins $\tilde{V}$ as $T_i[\tilde{V}]$ in the following.

The same procedure was performed for the heart rate data $h_{ri}[m]$. This data stream was normalized by the subject’s maximum training heart rate $h_{r_{max}}$. The heart rate bins were defined from 0.5 to $1 \cdot h_{r_{max}}$ with a fixed bin width of 0.02. This procedure resulted in the duration distribution for the heart rate $T_i[\tilde{H_R}]$.

**Grouping of workout features in time frames of four weeks:** In order to evaluate the training progress over time, we defined training blocks of four weeks similar to Berndsen et al. [147] and computed aggregated features for those training blocks. The partition of the blocks was defined based on the marathon date. Equation 32 defines the rules by which the $i$-th workout was assigned to training block $tb$:

$$i \begin{cases} \in tb_1 & \text{if 16 weeks } \leq t_{marathon}[0] - t_i[0] < 12 \text{ weeks} \\ \in tb_2 & \text{if 12 weeks } \leq t_{marathon}[0] - t_i[0] < 8 \text{ weeks} \\ \in tb_3 & \text{if 8 weeks } \leq t_{marathon}[0] - t_i[0] < 4 \text{ weeks} \\ \in tb_4 & \text{if 4 weeks } \leq t_{marathon}[0] - t_i[0] < 0 \text{ weeks} \end{cases}$$

for $i \in \{1, 2, 3, \ldots, W\}$

In this equation, $t_{marathon}[0]$ describes the first UTC timestamp of the marathon workout. For the $y$-th training block the total training time $T_{tb,y}$, the total training distance $D_{tb,y}$, the number of workouts longer than
90 minutes $I_{T90,tb_y}$ and further than 15 km $I_{D15,tb_y}$ could be computed by summing the values of the workouts within the training block.

\[
\begin{align*}
T_{tb_y} &= \sum_{i \in tb_y} T_i \\
D_{tb_y} &= \sum_{i \in tb_y} D_i \\
I_{T90,tb_y} &= \sum_{i \in tb_y} I_{T90,i} \\
I_{D15,tb_y} &= \sum_{i \in tb_y} I_{D15,i}
\end{align*}
\] (33)

The best 10 km velocity $v_{10,tb_y}$ for training block $y$ was chosen from all $v_{10,i}$ of workouts in $tb_y$:

\[
v_{10,tb_y} = \max_{i \in tb_y} v_{10,i}
\] (34)

The duration distribution curves for velocity $T_i[\hat{V}]$ and heart rate $T_i[\hat{HR}]$ were combined for the different training blocks and converted to probability distributions $P_{tb_y}[X = \hat{V}]$ and cumulative distributions $F_{tb_y}[X = \hat{V}]$ (Figure 43).

For the $y$-th training block, the duration distribution curve $T_{tb_y}[\hat{V}]$ was computed by summing the duration within the velocity bin $\hat{V}$ of all the workouts belonging to the training block:

\[
T_{tb_y}[\hat{V}] = \sum_{i \in tb_y} T_i[\hat{V}] \quad \forall \ \hat{V}.
\] (35)

From $T_{tb_y}[\hat{V}]$ we computed a probability distribution $P_{tb_y}[X = \hat{V}]$ by dividing the time spent in the velocity bins by the total training time in $tb_y$:

\[
P_{tb_y}[X = \hat{V}] = \frac{T_{tb_y}[\hat{V}]}{\sum_{\hat{V}} T_{tb_y}[\hat{V}]}. \] (36)
The cumulative distribution \( F_{tb_y}[X = \tilde{V}] \) can be computed from the probability distribution by
\[
F_{tb_y}[X = \tilde{V}] = \sum_{p=0.54}^{\tilde{V}} P_{tb_y}[X = p]
\]  
(37)

The same procedure was applied to the duration distribution of the heart rate to obtain the probability distribution \( P_{tb_y}[X = \tilde{HR}] \) and cumulative distribution \( F_{tb_y}[X = \tilde{HR}] \).

Using the probability distribution functions, we computed the normalized mean training velocity \( \bar{v}_{tb_y} \) and the normalized mean heart rate \( \bar{hr}_{tb_y} \) for the \( y \)-th training block:
\[
\bar{v}_{tb_y} = \sum_{\tilde{V}} \tilde{V} \cdot P_{tb_y}[X = \tilde{V}]
\]
\[
\bar{hr}_{tb_y} = \sum_{\tilde{HR}} \tilde{HR} \cdot P_{tb_y}[X = \tilde{HR}]
\]
(38)

We also used the distribution function of the velocity to compute the share of the workout time the subjects spent in different intensity zones. Similar to Kenneally et al. [205] and Billat et al. [206], we defined the zones based on the marathon velocity. The LIT zone was defined by velocities below \( v_{mp} \) and the HIT zone by velocities above \( 1.2 \cdot v_{mp} \) (Figure 43). Using the cumulative distribution functions \( F_{tb_y}[X = \tilde{V}] \), the share of time spent in the intensity zone for \( tb_y \) was computed as
\[
\text{LIT}_{tb_y} = F_{tb_y}[X = 1]
\]
\[
\text{threshold}_{tb_y} = F_{tb_y}[X = 1.2] - F_{tb_y}[X = 1]
\]
\[
\text{HIT}_{tb_y} = 1 - F_{tb_y}[X = 1.2]
\]
(39)

All the computations for the training block analysis were also applied to all \( W \) workouts leading up to the marathon in order to obtain each subject’s overall training statistics.
6.2 Methods

(a) (b)

Figure 43: Exemplary visualization of (a) duration distribution curve $P_{tb_1}[X = \tilde{V}]$ and (b) cumulative duration distribution curve $F_{tb_1}[X = \tilde{V}]$ for the normalized velocity $\tilde{V}$ for training block 1. The red lines indicate the barriers for the intensity zones defined based on the marathon performance velocity $v_{mp}$.

Filtering data set: An interquartile range (IQR) filter was applied to exclude all subjects, where parameters of subjects ($T_{mp}$, $D$, $T$, $hr_{max}$) exceeded thresholds of $1.5 \cdot \text{IQR}$ below or above the lower and upper quartile.

In order to create valid response groups, we also excluded all subjects who did not achieve a minimum average heart rate of $0.8 \cdot hr_{max}$ for the best 10 km runs in $tb_1$ and $tb_4$. $0.8 \cdot hr_{max}$ was chosen to ensure sufficient cardiopulmonary effort for an individual best 10 km performance as well as a sufficient availability of data.

Categorization of subjects in response and marathon performance groups: Conventional metrics to assess performance improvement (i.e. $V’O_{2\text{max}}$ or lactate thresholds) were not available for the unsupervised data set. Therefore, we used the improvement of the 10 km velocity $\Delta v_{10}$ from $tb_1$ to $tb_4$ as a surrogate to evaluate the response of subjects to training throughout the 16 weeks before the marathon.

$$\Delta v_{10} = v_{10, tb_4} - v_{10, tb_1}$$ (40)

A positive value for $\Delta v_{10}$ indicates an improvement and in turn a positive response to training and vice versa. The filter for the average heart rate stated in the data filtering section assured that those assessment runs were performed with a minimum cardiopulmonary effort. Despite the
absence of conventional metrics to assess performance improvement we believe that $\Delta v_{10}$ is a plausible surrogate since it should reasonably reflect an improvement in endurance capacity [74]. Also, research has shown that the velocity of 10 km races highly correlates to marathon performances [200, 211].

$\Delta v_{10}$ was used to categorize the subjects into three groups: high response, moderate response and low response. The borders separating the three response groups were computed at the 33.3rd and 66.6th percentile of $\Delta v_{10}$. We computed the percentiles for the three response groups separately on ten different $v_{10, tb_1}$ velocity groups due to decreasing improvement for subjects with higher initial $v_{10, tb_1}$ (Figure 44). The categorization of the subjects into the response groups was based on the distribution within the velocity group and not the absolute value of $\Delta v_{10}$. We decided for this approach to assure equally sized response groups across different performance levels.

![Figure 44: Visualization of the training response $\Delta v_{10}$ across ten velocity groups. Each dot represents one subject. The response categories are color coded. The vertical black lines are located at the percentile values of $v_{10, tb_1}$. The horizontal grey line indicates the zero line, where subjects showed neither a positive nor negative improvement. Due to the statistical approach in the response group definition, which assured equally sized groups, the low and moderate response group also included subjects with negative $\Delta v_{10}$.](image)

Independent of the response group, all subjects were also categorized in three equally sized groups based on their marathon performance times using the 33.3rd and 66.6th percentile. For our data set, the 33.3rd and 66.6th percentiles referred to marathon performance times of 3 h 44' and
Based on those values we assigned each subject to a fast, medium and slow marathon performance group.

### 6.2.3 Evaluation

The evaluation consisted of three parts. Firstly, we demonstrate plausibility of the data set by reproducing known distributions and trends from literature as recommended by Hicks et al. [30] for large unsupervised data sets. Plausibility was analyzed by plotting histograms for marathon performance times $T_{mp}$, maximum training heart rate $hr_{max}$, training improvement $\Delta v_{10}$ and a regression plot relating marathon average performance velocity $v_{mp}$ to the best 10 km velocity $v_{10}$.

Secondly, mean training velocity and mean heart rate throughout the training process were analyzed to evaluate $\Delta v_{10}$ as a reasonable surrogate to measure training response for each response group. Plausibility was assumed when normalized mean velocity $\Delta \bar{v}$ between $tb_1$ and $tb_4$ increases systematically across response groups without observing a difference in normalized mean heart rate $\Delta hr$.

$$
\Delta \bar{v} = \bar{v}_{tb_4} - \bar{v}_{tb_1}
$$

$$
\Delta hr = \bar{hr}_{tb_4} - \bar{hr}_{tb_1}
$$

(41)

Differences in $\Delta \bar{v}$ and $\Delta hr$ between response groups were analyzed using a one-way ANOVA.

Lastly, means and standard deviations were derived for training parameters describing the training volume. These parameters are total distance $D$, total training duration $T$, total number of workouts $W$ and number of workouts longer than 90 minutes $l_{T90}$ or 15 km $l_{D15}$ for the complete training period of 16 weeks. Additionally, means and standard deviations were derived for the training intensity parameters describing the share of time in the LIT, threshold and HIT zone. Finally, the performance indicators relative mean velocity $\bar{v}$, best 10 km velocity $v_{10}$ and relative mean heart rate $\bar{hr}$ were calculated.

Differences in the training characteristics between the response and marathon performance time groups were analyzed as follows: We computed a two-way ANOVA with the training parameter being the dependent variable and the response and marathon groups being the independent variables. We excluded $W$, $l_{T90}$ and $l_{D15}$ of the ANOVAs, because
the values of those training parameters were not continuous. For the intensity parameters LIT, threshold and HIT, we analyzed differences in the training process over time by computing repeated measure ANOVAs for the three training zones over the four training blocks. For all ANOVAs, we report partial $\eta^2$ effect sizes if the $p$-values showed statistical significance with a significance level of $\alpha < 0.05$. All statistical tests in this work were conducted using the Python package Pingouin [212].

6.3 Results

6.3.1 Plausibility of the data set

Figure 45 depicts the results for the plausibility of the data set. The distribution of the marathon performance reached from 2.5 up to 6 h. We noticed distinct peaks at the full and half hour marks (Figure 45 a). The histogram of the maximum training heart rate shows normally distributed values between 160 bpm and 220 bpm (Figure 45 b). A high correlation (Pearson’s $r = 0.77$) was found between marathon average velocity and the overall best average 10 km velocity detected within the 16 weeks leading up to the marathon (Figure 45 c). Lastly, a sorted distribution of $\Delta v_{10}$ is presented in Figure 45 d. Values of $\Delta v_{10}$ ranged between -1 and 2 m/s indicating a negative or no improvement in less than a third of the population.

6.3.2 Evaluation of response groups

Figure 46 depicts the verification of the response group definition. Subjects in the high response group, who showed the highest improvements in $\Delta v_{10}$, also showed the highest improvement in $\Delta \bar{v}$, while maintaining a similar mean heart rate. We found a large effect size for the differences of $\Delta \bar{v}$ ($\eta^2 = 0.136$) and a small effect sizes for difference of $\Delta hr$ ($\eta^2 = 0.001$) between the response groups.

6.3.3 Evaluation of training characteristics

Table 14 lists the mean values and standard deviations of the training parameters for subjects in the different response and marathon performance groups over all 16 weeks before the marathon. Besides, the effect sizes of the two-way ANOVA (response group, marathon performance group) for the main effects are reported in case of statistical significance ($\alpha < 0.05$).
Figure 45: Validation of data set. (a) Distribution of marathon performance time $T_{mp}$. (b) Distribution of maximum training heart rate $hr_{max}$. (c) Visualization of correlation between best 10 km velocity $v_{10}$ in the complete training period and marathon performance velocity $v_{mp}$. The blue dots indicate the individual subjects, the green line the linear regression function. (d) Adaptive potential for improvement of best 10 km velocity $\Delta v_{10}$. 
Figure 46: Visualization of the difference between $t_b_4$ and $t_b_1$ of (a) the normalized mean velocity and (b) the normalized mean heart rate for all subjects in the 3 response groups. The velocity values were normalized by the marathon performance velocity $v_{mp}$ and the heart rate by the maximum training heart rate $hr_{max}$.

We did not report effect sizes for the interaction effects, because they were not statistically significant. The results show small effect sizes for the differences between the response groups and higher effect sizes for the differences between the marathon performance groups. Our approach to categorizing subjects into response groups and marathon performance groups yielded a higher number of subjects with a fast marathon performance time in the high response group and in contrast a higher number of subjects with a slow marathon performance time in the low response group.

Figure 47 depicts the share of time spent in the three intensity zones during the four training blocks for the subjects in the different response and marathon performance groups. It shows differences in the intensity distributions between slow, medium and fast runners. We observe an increasing share of training time in the LIT zone from the slow to the fast marathon group. Within the marathon performance group, the overall amount of time spent in the individual zones remains constant.

However, differences in time spent in the intensity zones between the four training blocks were found. Especially subjects allocated in the high response group decreased the time in the LIT zone throughout the training process, while increasing the share of time in the threshold and HIT zone. This is underlined by the results of the repeated-measures
Table 14: Mean and standard deviation of the training parameters for the 16 week training process. The statistical values are reported for the different response groups and the different marathon time categories. $T$ indicates the total training duration, $D$ the total distance, $\overline{HR}$ the relative mean heart rate $\bar{v}$ the relative mean velocity, $v_{10}$ the best 10 km velocity, LIT the share of time in the LIT zone, threshold the share of time in the threshold zone, HIT the share of time in the HIT zone, $W$ the number of workouts, $I_{T90}$ the number of workouts longer than 90 minutes and $I_{D15}$ the number of workouts with a higher than 15 km of the 16 week training process. For all parameters except $W$, $I_{T90}$ and $I_{D15}$, we report the effect sizes $\eta^2_{resp}$ (response groups) and $\eta^2_{mp}$ (marathon performance groups) of the main effect of the two-way ANOVA if the $p$-value was below a significance level $\alpha = 0.05$. The last row of the table lists the number of subjects in each group.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\eta^2_{resp}$</th>
<th>$\eta^2_{mp}$</th>
<th>Marathon group</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$ [h]</td>
<td>-</td>
<td>0.037</td>
<td>slow</td>
<td>52.8 ± 18.7</td>
<td>52.5 ± 17.8</td>
<td>53.1 ± 19.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>55.9 ± 19.3</td>
<td>56.1 ± 17.4</td>
<td>56.8 ± 18.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>61.4 ± 21.3</td>
<td>62.3 ± 19.1</td>
<td>63.0 ± 19.3</td>
</tr>
<tr>
<td>$D$ [km]</td>
<td>0.002</td>
<td>0.132</td>
<td>slow</td>
<td>472.8 ± 164.5</td>
<td>482.9 ± 157.8</td>
<td>495.3 ± 185.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>547.1 ± 178.8</td>
<td>558.3 ± 167.0</td>
<td>559.4 ± 175.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>641.9 ± 213.9</td>
<td>671.6 ± 197.6</td>
<td>677.8 ± 206.7</td>
</tr>
<tr>
<td>$\overline{HR}$ [%]</td>
<td>0.004</td>
<td>0.030</td>
<td>slow</td>
<td>0.81 ± 0.04</td>
<td>0.81 ± 0.03</td>
<td>0.80 ± 0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>0.80 ± 0.03</td>
<td>0.80 ± 0.03</td>
<td>0.80 ± 0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>0.78 ± 0.04</td>
<td>0.80 ± 0.03</td>
<td>0.79 ± 0.04</td>
</tr>
<tr>
<td>$\bar{v}$ [%]</td>
<td>-</td>
<td>0.360</td>
<td>slow</td>
<td>1.09 ± 0.08</td>
<td>1.09 ± 0.09</td>
<td>1.10 ± 0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>1.01 ± 0.06</td>
<td>1.01 ± 0.06</td>
<td>1.01 ± 0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>0.96 ± 0.06</td>
<td>0.96 ± 0.06</td>
<td>0.95 ± 0.06</td>
</tr>
<tr>
<td>$v_{10}$ [m/s]</td>
<td>0.050</td>
<td>0.429</td>
<td>slow</td>
<td>3.05 ± 0.32</td>
<td>3.1 ± 0.27</td>
<td>3.33 ± 0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>3.40 ± 0.29</td>
<td>3.37 ± 0.24</td>
<td>3.49 ± 0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>3.78 ± 0.31</td>
<td>3.73 ± 0.30</td>
<td>3.87 ± 0.29</td>
</tr>
<tr>
<td>LIT [%]</td>
<td>0.003</td>
<td>0.303</td>
<td>slow</td>
<td>0.29 ± 0.17</td>
<td>0.28 ± 0.16</td>
<td>0.28 ± 0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>0.47 ± 0.19</td>
<td>0.47 ± 0.19</td>
<td>0.49 ± 0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>0.63 ± 0.18</td>
<td>0.62 ± 0.18</td>
<td>0.66 ± 0.17</td>
</tr>
<tr>
<td>thr. [%]</td>
<td>0.012</td>
<td>0.151</td>
<td>slow</td>
<td>0.47 ± 0.14</td>
<td>0.47 ± 0.14</td>
<td>0.43 ± 0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>0.43 ± 0.14</td>
<td>0.43 ± 0.15</td>
<td>0.40 ± 0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>0.31 ± 0.15</td>
<td>0.33 ± 0.16</td>
<td>0.29 ± 0.14</td>
</tr>
<tr>
<td>HIT [%]</td>
<td>0.002</td>
<td>0.296</td>
<td>slow</td>
<td>0.24 ± 0.18</td>
<td>0.24 ± 0.19</td>
<td>0.30 ± 0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>0.10 ± 0.09</td>
<td>0.10 ± 0.09</td>
<td>0.11 ± 0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>0.06 ± 0.06</td>
<td>0.06 ± 0.04</td>
<td>0.05 ± 0.05</td>
</tr>
<tr>
<td>$W$</td>
<td>-</td>
<td>-</td>
<td>slow</td>
<td>40.7 ± 13.7</td>
<td>41.5 ± 14.1</td>
<td>41.8 ± 14.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>44.2 ± 14.7</td>
<td>44.9 ± 13.6</td>
<td>46.1 ± 14.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>51.4 ± 18.7</td>
<td>52.7 ± 16.4</td>
<td>53.2 ± 17.5</td>
</tr>
<tr>
<td>$I_{T90}$</td>
<td>-</td>
<td>-</td>
<td>slow</td>
<td>10.8 ± 5.4</td>
<td>10.5 ± 5.1</td>
<td>10.6 ± 5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>11.3 ± 5.4</td>
<td>10.9 ± 5.2</td>
<td>11.1 ± 5.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>11.7 ± 6.1</td>
<td>11.4 ± 5.6</td>
<td>11.5 ± 5.5</td>
</tr>
<tr>
<td>$I_{D15}$</td>
<td>-</td>
<td>-</td>
<td>slow</td>
<td>9.0 ± 4.8</td>
<td>9.3 ± 4.6</td>
<td>9.6 ± 5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>11.0 ± 5.3</td>
<td>11.1 ± 5.3</td>
<td>10.9 ± 5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>12.9 ± 6.5</td>
<td>13.9 ± 6.5</td>
<td>14.1 ± 6.4</td>
</tr>
<tr>
<td>Subjects</td>
<td>-</td>
<td>-</td>
<td>slow</td>
<td>1121</td>
<td>744</td>
<td>394</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>medium</td>
<td>709</td>
<td>823</td>
<td>723</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast</td>
<td>429</td>
<td>685</td>
<td>1143</td>
</tr>
</tbody>
</table>
ANOVA for each combination of response and marathon time category in each zone over the training blocks. In Figure 47 the effect sizes of the statistical tests are indicated by asterisks. Subjects allocated in the high response group revealed the highest effect sizes for differences of time spent in the three intensity zones between the four training blocks.

Differences in training volume parameters between the four training blocks were also analyzed but did not show any significant differences between the three response groups (Figure 48).

6.4 Discussion

In this study, we performed a large-scale retrospective data analysis of runners’ training in the 16 weeks leading up to a marathon. The aim of the analysis was to evaluate differences in training characteristics between different response and marathon performance groups. The data used for the analysis were originated by members of the Runtastic fitness platform who used wearable sensors to track their training progress. From the initial data set of 14773 marathon finishers only 6771 subjects remained after applying filters to improve data quality. In particular, the filter ensuring that the subjects performed the 10 km effort in \( t_{b1} \) and \( t_{b4} \) with an average heart rate \( > 0.8 \cdot h_{max} \) reduced the number of subjects by 6845. We believe that this drastic reduction of more than 50% was necessary to ensure a conclusive analysis.

6.4.1 Plausibility of the data set

By reproducing known values and trends from literature as suggested by Hicks et al. [30], we could verify that our data set can be used for the analysis of differences in training leading up to a marathon. The distribution of marathon performance times is similar to the one presented by Allen et al. [213], including the peaks at the full and half hour marks. Thus, even though the data query only required a workout between 41 and 43 km, the marathon performance times indicate that the workouts were actual marathon races. This assumption is supported by the fact, that 98.6% of the marathon workouts were performed on the weekend. The distribution of the maximum training heart rate \( h_{max} \) shows realistic results similar to data observed by others [214, 215], who determined maximum heart rates using laboratory exercise tests. Thus, we believe that
6.4 Discussion

- slow marathon performance time
- medium marathon performance time
- fast marathon performance time

Figure 47: Visualization of the share of time spent in the intensity zones for (a) slow marathon performances, (b) medium marathon performances and (c) the fast marathon performances. For each marathon performance group, we provide three plots showing the share of time spent in the three zones for the low response, moderate response and high response group. The individual boxes in each plot visualize the IQR within the training blocks. The black horizontal lines within the boxes indicate the median. The whiskers extend to 1.5 IQR. We computed repeated measure ANOVAs for each response and marathon performance category for each intensity zone. The asterisks indicate the effect size of the results: * for $0.01 \leq \eta^2 < 0.05$, ** for $0.05 \leq \eta^2 < 0.12$, *** for $\eta^2 \geq 0.12$. 

139
6 Big Data Analysis on Fitness platform data

Figure 48: Visualization of (a) workout duration $T$, (b) workout distance $D$ and (c) number of workouts $W$ over the four training blocks.
the maximum training heart rate $hr_{max}$ also reflects the actual maximum heart rate well.

Strong correlations between average marathon velocity and average 10 km velocity have been reported by others [200, 211] and are verified by our data. The sorted values for $\Delta v_{10}$ show a heterogeneity in response to training. In comparison to the findings from Bouchard and Rankinen [208], the portion of the population who showed a negative or no improvement in our investigation was higher. We believe that the higher portion was due to the unsupervised nature of the data as well as the low threshold of $> 0.8 \cdot hr_{max}$ we set to verify the best 10 km performances. However, increasing the threshold of $hr_{max}$ to elevate the cardiopulmonary effort for the best 10 km velocities did not change the proportion of training responses.

In comparison to supervised studies from Gordon et al. [66] and Hagan et al. [198], we observed lower weekly mean values in number of work-outs, total training duration and total distance. However, reduced mean values in training volume have also been shown in other unsupervised investigations [146, 216]. Lower training volumes might be caused by the heterogeneous nature of the larger data set itself.

### 6.4.2 Evaluation of response groups

We introduced an approach to assess physical fitness based on the best 10 km velocity $v_{10}$ that was accompanied by a heart rate $> 0.8 \cdot hr_{max}$. We classified three equally large response groups based on observed changes in the average 10 km velocity in $tb_1$ and $tb_4$. The idea of frequently monitoring typical training sessions to evaluate the response to training has already been proposed by Boullosa et al. [209] and appears very practical. This is especially the case when data from recreational runners are analyzed, where laboratory fitness assessments are usually not part of the individual training routine. The 10 km velocity was chosen due to its high correlation to the marathon average velocity [200, 211]. Therefore, we assume that an improvement of $v_{10}$ should also positively influence the marathon performance velocity $v_{mp}$. In addition, a systematic increase in mean normalized running velocity was found when comparing the three response groups from low to high response while no systematical differences in mean normalized heart rate were present. This provides further evidence that in general $\Delta v_{10}$ likely reflects an improved physical fitness, even though the cause for the improvement may vary between
individuals (e.g. improvement due to following a specific training structure with fast runs at the end of the 16 weeks training period). Ultimately, the fact that there were more subjects with a fast marathon performance time allocated in the high response group gives final confirmation that our approach to classify the three response groups based on $\Delta v_{10}$ is reasonable.

### 6.4.3 Evaluation of training characteristics

The evaluation of training characteristics between marathon performance groups revealed differences with medium to large effect sizes. The mean values of all parameters describing the training volume ($D, T, W, I_{T90}, I_{D15}$) are systematically higher for the faster marathon performance time group. Similar relationships were also reported elsewhere [66, 198, 200].

In accordance with others, our results also demonstrate that polarized training with maximized volumes below the aimed marathon velocity in the LIT zone yield better marathon performances [201]. While slow marathon performance times were associated with the largest shares of training time in the threshold zone, fast marathon finishers spend on average more than 60% of their training time in the LIT zone below their average marathon velocity. The larger shares in the threshold zone for the medium and slow marathon groups might be due to the fact that recreational runners cannot control intensity well and tend to run too fast even for prescribed training plans [217].

The mean training parameters in Table 14 showed no differences between the response groups (all $\eta^2 < 0.012$). This implicates that high training volumes do not influence the response to training in general. This should be of interest to novice runners, who are at higher risk to be injured from too high training loads [218, 219]. Nevertheless, the response groups differed regarding the shares of time spent in the three intensity zones throughout the four consecutive training blocks. Independent of the marathon performance time, we observed strong effect sizes for decreasing duration in the LIT zone across the four training blocks for subjects in the high response group. While this observation of course is partly a result of our definition of the response groups, the analysis demonstrates that those subjects who started to train at very low velocities and continuously increased their training velocity up to the actual marathon velocity throughout the 16 weeks responded to the highest extent, leading up to at least an average ($<4$ h 14') or even a fast marathon time ($<3$ h 44').
6.4.4 Limitations

Despite all the filters applied to improve data quality, a study with unsupervised data from fitness platforms cannot be as controlled as a supervised study. For our investigation, we are not able to guarantee that all subjects logged and uploaded all physical activities which could have influenced their 10 km or marathon performance. Contextual information affecting the performance of runners like humidity and temperature during a workout or an injury of a runner were not available. The results are also influenced by the varying accuracy of the different wearable sensors recreational runners use to track their workouts. Running velocity was not adjusted to the elevation profile of the running route, which neglects the impact of inclines and declines to training load. Additionally, phenomena like “hitting the wall” during a marathon [152] were not controlled for, which might cause subjects to be classified in a worse marathon performance category despite a good training process. We acknowledge that these limitations might affect the results of individuals in our analysis. However, we believe that the number of those individuals is low compared to the overall number of subjects and that the effect for most of the limitations are equally distributed over the response and marathon performance groups. Thus, differences between or within groups should not be affected. Nevertheless, a detailed analysis of the influence of those limiting factors on the response to training and the marathon performance shall be conducted in future work.

6.5 Conclusion

In this work, we retrospectively analyzed 16 weeks of training for 6771 marathon finishers. We showed that unsupervised data recorded by wearable sensors are suitable for performing such an analysis by reproducing known trends and values from literature. Our analysis demonstrated that a combination of maximized training volume at velocities below an individual’s marathon velocity, a continuous increase in average running velocity along the complete training period up to final average marathon velocity and high velocity runs ($> 1.2 \cdot v_{mp}$) not accounting for more than 5% of the overall training volume was associated with a higher $\Delta v_{10}$ which likely benefited the marathon performance as well. We also demonstrated that a high training volume does not generally influence the response to training.
The large variances in both the training characteristics and the corresponding responses indicate that the most effective training plan for an individual has yet to be developed. However, coaches and athletes also have to acknowledge that, even with the best and most effective training plan, the potential to improve performance is limited and partially genetically determined.

This study also showed that data recorded by wearable sensors and stored on various fitness platforms are an extremely valuable source for investigating different training regimes retrospectively on large sample sizes. Especially for longitudinal investigations, the limitation of low sample sizes can be overcome. This might enable sport scientists and training physiologists to draw more generalizable conclusions in the future.

### 6.6 Author Contributions

MZ designed the study, implemented the methodology, interpreted the results and wrote the manuscript. CH interpreted the results, wrote and reviewed the manuscript. BD designed the study and reviewed the manuscript. SD exported and anonymized the data set and reviewed the manuscript. KR interpreted the results and reviewed the manuscript. BME designed the study, interpreted the results and reviewed the manuscript. All authors have read and approved the final version of the manuscript and agree with the order of presentation of the authors.
Part IV

Perspectives
7 Discussion

While the content of the contributions was already discussed within the individual papers, I want to discuss some additional ideas concerning the objectives of this thesis. I will start with discussion items regarding the IMU contributions before finishing with items about the Big Data contribution.

7.1 IMU analysis in endurance running

In the contributions in chapter 4 and 5, I showed that running-related spatio-temporal and kinematic parameters can be computed accurately based on foot-mounted IMUs. This advancement will enable sport scientists to conduct interesting research both in the near and far future.

In the near future, I believe that the scientific usage of data from foot-mounted IMUs, but also wearables, in general, will increase because it enables researchers to monitor athletes in the field. While highly accurate gold standard systems like motion capture systems enabled scientists to gain deep insights into the biomechanics of runners in the last decades, it has yet to be proven that these insights also generalize to in-field scenarios. Now, that the accuracy of spatio-temporal and kinematic parameters based on IMU data has been validated not only in this thesis but also by other data scientists [90, 92, 118], sport scientists can start to use these sensors for in-field studies. First studies in this realm already showed promising results: IMUs were used to continuously measure the effect of exhaustion on stride parameters with a higher temporal resolution than ever before during marathon races [220], to create insights into the comfort of running shoes [93] or to show differences in running kinematics on different surfaces [P10, 221]. Besides, I believe that IMUs can lead to new insights into running economy. Running economy is currently assessed using spirometry within the lab and describes the oxygen consumption during a predefined velocity [73]. Research has shown that running economy is influenced by several biomechanical factors [44] and consequently links biomechanics and exercise physiology. The changes of the biomechanical parameters during long races [222, 223] will also influence running economy. Thus, monitoring these changes with IMUs in combination with physiological parameters from other
wearable sensors will also enable a deeper analysis of running economy in-field and within races.

In the far future, I believe that either new platforms will be introduced which also include spatio-temporal and kinematic data from IMUs or that current fitness platforms will be extended by these parameters. While this enables further interesting research questions (see next section), I want to discuss the results of my IMU contributions in this context. Both results showed that the accuracy of the parameters is sensitive to the underlying algorithms and the sensor position. This implicates that only parameters which are computed with the same algorithms based on data of sensors at the same sensor position can be fairly compared. Similar results were found in other sports as well. Mitter et al. [224] showed that the accuracy of devices measuring velocity in power lifting exercises differ and that the devices can consequently not be used interchangeably throughout a training process. This problem can be solved by adding context information on the device the parameters originate from to the databases. Using this information, only parameters of specific positions can be aggregated and compared for specific research questions. The problem with different algorithms can be solved similarly, however, there is another aspect to consider. As mentioned before, most of the current fitness platforms contain GPS and heart rate data. This data is usually acquired and transferred to the platform with a sampling rate of 1 Hz. Based on the transmitted raw data, the platforms use their proprietary algorithms to determine the workout metrics they display to the user. However, IMUs require a sampling rate of at least 200 Hz to accurately determine running-related parameters (see section 3.1.3.1). This will either increase the amount of raw data stored on the platforms drastically or result in a computation of the parameters before being transmitted to the platform. While the first option might be desirable for data scientists, because it would create a large database with IMU data which could be used to further improve algorithms, the second option requires the already described indicator for the device the parameter originates from. With this indicator, the data can be aggregated with respect to the device and it can be assumed that the underlying algorithms are the same.

Another point to bear in mind when evaluating such kind of data at a large scale is that the study cohort is most likely biased and not as broad as the one that is currently transmitting GPS and heart rate data. While most of the current platform data originate from smartphones and smartwatches, which have a large user base because they provide many other services to
the user, a foot-mounted IMU to measure spatio-temporal and kinematic data has (at least currently) only the purpose to extract the parameters. Besides, many recreational athletes might lack the knowledge to interpret these parameters in comparison to running velocity, distance, and heart rate. Thus, only ambitious athletes will take the additional costs to buy such a wearable device and get further insights into their biomechanics, which makes the data biased.

For the accuracy of the parameters, I believe that the accuracy of the trajectory-based parameters is and will remain limited by the zero-velocity update. While minor improvements in the event detection algorithms, in the integration process, and in the sensor hardware will potentially improve the accuracy of the parameters, the problem with the invalid assumption of zero velocity during MS at higher running velocities will remain. Due to this fact, the deep learning algorithms, which do not rely on such an update, are an interesting approach and have the potential to further increase accuracy. As I mentioned in the contribution in chapter 4, the DCNN I used did not generalize well. This issue might be improved by acquiring additional labeled data. However, generating training data is very time-consuming. While in other disciplines like image or speech recognition, the signals can be well interpreted by looking at the images or listening to speech, labeling IMU data by hand is not as easy. However, data augmentation might help to solve this problem [225, 226]. An example for data augmentation for IMU running data was proposed by Dorschky et al. [227] who augmented IMU data by biomechanical simulations. They showed that they could partly improve DCNN models which compute biomechanical parameters from IMU data if they added the simulated data to the training data set.

Besides, other network architectures might also improve the results. While I reimplemented the DCNN network of Hannink et al. [140], RNNs are another type of network architectures, which might provide accurate estimates. RNNs, and especially the LSTM architecture, are currently the gold standard in speech recognition [228]. These networks provide good results in time series analysis because they can encode past information through feedback connections. That this type of network can be successfully applied to IMU data has already been shown for activity classification task [229] or denoising of raw signals [230].

Lastly, I want to briefly mention and discuss the differences between the two publications as well as their limitations. First of all, the number of
subjects differed between the first and the second contribution. While the different algorithms were evaluated on 27 subjects in the laboratory data set, the sensor positions were evaluated on 29 subjects even though the raw data in both studies originated from the same data acquisition. The reason, why the second contribution has two additional subjects, was that problems in the processing of the motion capture ground truth data made me exclude the two subjects from the algorithm evaluation. I was able to solve the problem later, which made it possible to evaluate the sensor position on 29 subjects.

One limitation of both contributions is that the algorithms were all evaluated for runs on flat surfaces. Especially the trajectory algorithm, which relies on the fact, that the accelerometer during midstance can be used to estimate the orientation of gravity, might be affected by a tilted running surface. Similarly, the DCNN network was only trained on data on a flat surface, which will likely cause problems when running data on inclined surfaces is fed to the network. Besides, the stride segmentation algorithms used in the publications might have to be improved for real-world applications. As the evaluation of the segmentation algorithms was not in the scope of this thesis, the sensitivity and specificity of the IC detection algorithms are not known. Especially a high specificity is required because the algorithms were all designed to work with running data. In the case of false-positive IC detections, the resulting parameters computed with the evaluated algorithms might not be precise.

7.2 Big Data analysis in endurance running

With the third paper presented in chapter 6, I could contribute another proof that the data of fitness platforms can be used for scientific investigations. In this context, one integral part of my work was to show the validity of the data set by reproducing known values and trends from literature in order to proof good data quality. Due to the fact that the research branch of wearable-based Big Data analysis is rather new, these validity checks are important for sport scientists to accept the results as robust. The sports science community has traditionally used very accurate and expensive measurement tools to base their findings on. In comparison to the data quality of these tools, the data quality of wearable data of fitness platforms is lower. Also, some of the measures, that sport scientists traditionally use, can not be measured with wearables. However, other metrics might be detectable using wearables which correlate
to the highly accurate ground truth parameters. Showing the validity of these surrogates and providing references for the correlation to the traditional parameters can help to increase the acceptance of wearable data.

Similarly, providing detailed insights about the inclusion/exclusion criteria for the study as well as applied data filters helps to gain trust in the results. Within my work, the first inclusion criteria was to only query data of people who had one workout with a marathon distance between 2017 and 2019. Then I filtered for users, who had at least 16 workouts within the 16 weeks leading up to the marathon. As another inclusion criterion for my study, I only used the data of subjects who performed a 10 km workout with more than 80 % of their maximum detected heart rate $h_{r_{\text{max}}}$. Lastly, I removed outliers using an IQR filter. All these steps drastically reduced the data set. However, this data cleaning was necessary to prevent incorrect conclusions from faulty data. Despite the reduction of the data set by more than 50 %, the number of subjects included in the study is still very high. To the best of my knowledge, the response of training has never been evaluated on a higher number of subjects before.

Hicks et al. [30] identified the access to data to be a significant barrier to Big Data research. Hopefully, more fitness platform providers will follow the trend of recent years and will be open for research collaborations. So far, all data was shared under limited research licenses, which permitted sharing of data with the research community. It would be even better if platform providers would provide an anonymized benchmark data set which researchers could use to conduct studies with different kinds of research questions. An advantage of such a benchmark data set would be that the step of proving validity would only have to be performed once. Also, with the increasing number of published papers based on such a data set, the trust in Big Data based on wearable sensors would grow. This can potentially have a positive impact on other domains like health and medical science.

Another option for a scientific data set would be that a research platform arises, which users of other platforms can donate their data to. An advantage of this approach would be that subjects could sign consent forms when registering for the platform, which enables researchers to have more detailed information on the study cohorts. In the anonymization process of the data, valuable information about subjects is often lost,
which prevents a more detailed analysis of user groups. However, realizing such a platform would require funding for hardware, maintenance, and also marketing to get users engaged.

No matter whether researchers get access to data under a limited research license or via a publicly available data set, information about the wearables the data originates from is beneficial for evaluation purposes. One limitation of my study presented in chapter 6 was that I did not have this information. Thus, I could not address the accuracy of the wearable sensors in detail. Especially for the heart rate recordings, I could not differentiate whether the heart rate data originated from a chest strap or a PPG sensor in a smartwatch. Because the heart rate data was stored as a time series in beats per minute, not only differences in the sensor technology affected the time series, but also the proprietary algorithms the manufacturer used to compute the heart rate time series. Despite the fact that I proved the maximum detected heart rate to be in a reasonable range, I believe that either using the heart rate recordings of a specific sensor type (chest strap, PPG sensor) or a specific manufacturer might have improved the comparability of the heart rate recordings. As a positive example, Emig and Pelton [149] only used data of one type of smartwatch in their analysis. Thus, they could assure that both GPS and heart data of the study cohort are comparable.

In the future, I believe that the data of fitness platforms will be enriched with additional information. Both additional data from wearables (spatio-temporal, kinematic and kinetic) as well as context information about the run (weather, surface of running track based on GPS) will enable even more detailed insights into the biomechanics and physiology of endurance runners in specific conditions [231, 232]. Especially longitudinal studies based on this kind of data will provide new insights into the response of runners to training or running-related injuries, which have not been possible before due to a lack of data [21].
8 Conclusion & Outlook

Within this work, I contributed to the state of the art in wearable computing and Big Data analysis for endurance running. On the one hand, I showed that data of IMUs mounted to the foot can be used to accurately compute spatio-temporal and kinematic parameters. I contributed a study with 29 subjects, which were equipped with four IMUs on the foot and ran over a wide range of velocities from 2-6 m/s. Using this data, I could evaluate the accuracy of different algorithms which were presented in literature for the computation of stride length and running velocity. The strap-down integration approach, which reconstructed the trajectory of the IMU during one stride cycle was found to be the most accurate algorithm. Another advantage of this algorithm in comparison to the other approaches is that both translation and orientation of the foot during a gait cycle are computed, which allows the computation of additional stride features. However, also the deep learning approach using a DCNN showed promising results, which might be further improved by gathering more data and using other network architectures.

Besides, I used the data set to evaluate whether the sensor position of the foot influences the accuracy of spatio-temporal and kinematic parameters. I could show that the sensor position indeed influenced the acquired raw IMU data and, consequently, also on the accuracy of the resulting parameters. In my analysis, a sensor in the cavity of the sole of a running shoe showed the highest accuracy.

In summary, both contributions proved that spatio-temporal and kinematic data from IMUs attached to the foot can accurately be determined. Thus, these parameters can potentially extend the database of fitness platforms, which enables new types of interesting research questions. For example, new insights into running economy might be deduced as running economy is influenced by the biomechanics of a runner. Thus, running economy can now be investigated in-field, which was not possible before. However, my results also show that researchers answering those questions must take care when preparing the data for a Big Data analysis. Ideally, they can work on raw data of one specific sensor position and use one common algorithm to evaluate the data. Otherwise, differences in the data due to varying sensor positions or differences in the algorithms used to compute the parameters of interest can make the interpretation of results difficult.
Additionally, I showed the great potential of data originating from running platforms for answering research questions in sports science. I analyzed a data set of the adidas Runtastic platform, which included GPS and heart rate data of the marathon training process of more than 6771 runners. By monitoring the best 10 km velocity of runners, I created a measure for the individual training response of marathon runners. The results showed, that there is not a single training strategy, that all runners respond to well, but that training response is highly individual. Nevertheless, the polarization of training seemed to positively influence the training response.

In the future, I believe that the amount of data from different kinds of wearable sensors stored on fitness platforms will continuously grow. If the companies running the platforms are open for research collaborations, more and more studies using this Big Data of wearable sensors will emerge. This will add a new dimension to the research fields of biomechanics and exercise physiology because a more detailed analysis of longitudinal effects will be possible. In this manner, insights into running injuries could be evaluated, especially when kinematic and kinetic features can be evaluated over a long time. However, researchers in the respective fields have to accept that the data from wearables are often unsupervised and potentially of lower quality than their current gold standards. Due to this fact, proving the validity of the data in each study will be necessary to gain trust in the results.

Based on my contributions I can conclude that spatio-temporal and kinematic parameters can be accurately determined in endurance running using foot-mounted IMUs. By extending current fitness platforms by this kind of data, a more holistic view of the runners will be possible. This holistic view of large amounts of endurance athletes will allow longitudinal studies, which have not been possible before. Thus, Big Data based on wearable sensors has a large potential to yield exciting new insights into both biomechanics as well as exercise physiology in the near future.
Bibliography


[38] Cavanagh, P. R.: *Biomechanics of Distance Running*. ERIC, 1990.


Smyth, B.: How recreational marathon runners hit the wall: A large-scale data analysis of late-race pacing collapse in the marathon. In: *PloS one* 16.5 (2021), e0251513.


Bibliography


Bibliography


Own publications referring to this work


Body-worn sensors, so-called wearables, are getting more and more popular in the sports domain. Wearables offer real-time feedback to athletes on technique and performance, while researchers can generate insights into the biomechanics and sports physiology of the athletes in real-world sports environments outside of laboratories. One of the first sports disciplines, where many athletes have been using wearable devices, is endurance running. With the rising popularity of smartphones, smartwatches and inertial measurement units (IMUs), many runners started to track their performance and keep a digital training diary. Due to the high number of runners worldwide, which transferred their data of wearables to online fitness platforms, large databases were created, which enable Big Data analysis of running data. This kind of analysis offers the potential to conduct longitudinal sports science studies on a larger number of participants than ever before.

In this dissertation, both studies showing how to extract endurance running-related parameters from raw data of foot-mounted IMUs as well as a Big Data study with running data from a fitness platform are presented.