An Open Data Set of Inertial, Magnetic, Foot–Ground Contact, and Electromyographic Signals From Wearable Sensors During Walking

Desiree Camara Miraldo, Renato Naville Watanabe, and Marcos Duarte
Federal University of ABC

This study describes an open data set of inertial, magnetic, foot–ground contact, and electromyographic signals from wearable sensors during walking at different speeds. These data were acquired from 22 healthy adults using wearable sensors and walking at self-selected comfortable, fast and slow speeds, and standing still. All data are publicly available in the Internet (https://doi.org/10.6084/m9.figshare.7778255). In total, there are data of 9,661 gait strides. This data set also contains files with the instants of the gait events identified using the foot–ground contact sensors and notebooks exemplifying how to access and visualize the data. This data set gives the opportunity to all interested researchers to work with such data, for example, making tests of algorithms for gait event estimation against a common reference, possible.

Keywords: biomechanics, detection, gait event, motion analysis

Falls and injuries while walking are recurrent problems for people who suffer from foot drop. This condition refers to weakness of the ankle dorsiflexor muscles, of which the primary one is the tibialis anterior (TA) impairing, for example, the ability to raise the foot and toes to prevent them from hitting the ground during the swing phase of walking (Stewart, 2008; Westhout, Paré, & Linskey, 2007). When the lesion affects the central nervous system, the electrical excitability of the associated peripheral nerves is likely preserved, so functional electrical stimulation may be used to restore adequate movement patterns among people who suffer from foot drop (Lyons, Sinkjaer, Burridge, & Wilcox, 2002). In such case, a person with foot drop could wear a portable functional electrical stimulation device that would electrically stimulate the TA muscle just before the expected swing phase of the affected inferior limb, evoking a flexion of the ankle and foot during walking. In this scenario, it is fundamental to correctly estimate the moment at which to trigger the stimulus to the muscle (in this case, the instant when the foot should leave the ground under normal conditions).

The authors are with the Biomedical Engineering Program, Federal University of ABC, São Bernardo do Campo, São Paulo, Brazil. Duarte (marcos.duarte@ufabc.edu.br) is corresponding author.
Temporal events that identify different phases in a person’s gait can be consistently defined by the interaction forces between the left and right feet and the ground (see “Methods” section). Accordingly, force sensors embedded at the ground (e.g., force platforms) and wearable sensors (e.g., portable foot switches at the sole) have been used as the gold standard method to detect gait events. Another method to estimate gait events, convenient for its portability, low cost, and practicality, has attracted increasing interest. The method is based on inertial and magnetic wearable sensors integrated as a single inertial measurement unit (IMU) (Rueterbories, Spaich, Larsen, & Andersen, 2010; Trojaniello et al., 2014; Truong et al., 2019). A typical portable IMU that can be applied in gait analysis consists of a microelectromechanical system with gyroscope (angular position rate sensor), an accelerometer (acceleration sensor), and a magnetometer (orientation sensor) with one, two, or three axes at each sensor. Several algorithms based on some or all signals of one or multiple IMUs have been proposed in the literature to estimate specific events during normal and impaired gait for potential real-time use in daily living situations (see the Supplementary Material [available online] for a brief review of recent articles on this topic since 2010 and [Rueterbories et al., 2010] for review of earlier articles). Nevertheless, despite the intense development in this field, there is no accepted robust algorithm for gait event estimation based on IMUs for real-time applications.

A limitation to the development of robust methods for gait event estimation is that there is no public data set consisting of raw signals to enable gait analysis (e.g., force data from foot–ground interactions) and raw IMU signals that researchers could use to test algorithms. We are aware of two open data sets with some of these characteristics: the MAREA gait database (Khandelwal & Wickström, 2017) contains data from foot-switch sensors (foot–ground contact data) and from accelerometers, but not gyroscope or magnetic sensor data. The OSHWSP gait data set (Llamas, González, Hernández, & Vegas, 2016) contains data from triaxial accelerometers and gyroscopes, but not foot–ground contact data. Thus, there is a demand for an open data set of gait event-related signals, which should contain data on different walking speeds, because various speeds are present in daily life activities. Data on the timing of TA muscle activation in healthy subjects during walking could also be useful for gait event estimation and the development of a functional electrical stimulation device for people with foot drop. This activation timing can be measured via surface electromyography. A robust public gait data set would make it possible to test algorithms against a common data set, improving the replicability and transparency of such studies and attracting research groups around the world to the problem of gait event estimation, which has otherwise been inaccessible owing to lack of data.

In this context, the goal of this study was to create an open data set of inertial, foot–ground contact, and electromyographic data during walking at different speeds.

**Methods**

**Subjects**

A convenience sample recruited from students and employees at the Federal University of ABC comprised 22 healthy subjects (10 males and 12 females) who voluntarily participated in this study. These subjects averaged (±1 SD) 28.1 ±
7.4 years of age, 71.1 ± 12.0 kg of body mass, 169.6 ± 10.5 cm of height, and 24.7 ± 2.8 kg/m² of body mass index. Data for each subject are presented in the open data set (see “Results” section on how to access it). In addition, using the same methods, we also collected data for one adult with a foot drop gait abnormality. Since this study focuses on the data of healthy subjects, information about this additional subject and the acquired data are presented only as Supplementary Material (available online) and in the open data set, and it will not be discussed any further here. This study was approved by the local ethics committee of the Federal University of ABC (CAAE: 53063315.7.0000.5594), and all subjects signed a consent form prior to data collection.

Gait Events

The gross movement patterns of a healthy person’s gait are cyclic, so the sequence of events that describes the gait is repeated after a certain period (Alexander, 2004; Whittle, 2007). The movement patterns of a walking gait can be described in more detail when divided into events and phases, as illustrated in Figure 1. A complete normal gait cycle (or stride) begins and ends with the same event, usually the initial ground contact of the leading limb, a heel strike (HS). Then the leading limb takes over the body weight by placing the whole foot on the ground, in a subphase called loading response. The moment when the toes touch the ground is the toe strike (TS). Next, in the midstance subphase, the body is moved forward while the opposite limb is in the swing phase. The heel loses ground contact, the heel-off (HO) event, and the body continues to be propelled forward until the preswing subphase starts. Still due to this propulsion, the toe-off (TO) event occurs when the toes leave the ground, starting the swing phase. During the swing phase the swinging limb is accelerated forward, then it passes the opposite limb (midswing subphase) and is decelerated until the next HS event, which will finish the swing phase. The phase from the HS event until the TO event of the same lower limb is

Figure 1 — Gait events, phases, and subphases of a typical walking gait cycle. Letters L or R before the event names indicate the lower limb side, left (L) or right (R), respectively. The horizontal stripes illustrate the stance phase of each lower limb. LHS = left heel strike; LHO = left heel-off; LTO = left toe-off; LTS = left toe strike; RHS = right heel strike; RHO = right heel-off; RTO = right toe-off; RTS = right toe strike.
the support phase for that lower limb. A stride starting and ending with the left HS (LHS) consists of one right step followed by one left step; the inverse occurs for a gait stride starting and ending with the right HS (RHS).

Data Acquisition

To measure the inertial variables and electrical activity of the TA of both legs and the contact of the heel and toe of both feet with the ground, we employed a commercial off-the-shelf integrated solution composed of six wireless wearable units (three in each leg/foot) and one portable data logger (Trigno EMG System, Trigno Personal Monitor; Delsys Inc., Natick, MA), as shown in Figure 2. The first unit (Trigno IMU with 10 channels; Delsys Inc.), referred here as IMU + electromyographic (EMG) unit, had a triaxial accelerometer (with a sampling period of 6.75 ms/sample per channel), a triaxial gyroscope (6.75 ms/sample per channel), a triaxial magnetometer (13.5 ms/sample per channel), and an EMG channel (900 ms/sample per channel). All these sampling periods are preset by the manufacturer of the IMU + EMG sensor and cannot be altered. This IMU + EMG sensor was fixed to the shank, over the belly of the TA muscle (referred as “taR” or “taL,” at the right or left side, see Figure 2). Skin preparation and sensor placement were performed according to the Surface Electromyography for the Non-Invasive Assessment of Muscles recommendations (Hermens, Freriks, Disselhorst-Klug, & Rau, 2000). A second IMU + EMG sensor was fixed to the forward flat part of the tibia bone, aligned with its long axis at the same height of the first IMU + EMG sensor (referred as “tbR” or “tbL,” at the right or left side, see Figure 2). The third

Figure 2 — (a) Placement of the wearable units on the subjects’ legs. “taR” and “taL,” are the IMU + EMG units over the right and left tibialis anterior muscles, “tbR” and “tbL,” are the IMU + EMG units over the tibia bones, and the other two units are the FSR adapters. The coordinate directions x, y, and z of the IMU + EMG units are indicated for the “tbR” unit. In a standing position with the feet parallel to each other, the y-axis was approximately vertical, and the positive direction pointed downward, the z-axis was approximately horizontal and the positive direction pointed forward, and the x-axis direction can be found by the right-hand rule. (b) Placement of FSR sensors on the heel and toe of the right foot. FSR = force-sensitive resistor; EMG = electromyographic; IMU = inertial measurement unit.
unit (Trigno 4-Channel FSR Adapter; Delsys Inc.) was connected with two force-sensitive resistor (FSR) sensors to measure the heel and toe contacts with the ground, with a sampling period of 6.75 ms/sample per channel. Each FSR sensor (a 1.5-cm diameter circular pad) was fixed underneath the toe and heel with double-sided adhesive tape. The remaining three wearable units were fixed to the other leg and foot in a similar configuration. The data from the extra IMU + EMG unit attached on each leg (“tbR” and “tbL,” see Figure 2) are intended for potential future studies, for example, about signal reproducibility and sensitivity to sensor positioning for gait event estimation with inertial and magnetic signals. The data from these extra units are also available in the open data set and will not be further explored in this text. The exact axes of orientation of the IMU + EMG sensors depended on the subject’s leg shape and how he or she walked, and in a standing position with the feet parallel to each other, the y-axis was approximately vertical and the positive direction pointed downward, the z-axis was approximately horizontal and the positive direction pointed forward, and the x-axis direction can be found by the right-hand rule (see Figure 2). The data logger was fixed with a belt to the subject’s waist. A software code for the data logger managed the data acquisition, and after the session, the data were uploaded to a computer in a single file for each trial via the data logger’s software (version 4.3; EMGworks, Delsys Inc.). The data logger acquired signals from the different sensors at the different sampling periods cited earlier, but for each trial, the data stored in the file have the respective timestamps (a column with time in seconds for each signal).

Task

After the sensors were attached to the subject and the task was explained, the subject walked six times at each of three self-paced speeds (comfortable, slow, and fast) on a 40-m long and 2-m wide walkway with a flat, rigid surface. (Trials 1, 3, and 5 were in one direction, and Trials 2, 4, and 6 were in the opposite direction.) The order of the speeds was randomized among subjects. Each trial lasted from 30 to 60 s. In addition, one trial was acquired with the subject standing upright and as still as possible for 10 s, in case someone wants to perform a calibration of the sensors. Data collection for each subject was performed in a single session, which lasted about 40 min.

Preprocessing

All subsequent steps, including file reading and writing, data processing, analysis, and visualization, were implemented in the Python language. There were a few instances of missing data for short periods (less than 100 ms) during the data collection, probably related to wireless transmission, and these missing data appear as zeros in the files. Missing values were identified, and the data were reconstructed by linear interpolation. After this reconstruction, the data were filtered with different frequency cutoffs based on the original sampling and data characteristics: accelerometer and gyroscope data were low-pass filtered with a 60-Hz cutoff frequency; data from the magnetometer were low-pass filtered with a 30-Hz cutoff frequency, and EMG data were band-pass filtered between 20 and 450 Hz. These four signals were filtered using a fourth-order zero-phase Butterworth filter. Due to
its impulse response characteristic, the data from the FSR sensors were low-
pass filtered with a second-order zero-phase critically damped filter with a 30-Hz
cutoff frequency. Subsequently, all data were resampled to a common frequency
of 1000 Hz using a polyphase algorithm (function “resample poly” from the
SciPy Python library). The amplitude of the FSR data was normalized to the
interval 0–1 for each trial. Finally, all data for each trial were saved in a text file
and are available in the open data set. A flowchart with the preprocessing steps is
presented in Figure 3, and an example of the data from the open data set is shown in
Figure 4.

Figure 3 — Flowchart of signal preprocessing steps applied to the data. Note that all
files with raw data (the data collected) and all files recorded in the last step indicated in
the flowchart are available in the open data set. EMG = electromyographic; FSR =force-
sensitive resistor.
Detection of Gait Events

Data from the FSR sensors under the heel and toe of the right and left feet were employed to identify the following gait events (see “Methods” section): right and left heel strike (LHS) and left toe-off (LTO). The figure illustrates the raw data of part of one trial for the measured signals: FSR = contact forces at the heel and toe; EMG TA = tibialis anterior electromyographic activity; Accel. = acceleration; Ang. vel. = angular velocity; Mag. field = magnetic field; a.u. = arbitrary units. Data are from subject “s03,” Trial 1, comfortable speed, left side.

**Figure 4** — Exemplary raw data of part of one trial for the measured signals. FSR = heel and toe contact forces; LHS = left heel strike; LTO = left toe-off; EMG TA = tibialis anterior electromyographic activity; Accel. = acceleration; Ang. vel. = angular velocity; FSR = contact forces at the heel and toe; Mag. field = magnetic field; a.u. = arbitrary units. Data are from subject “s03,” Trial 1, comfortable speed, left side.

**Detection of Gait Events**

Data from the FSR sensors under the heel and toe of the right and left feet were employed to identify the following gait events (see “Methods” section): right and
left HS, HO, TS, and TO. However, the FSR data often presented fluctuations at the baseline, so before the event detection, this fluctuation was subtracted from the original data, yielding moving minimum filtered data (using the function “move_min” of the Bottleneck Python library with a window size of 500 points [0.5 s] in a dual-pass forward and backward filtering to not introduce any phase lag). To detect gait events from the FSR data, we employed a Python function “detect_onset.py” (available with the data set), which performs onset detection based on an amplitude threshold method with a parameter specifying a minimum number of samples above threshold to detect as onset, and other parameters to tune the detection. (After a few trials with visual inspection of the function output, the following parameters were chosen: threshold = 2, n_above = 1, n_below = 2, threshold2 = 3, n_above2 = 1.) The gait events are stored in a separate file as indices corresponding to the rows in the sensor data files for a given trial, and they are also made available in the open data set.

Data Visualization

To visualize patterns in the measured signals, we computed average data across subjects. First, an estimation of the EMG power was calculated using a moving RMS filter with a window of 100 points (0.1 s). The other signals were low-pass filtered with a fourth-order zero-phase Butterworth filter and a 10-Hz cutoff frequency. Second, for each trial, we segmented the data in cycles or strides (see “Methods” section). The data of each stride were normalized in time from 0% to 100% in steps of 1% and averaged across trials to obtain the mean gait cycle for the given subject/condition. The mean and SD of the gait cycle across subjects were calculated, repeating the same process with the data of all subjects.

Results

The data set, consisting of all the data of the 22 healthy subjects, is available as an open repository accessible on the Internet (https://doi.org/10.6084/m9.figshare.7778255), under the CC0 license (https://creativecommons.org/publicdomain/zero/1.0/). The data set contains two types of data: (a) the raw data of all subjects compressed in a single file (named “all_subjects_raw_data.zip”), and (b) processed data of all subjects as described in the “Methods” section (with filtering, resampling, and event identification). These processed data are described next.

The data set for the subjects contains data for a total of 9,661 gait strides, with walking speed varying from 0.63 to 2.46 m/s and stride length varying from 0.93 to 2.22 m (see Figure 5 for box plots of these data). Figure 6 shows plots of the ensemble averages over the gait cycle at comfortable speed for the following variables: EMG activity of the TA muscle, three-dimensional acceleration, and angular velocity of the left and right legs. Because the subjects walked six times each, with Trials 1, 3, and 5 in one direction and the other trials in the reverse direction, we did not compute the ensemble average across trials for the earth’s magnetic field data.
All the data in the data set are stored in ASCII (text) format and can be downloaded separately for each subject or as a single compressed file. The data set has three types of contents: data of the measured signals (data files), data of the gait events (event files), and metadata about the subjects (metadata file).

The data file contains a time column ("Time," in seconds), along with tab-separated columns with data from the four IMU + EMG units over the TA ("ta") muscle and over the tibia bone ("tb") of the right ("R") and left ("L") legs with triaxial ("x," "y," "z," see “Methods” section for the axis convention) accelerometers ("ACC," in units of gravitational acceleration); triaxial gyroscopes ("GYR," in degrees/s); triaxial magnetometers ("MAG," in mT); EMG of the TA muscle ("EMG," in mV); and from the FSR (in arbitrary units, normalized from Figure 5 — Violin plots (box plots plus kernel density estimations) across the 22 subjects from the open data set for the spatiotemporal parameters: stride duration, support duration, stride length, and speed calculated using the data from the force-sensitive resistor under the right foot for the different walking speeds. The numbers shown at the top of each column indicate the total number of gait strides available in the data set at each speed (and used to generate these plots). For each variable, the curve represents an estimation of the data distribution, the vertical black line represents the interval for 95% of the data, the black box represents the interquartile range, and the central dot represents the median value.
0 to 1) under the heel (“hs”) and toe (“to”) of both feet, resulting in a total of 43 columns, all sampled at 1000 Hz. Accordingly, each file has the following header indicating the type of data in each column:


Figure 6 — Mean ±1 SD across the 22 subjects from the open data set for the measured signals EMG TA, Accel., and Ang. vel., over the left and right gait cycles walking at a comfortable speed. The mean ±1 SD termination of the gait support phase, indicated by the LTO or RTO events, are shown by the vertical lines and shaded areas of the plots. The gait events were determined using the data from the force-sensitive resistor under the right and left feet. These curves are based on a total of 3,222 gait strides. EMG TA = tibialis anterior electromyographic activity; Accel. = three-dimensional acceleration; Ang. vel. = angular velocity; LHS = left heel strike; RHS = right heel strike; LTO = left toe-off; RTO = right toe-off.
These files are named “s<nn><c><t>.txt,” where <nn> refers to the number of the subject from “01” to “22”; <c> refers to the walking speed (“c”: comfortable, “s”: slow, or “f”: fast); and <t> refers to the trial (from “1” to “6”). For each of the 22 subjects, there are three speeds and six trials (18 files), plus one file for the standing still task (named “s<nn>up.txt”).

The event file contains the indices (the line numbers in the corresponding data file) for the following eight gait events (see “Methods” section) identified in the header: right HS, right HO, left TS, left TO, left HS, left HO, right TS, and right TO. The name of the event file follows the same convention as that of the data file, followed by the letters “ev” at the end of the file name. There is one event file for each walking trial.

The metadata file, named “info.txt,” contains the subjects’ numbers and information from their anamneses. Following is the coding for the metadata (the first word identifies the name of the column in the header):

- Subject: number of the subject (from “01” to “22”).
- Gender: gender (‘F’ or ‘M’).
- Date of Birth: date of the subject’s birth (dd/mm/yyyy)
- Age: subject’s age in years, months, and days.
- Illness: whether the subject has any self-declared illness (“Yes” or “No”).
- Illness 2: type of illness (“No” if the subject does not have any illness).
- Mass: mass in kg (measured with a calibrated scale).
- Height: height in cm (measured with a calibrated stadiometer).
- BMI: body mass index in kg/m².
- Date Acquisition: date of the subject’s evaluation (dd/mm/yyyy).

**Discussion**

In this study, we presented an open data set of inertial, magnetic, electromyographic, and foot–ground contact data from wearable sensors placed on both legs and feet during walking at different speeds and standing still. Data were generated by 22 healthy subjects. The open data set contains both raw and processed data sampled at 1000 Hz totaling 9,661 gait strides of the 22 healthy subjects, along with data from the same sensors of each subject standing still.

All subjects exhibited similar intra- and interindividual patterns in terms of spatio-temporal parameters, EMG activity of the TA muscle, and three-dimensional linear acceleration and angular velocity of the left and right legs at different gait speeds. All variables were affected by walking speed, as a whole, higher speed, shorter stride and support duration, longer stride length, and greater peak-to-peak amplitude of EMG and inertial variables, consistent with the literature (Fukuchi, Fukuchi, & Duarte, 2019). Qualitatively, the EMG activity of the
TA muscle was similar between the left and right sides and exhibited peaks of activation at the HS and TO gait phases, also consistent with the literature (Winter, 1991). The angular velocity and linear acceleration of the two legs depend on the exact alignment of the sensors (which is generally not a concern for the development of algorithms for estimating gait events). In this study, the axes of the sensors were not aligned to the main planes of movement, and a direct comparison with the literature is difficult.

Also included in the data set are data files with indices of the actual gait events for each stride and a file with information about the subjects’ health characteristics. This open data set can be used in future studies related to gait event estimation based on inertial sensors. In addition, using the same methods, we collected data for one adult with a foot drop gait abnormality, and these data are also available in the open data set (see Supplementary Material [available online] for more information).

The only two other data sets of inertial gait signals openly available are useful but include only limited signal types, limiting their application. The MAREA gait database (Khandelwal & Wickström, 2017) comprises triaxial accelerometer data and foot–ground contact of 20 healthy subjects in different gait activities, but no other inertial signals. The OSHWSP gait data set (Llamas et al., 2016) contains data from triaxial accelerometers and gyroscopes of 12 healthy subjects walking at self-selected speed, but not foot–ground contact data.

This study only acquired data from individuals walking in a straight line at three constant speeds and in a very predictable environment. Although these controlled characteristics certainly helped to obtain very consistent patterns for the investigated variables and can facilitate the reproducibility of the data, the current data are very limited for the estimation of gait events if we consider the variety of ways we walk day by day. Another limitation was the use of commercial and relatively expensive sensors that may impose difficulties in extending the data set by other research groups.

The open data set we presented in this study, with data of inertial, magnetic, electromyographic, and foot–ground contact data from wearable sensors during walking at different speeds and standing still, is unique in the literature given the number of different sensors used, tasks investigated, and the number of gait strides available. This open data set will give the opportunity to all interested researchers to work with such data and will enable them to develop and test algorithms for gait event estimation against a common reference, potentially improving the replicability and transparency of those studies.

**Acknowledgments**

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP/Brazil) under grant 2015/14810-0 and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brazil).

**References**


(Ahead of Print)


(Ahead of Print)