Article:
Ena, A., Mazzà, C., Rodríguez-Romero, A. et al. (2026) Accurate quantification of steps from multiple smartphone positions. Scientific Reports, 16 (1). 4143. ISSN: 2045-2322

https://doi.org/10.1038/s41598-025-34270-2

scientific reports

Check for updates

OPEN

Accurate quantification of steps from multiple smartphone positions

Alejandro Ena[1], Claudia Mazzà[1,2], Antonio Rodríguez-Romero[1], Tim Woelfle[3,4], Jannis Mueller[3,4], Cristina Granziera[3,4,5], Ludwig Kappos[3,4], Emmanuel Bartholomé[1], Corrado Bernasconi[1], Shibeshih Belachew[1] & Óscar Reyes[1]✉

A key challenge in smartphone-based assessment of motor capacity is that patients may wear their smartphone in varying positions, while state-of-the-art algorithms have not been designed and validated for multiple device locations. This paper proposes a solution to estimating foot-to-ground initial contacts (ICs) during gait using inertial measurement unit (IMU) sensor data collected from a smartphone agnostic of its location in a cloth front or back pocket. FAIR-Q, an algorithm originally validated for data collected from the lower trunk was further tuned for this intended use, and tested on IMU data collected in cloth pocket positions from both healthy adults ($n = 83$, age range: 20–83 y.o.) and people with Multiple Sclerosis ($n = 50$, age range: 22–61 y.o., EDSS score: 0–6) during a 30s walk test. The performance of FAIR-Q was compared against a gold standard multi-sensor device in terms of sensitivity and measurement error in identifying ICs and measuring step and stride durations. Excellent performance was achieved for both groups in all tested conditions (test-level relative errors for duration measures < 1%) and using data from a large variety of smartphone devices, supporting the method's suitability for high-frequency smartphone-based assessment of gait capacity in neurological diseases.

Keywords Gait analysis, Initial contacts, Pocket smartphone placement, Time frequency analysis, Multiple sclerosis

Smartphone-based assessment of motor capacity by means of inertial sensor data recorded during the execution of structured motor tasks has gained popularity for the longitudinal monitoring of people with neurological and muscular diseases[1]. This is particularly true in the case of Multiple Sclerosis, a chronic immune-mediated disease of the central nervous system characterised by demyelination and progressive impairment of motor function. Because gait disturbances are among the most common and disabling symptoms in people with Multiple Sclerosis (pwMS), objective, sensor-derived measures of walking performance have become increasingly important for supporting clinical decision-making, monitoring rehabilitation progress, and evaluating response to disease-modifying therapies[2–5].

This assessment typically entails the development of an application (App) that is used to provide detailed instructions to guide the patients through the execution of structured smartphone-instrumented tests (similar to those typically included in a clinical assessment of the same functions). The same App is then used to record and remotely transfer the data collected during the tests by the inertial measurement units (IMU) sensors embedded in the smartphones. This approach offers the opportunity of more frequent assessments, which better reflect patient's performance in their natural environment. One of the challenges, however, is that in unsupervised conditions smartphone handling becomes a critical factor to control for. In gait measurement applications, for example, patients might find it preferable to place their smartphone in a pocket, even when instructed to use a provided belt to enable a median location in the lower trunk[1,4,6]. This implies that to extract relevant information, the data need to be processed by algorithms capable of detecting steps (interval between initial contact with the ground of one foot and initial contact of the contralateral foot) as well as strides (the interval between two subsequent initial contacts of the same foot) from highly variable input raw signals collected from different smartphone locations.

[1]Indivi AG, Basel, Switzerland. [2]School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, UK. [3]Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel, Basel, Switzerland. [4]Department of Neurology, University Hospital Basel, Basel, Switzerland. [5]Department of Biomedical Engineering, University of Basel, Basel, Switzerland. ✉email: oscar.reyes@indivi.io

Many algorithms have been proposed and validated for detecting foot-to-ground initial contacts (ICs) and quantifying both step and stride dynamics from an IMU sensor located on the lower trunk in a median position. Simpler methods such as thresholding or zero-crossing detection[7,8] rely too heavily on predefined thresholds or peak detection algorithms. Template matching methods[9] are a more robust alternative but require predefined step cycle templates. Wavelet-based methods proved accurate to capture gait phases[10] but still have limitations when applied to free-living gait data in a real-world environment[11]. Advanced machine learning methods[12] can offer higher accuracy and adaptability for irregular gait but require substantial computational resources and large-annotated datasets. Overall, state-of-the-art methods for step detection still suffer from the assumption of a steady walking cadence and typically rely on the IMU sensor being placed near the body's center of mass. This poses significant challenges when applied to impaired gait patterns and unsupervised recordings[13].

Comparatively, the current state-of-the-art in detecting steps and strides from IMU data collected with a sensor positioned proximally on the thigh (i.e. when a smartphone is in a pocket) remains underdeveloped. Recent studies have demonstrated the feasibility of detecting stride events using smartphones placed in the trouser pocket during both treadmill walking[14] and level-ground walking[15,16]. Some of these approaches have also been applied to pathological populations, including individuals with lower limb joint pain[17] and with Parkinson's disease[18]. While these methods have demonstrated satisfactory accuracy in stride detection, they are generally designed to identify initial contacts (ICs) from a single leg, hence only allowing for stride detection, or require the use of two devices - one per leg - for bilateral detection. Furthermore, some of these approaches rely on long, uninterrupted walking bouts to achieve acceptable performance[14], limiting their applicability both in free-living and in clinical environments. Despite their promise, these methods have not yet been validated for use with data collected from pwMS.

A Frequency-Adaptive, Iterative and Robust Quantification (FAIR-Q) method was recently validated on lower-trunk real-world gait data for the purpose of segmenting steps without relying on specific signal shapes or templates[19]. When adequately tuned, the method proved accurate when applied to real-word data collected from a research-grade device, across normal and impaired gait patterns. This method leverages a time-frequency analysis that enhances the temporal allocation of events over gait cycles. This characteristic makes it an excellent candidate for adaptation to other gait cyclic signals, with enhanced robustness to the location and quality of the sensors embedded in the device. This paper aims at testing this hypothesis by assessing the accuracy of FAIR-Q in detecting the foot-to-ground ICs and consequently calculating step and stride duration and cadence in IMU data collected from a smartphone located either in a front pocket (FrP) or in a back pocket (BP). More specifically, the performance of the algorithm is evaluated against a reference motion capture system in both healthy controls (HCs) and pwMS.

Methods

Subjects and experiments

Data included in the study derived from two cross-sectional studies, involving 83 HCs and 50 pwMS. Expanded Disability Status Scale (EDSS) total scores, functional system subscores (FSS), and Ambulation Scores were assessed by trained and certified raters in accordance with Neurostatus standardized examination and scoring procedures[20].

All participants (demographics in Table 1) performed a supervised, structured 30-second walk test on a flat indoor surface. The walking path differed between groups: pwMS walked in a straight line along a long hospital corridor, while HCs followed a curvilinear path within a 5-meter-sided square room. Participants in both groups were instructed to walk at their self-selected pace without stopping or interruptions, and to avoid abrupt turns as much as possible. To ensure safety, particularly for pwMS with higher levels of disability, all assessments were closely monitored by clinical staff walking alongside the patients.

During these tests, inertial sensor data were simultaneously collected by a reference system (Awinda, Xsens[21], sampling frequency 60 Hz) and a smartphone (sampling frequency 50 Hz). The Xsens reference system consists of 17 IMUs, each equipped with 3D accelerometers, gyroscopes, and magnetometers. These sensors were attached to specific body segments - head, shoulders, upper arms, forearms, hands, upper legs, lower legs, and feet - using straps, and these units provided the Xsens system with the data required to reconstruct full-body 3D kinematics. Gait events were derived primarily from the lower-limb sensors by processing the foot-contact information provided by the Xsens system. Additionally, one IMU was rigidly attached to the back of the smartphone for synchronization and alignment purposes.

Each HC performed four 30-second walking tests, corresponding to four smartphone placement conditions: FrP left, FrP right, BP left, and BP right. For each participant, two tests were recorded with an Android device and two with an iPhone, ensuring that each type of hardware and operating system was used once in the front pocket and once in the back pocket. The assignment of Android and iPhone devices to the left or right side was randomized across participants but stratified at the group level to ensure a similar number of recordings for left and right legs. Devices were randomly selected from a pool of 10 Android smartphones (from 7 different manufacturers, Android version ≥ 9 and ≤ 12) and 7 different iPhone models (iOS version ≥ 15.7 and $\leq 16.4.1$). This procedure resulted in a total of 332 smartphone recordings (BP, $n = 166$; FrP, $n = 166$), matched by an equal number of Xsens IMU recordings.

The pwMS performed a single 30-second walking test, during which IMU sensor data were recorded using the participant's own smartphone, placed in the front pocket - either left or right leg, without standardization. Among the 50 pwMS, 33 used iPhone (12 different models, iOS version $\geq 14.7.1$ and $\leq 16.1.2$) and 17 used Android smartphones (from 5 different manufacturers, Android version ≥ 8 and ≤ 13). This procedure resulted in 50 smartphone recordings, matched by an equal number of Xsens IMU recordings.

	HCs		PwMS	
	Female (*N* = 45)	Male (*N* = 38)	Female (*N* = 28)	Male (*N* = 22)
Height (cm)				
Mean (SD)	164 (5.90)	175 (7.06)	168 (6.48)	183 (5.05)
Median	165	175	168	182
Range [Min, Max]	[149, 175]	[160, 188]	[156, 184]	[176, 194]
Weight (kg)				
Mean (SD)	67.3 (13.6)	80.7 (14.8)	78.0 (20.7)	85.0 (10.4)
Median	64.0	79.1	76.0	86.0
Range [Min, Max]	[48.6, 106]	[53.5, 122]	[43.0, 133]	[63.0, 102]
Age (year)				
Mean (SD)	43.7 (14.8)	42.4 (16.8)	43.4 (10.5)	42.2 (10.5)
Median	44.0	45.0	43.0	42.5
Range [Min, Max]	[22, 77]	[20, 83]	[25, 61]	[22, 61]
Disease duration (year)				
Mean	-	-	12.00	8.44
Median	-	-	12.29	6.98
Range [Min, Max]	-	-	[0.42, 32.12]	[0.13, 27.94]
EDSS				
Mean (SD)	-	-	2.3 (1.3)	2.1 (1.3)
Median	-	-	2.25	1.75
Range [Min, Max]	-	-	[0, 5.5]	[0, 6.0]
Pyramidal FSS				
Mean	-	-	1.29	1.14
Median	-	-	1.0	1.0
Range [Min, Max]	-	-	[0.0, 4.0]	[0.0, 4.0]
Cerebellar FSS				
Mean	-	-	0.75	0.59
Median	-	-	1.0	0.0
Range [Min, Max]	-	-	[0.0, 3.0]	[0.0, 3.0]
Ambulation score				
Mean	-	-	0.43	0.45
Median	-	-	0.0	0.0
Range [Min, Max]	-	-	[0.0, 4.0]	[0.0, 7.0]
T25FW (second)				
Mean	-	-	5.29	4.66
Median	-	-	4.74	4.33
Range [Min, Max]	-	-	[3.25, 11.28]	[3.27, 12.3]
SDMT score				
Mean	-	-	61.34	60.82
Median	-	-	59.0	57.0
Range [Min, Max]	-	-	[40.0, 100.0]	[43.0, 88.0]

Table 1. Demographic characteristics of the healthy controls (HCs) and people with multiple sclerosis (pwMS). SD, standard deviation; EDSS, expanded disability status scale; FSS, functional system Scale; SDMT, symbol digit modalities Test; T25FW, timed 25-Foot Walk.

Data technical verification

The reference system data were downsampled to 50 Hz, and temporal synchronization between the Xsens and smartphone signals was achieved using a maximum cross-correlation approach applied to sliding windows of the three gravity acceleration components obtained from the smartphone and the reference IMU sensor rigidly attached to it.

After signal temporal alignment, a comprehensive technical verification was conducted to assess data quality and select the final test cases. This included checks for missing values, signal duration, sampling frequency stability, and synchronization among the Xsens IMUs[22]. Additionally, segment kinematics reconstructed by the Xsens 3D model and the detected foot contacts were visually inspected to ensure the absence of anomalies during task execution. The same checks were applied to both HCs and pwMS datasets and were performed by at least two operators. The ICs provided by the reference system served as the reference for evaluating the accuracy

of ICs estimated by FAIR-Q. Following individual case verification, 330 test execution recordings from HCs (FrP, n = 165; BP, n = 165) and 49 from pwMS were selected for subsequent analyses.

Frequency-adaptive, iterative and robust quantification (FAIR-Q) method for step detection

Figure 1 summarises the key steps of the FAIR-Q method, which was used to identify the occurrence and timing of the ICs between the foot and the floor. The method has been described in detail elsewhere[19], where it was tuned and validated to process data collected from an IMU sensor located on the lower back. In brief, a Synchrosqueezed Wavelet Transform (SSWT)[23] is applied to a previously filtered vertical component of the acceleration signal (Fig. 2a). This method has been chosen due to its performance localising the energy along the time, enabling accurate tracking of frequencies along the complete signal. The obtained time-frequency representation (TFR) is then processed to build a "*ridge*" (Fig. 2b), defined as the sequence of points in the TFR that effectively captures the evolution of motion-related frequencies and preserves the most significant spectral structures of the signal. The ridge is further analyzed to detect large discontinuities, or jumps, which are used to segment the signal into frequency-stable intervals. The subsequent analysis of this ridge allows identifying local frequencies (Fig. 2c) within each segment. These frequencies are used to adaptively smooth the original signal according to the actual cadence, resulting in a series of potential ICs, which are then pruned of false positives using a peak detection and prominence approach (Fig. 2d).

Previous work[19] provided a systematic framework for selecting the parameters shown in Fig. 1, and an extensive sensitivity analysis demonstrated that the algorithm maintains reliable step-detection performance across a wide range of parameter values. Using this framework, the optimal parameter values for real-world gait analysis with a lower-back sensor were identified as follows: $\boldsymbol{f_l}$ = 0.25 Hz; $\mathbf{f_u}$ = 2Hz; $\mathbf{w_d}$ = 3s; $\boldsymbol{d_t}$ = 2, $\boldsymbol{l}$ = 2.5 s, $\mathbf{w_{MA}}$ = 0.8s, $\boldsymbol{j}$ = 0.4 Hz, $\mathbf{P_T}$ = 0.1 m/s^2.

However, the power spectrum of gait signals differs significantly between the thigh and lower trunk due to distinct biomechanical patterns and can be further influenced by variations in device orientation and placement (e.g., loosely in a pocket vs. tightly attached). Therefore, to ensure the method's effective application to the current dataset, it was necessary to adjust the upper frequency cutoff $\mathbf{f_u}$. To this end, a frequency analysis was conducted using data from HCs only, wherein Fast Fourier Transforms (FFT) were computed for the vertical acceleration signals from the 330 test executions. The data was then visualized (Fig. 3) by overlaying all individual power spectra across all test executions and this visualization was used to determine the upper frequency cutoff $\mathbf{f_u}$, separating gait-related signal components from the ones that are assumed to be noise. Based on this analysis, $\mathbf{f_u}$ was set to 4 Hz for both FrP and BP positions, and FAIR-Q algorithm was subsequently applied using this cutoff to extract the ICs.

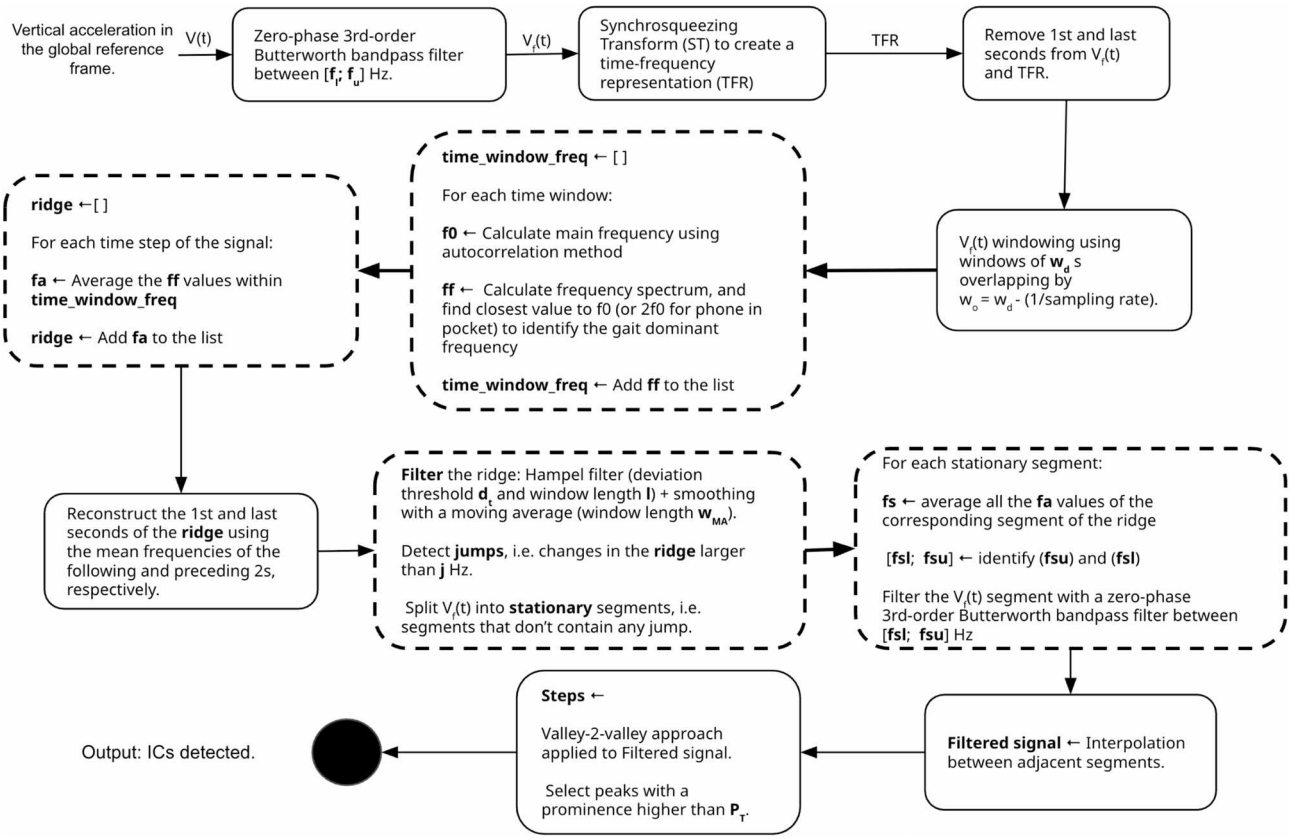

Fig. 1. FAIR-Q algorithmic steps.

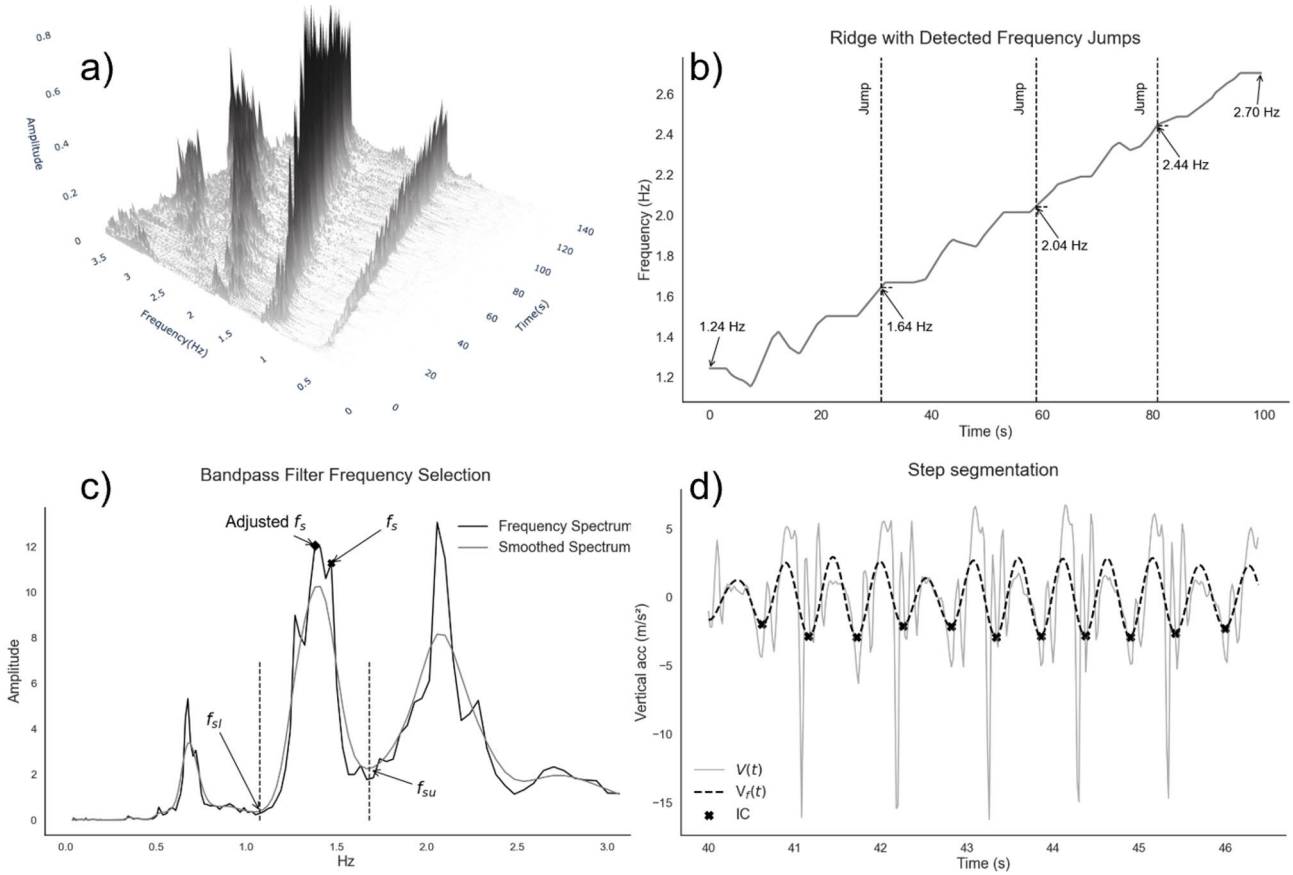

Fig. 2. Representative results from key steps of Frequency-Adaptive, Iterative and Robust Quantification (FAIR-Q): (**a**) Synchrosqueezed Wavelet Transform (SSWT), (**b**) Ridge construction and filtering, (**c**) Stationary segments analysis and (**d**) Detection of Initial Contacts (ICs) and pruning of false positives. Frequency-adaptive, iterative and robust quantification (FAIR-Q) method for step detection

Finally, the previous application of the FAIR-Q algorithm to lower back data[19] revealed a consistent temporal bias of 90 milliseconds in the IC estimates. To account for this potential aspect also in the FrP and BP configurations, a bias analysis was performed on the data collected from the HCs using a Bland-Altman approach[24], wherein IC timings obtained from FAIR-Q were compared against the reference measurements from Xsens. The identified bias was then systematically corrected also in the final ICs estimated in both HCs and pwMS.

Measurements and statistical analysis

The unbiased IC timings calculated by FAIR-Q were used to derive gait temporal parameters, including step duration (**T1**), stride duration (**T2**) and *Cadence*, which was computed according to the definition provided in[13]:

$$Cadence = 2 * \frac{\sum_{k=1}^{N} \frac{60}{T2_k}}{N} \quad (1)$$

where N is the total number of strides and $T2_k$ is the stride duration of the k_{th}–stride in a given test.

Ground truth ICs from the Xsens system (IC_{REF}) were used as a reference to evaluate the performance of FAIR-Q. A predicted IC from FAIR-Q was considered a True Positive (TP) if it occurred within a 0.32s window centered on a corresponding IC_{REF}, which is a narrower window than that used in[13]. Predicted ICs that did not fall within any IC_{REF} windows were labeled as False Positives (FP). Conversely, IC_{REF} events without corresponding predicted ICs were labeled as False Negatives (FN). This information was used to calculate the following metrics:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP + FP} \quad (3)$$

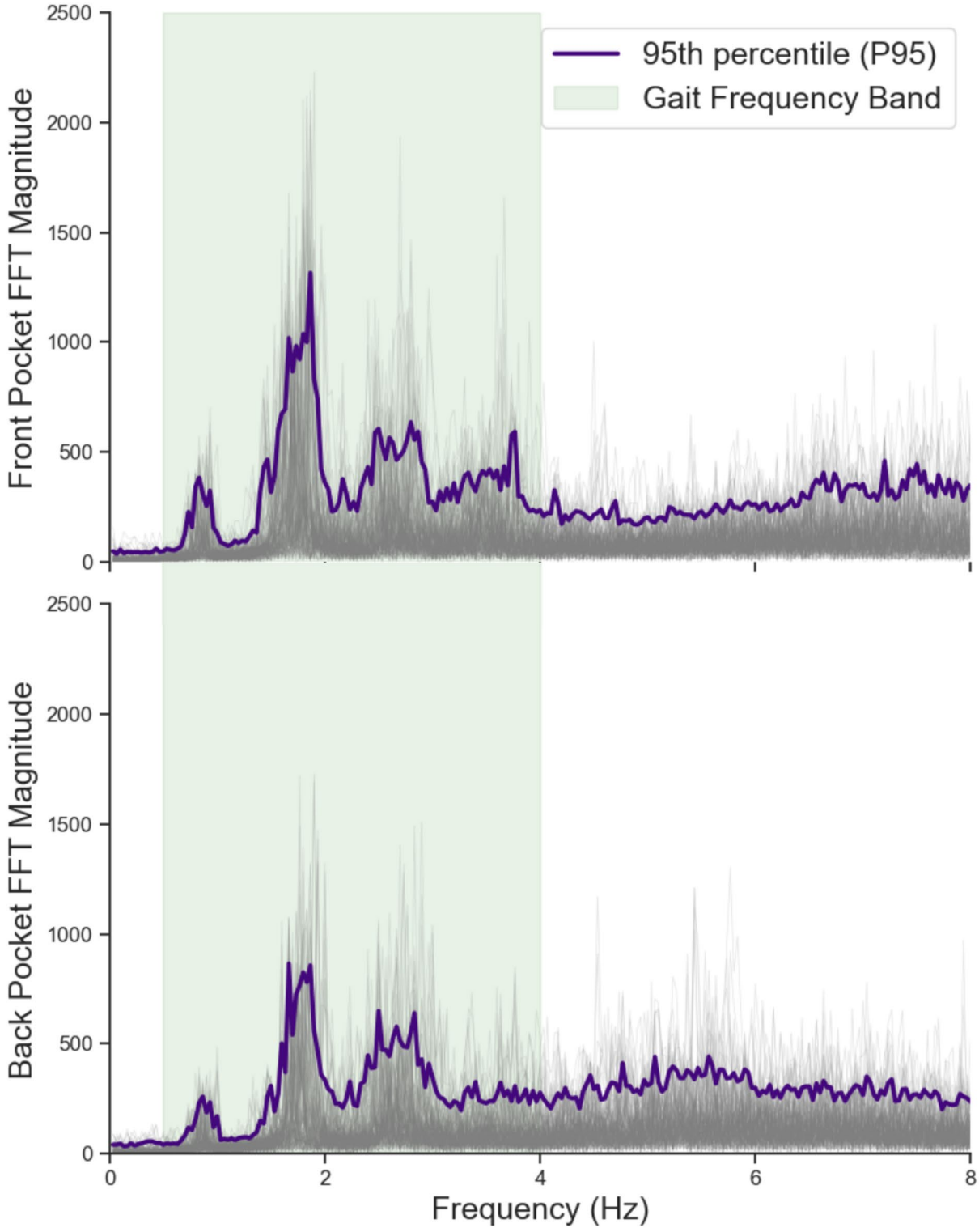

Fig. 3. Aggregated Fast Fourier Transforms (FFTs) of the vertical acceleration and 95th percentile curve in FrP and BP positions, zoomed to the range 0–8 Hz to focus on gait-related frequencies and enhance visualization.

$$F1\ Score\ =\ \frac{TP}{TP\ +\ 0.5(FP+FN)} \tag{4}$$

The accuracy in detecting the exact timing of individual ICs, the duration of individual steps (**T1**) and strides (**T2**) was established using only the TP events in terms of both absolute and relative errors:

$$Median\ Absolute\ error\ TP\ IC\ (s)\ =\ med\,|IC_{REF_{TP}} - IC_{FAIR-Q_{TP}}| \tag{5}$$

$$Median\ Absolute\ error\ in\ TP\ step\ duration\ (s)\ = |med\,T1_{REF_{TP}} - med\,T1_{FAIR-Q_{TP}}| \tag{6}$$

$$Median\ Absolute\ error\ in\ TP\ stride\ duration\ (s)\ = |medT2_{REF_{TP}} - medT2_{FAIR-Q_{TP}}| \tag{7}$$

$$Median\ relative\ error\ in\ TP\ step\ duration\ (\%)\ = \frac{|medT1_{REF_{TP}} - medT1_{FAIR-Q_{TP}}| * 100}{medT_{REF}} \tag{8}$$

$$Median\ relative\ error\ in\ TP\ stride\ duration\ (\%)\ = \frac{|medT2_{REF_{TP}} - medT2_{FAIR-Q_{TP}}| * 100}{medT_{REF_{TP}}} \tag{9}$$

Additionally, the accuracy of the method when applied to the analysis of the whole test was assessed using all steps and strides detected by each method and calculating the median relative error over the whole test:

$$Median\ relative\ error\ in\ step\ duration\ at\ test\ level\ (\%)\ \frac{|medT1_{REF} - medT1_{FAIR-Q}| * 100}{medT_{REF}} \tag{10}$$

$$Median\ relative\ error\ in\ stride\ duration\ at\ test\ level\ (\%)\ \frac{|medT2_{REF} - medT2_{FAIR-Q}| * 100}{medT2_{REF}} \tag{11}$$

Lastly, agreement between paired FAIR-Q and Xsens measurements for the total number of detected ICs, T1, T2, and cadence was assessed using Bland–Altman plots, including all four tests for each HC and the single test for each pwMS. The Intra-Class Correlation Coefficient (ICC(2,1))[25] was computed to provide a quantitative measure of agreement between paired FAIR-Q and reference system measurements. Additionally, to quantitatively assess agreement between Android and iPhone smartphones devices, for each HC and phone position (FrP or BP) the measurements from separate repetitions were compared using Bland–Altman analysis (bias, mean difference, and percentage of bias/mean ratio) to characterize the direction and magnitude of any deviations.

Results

IC timings bias analysis

When comparing the estimates of the TP IC timings obtained for the HCs from FAIR-Q and the reference system using a Bland Altman approach (Fig. 4), a constant bias of 0.03s and 0.07s was observed for the FrP and BP positions, respectively. These values were hence compensated by subtracting these values as a last step in the ICs timing, before computing the other derived metrics. The 0.03 bias in FrP position was also accounted for in the analysis of the data derived from pwMS.

Analysis of performance metrics

Regarding the number of ICs detected per test execution, only 7 out of 165 FrP (4%) and 9 out of 165 BP (5%) in the HC group, as well as 2 out of 49 FrP (4%) in pwMS, fell outside the limits of agreement. Visual inspection of these signals seemed to indicate that they contain portions with high varying cadence. These outliers, however, did not heavily reflect on the median step, stride and cadence level, as shown by the Bland Altman Plots calculated for the metrics of interest in both positions and groups (Fig. 5). Figure 5 also highlights an overall excellent agreement between the FAIR-Q and the reference system values, leading to negligible biases, narrow limits of agreements and very few outliers. At the level of the whole 30-second test, the IC detection sensitivity and positive predictive values were excellent with median values for all groups being 1.00 and 0.98, respectively (Table 2). When looking only at TP events, step duration had very similar accuracy in the three conditions, and a mean null error was observed in all three groups.

The plots in Fig. 5 also show that, when looking at TP values, the null bias in the duration estimates holds for both shorter and longer steps and strides. The limits of agreement at test level were also very narrow for the steps estimates ([−0.02,0.01s] in HC BP, [−0.02,0.02s] in HC FrP and [−0.02,0.02s] in pwMS), with only 3% of outliers in the HC and 4% in pwMS. Overall, very similar results were obtained at stride level, where the mean error was close to zero and the 6 outliers in the pwMS group were very close to the 0.01s lower limits of agreements.

The highly accurate behavior of the algorithm was confirmed by all calculated metrics. Table 2 summarizes the results for the assessment of the FAIR-Q performances both at the level of single IC detection, at the level of single step and stride duration quantification and when using these to evaluate gait metrics over the whole 30-second test (i.e. Test-Level performance metrics). Results showed an overall high agreement with the reference system for all estimates. Almost identical performance was recorded in terms of ability of FAIR-Q to detect the ICs in the two smartphone locations, as well as in HCs versus pwMS.

The algorithm proved very accurate at both step and stride level (<4% and <2% median relative error, respectively) when looking at TP only. These differences between step and stride duration accuracy at single step/stride level were no more noticeable when results were considered at whole-test level, where the errors became negligible for both step and stride duration.

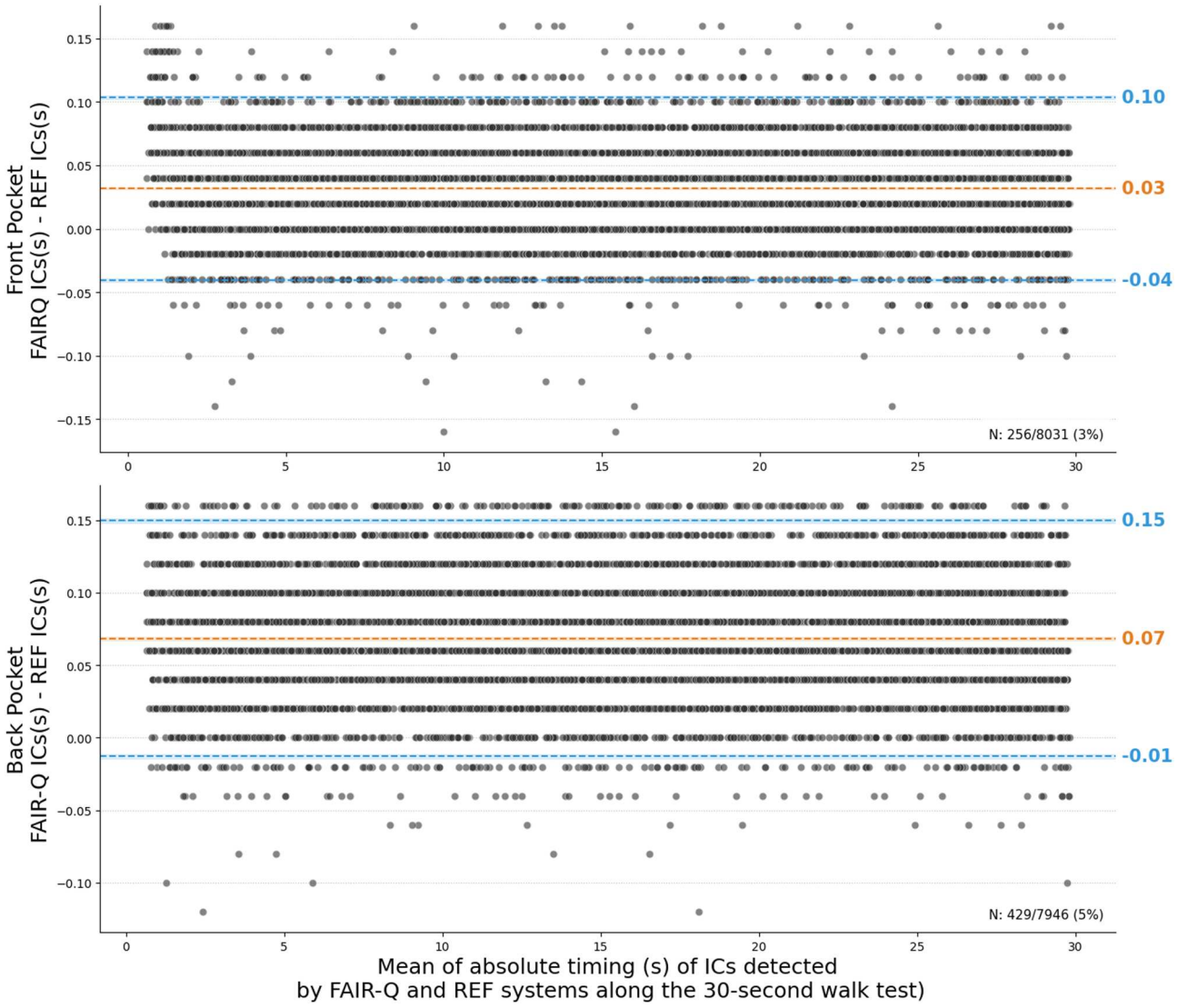

Fig. 4. Bias analysis of the True Positive Initial Contact timings between FAIR-Q and the reference system when applied to data from Front (top panel) and Back (bottom panel) pockets. N, number of ICs outside the limits of agreement; Blue dotted lines, Upper and Lower Limits of Agreement; Orange dotted lines, Mean of values.

Excellent performances of the FAIR-Q algorithm were also confirmed by ICC calculation (Table 3), showing robust agreement for both smartphone pocket positions and in both HCs and pwMS groups for all gait measures, i.e. number of ICs, step/stride duration and cadence, where all average ICCs remained ≥ 0.90.

Discussion

This paper aimed to propose and validate an approach for gait IC detection and step and stride segmentation applicable to data collected from different smartphone locations. To this end, ICs detected by FAIR-Q from smartphones placed in a back and/or front pocket were compared to those collected from a reference system for ground truth determination. FAIR-Q method proved to be highly accurate for both smartphone positions and when applied to gait data from both healthy controls and pwMS. This makes it particularly suitable for the analysis of gait IMU data from studies in which a participant might decide to swap smartphone locations between different observations, which is at risk to occur at a significant rate. Moreover, FAIR-Q method was tested using gait IMU data from a wide variety of smartphone brands and models, ensuring its suitability for trials collecting data in a "bring-your-own-device" setting.

The Bland-Altman plots generated for the analysis of the True Positive (TP) ICs (Fig. 4), built with data from the HCs only, showed the presence of a constant minor bias between FAIR-Q-detected IC and the IC_{REF} timing, which confirmed previous results obtained for a sensor in the belt[19]. When accounting for this bias by subtracting it from all IC estimates obtained from the FrP and BP smartphone positions, the resulting IC estimations showed only a negligible median absolute error of 0.02 s in all tested groups (HCs FrP and BP, and

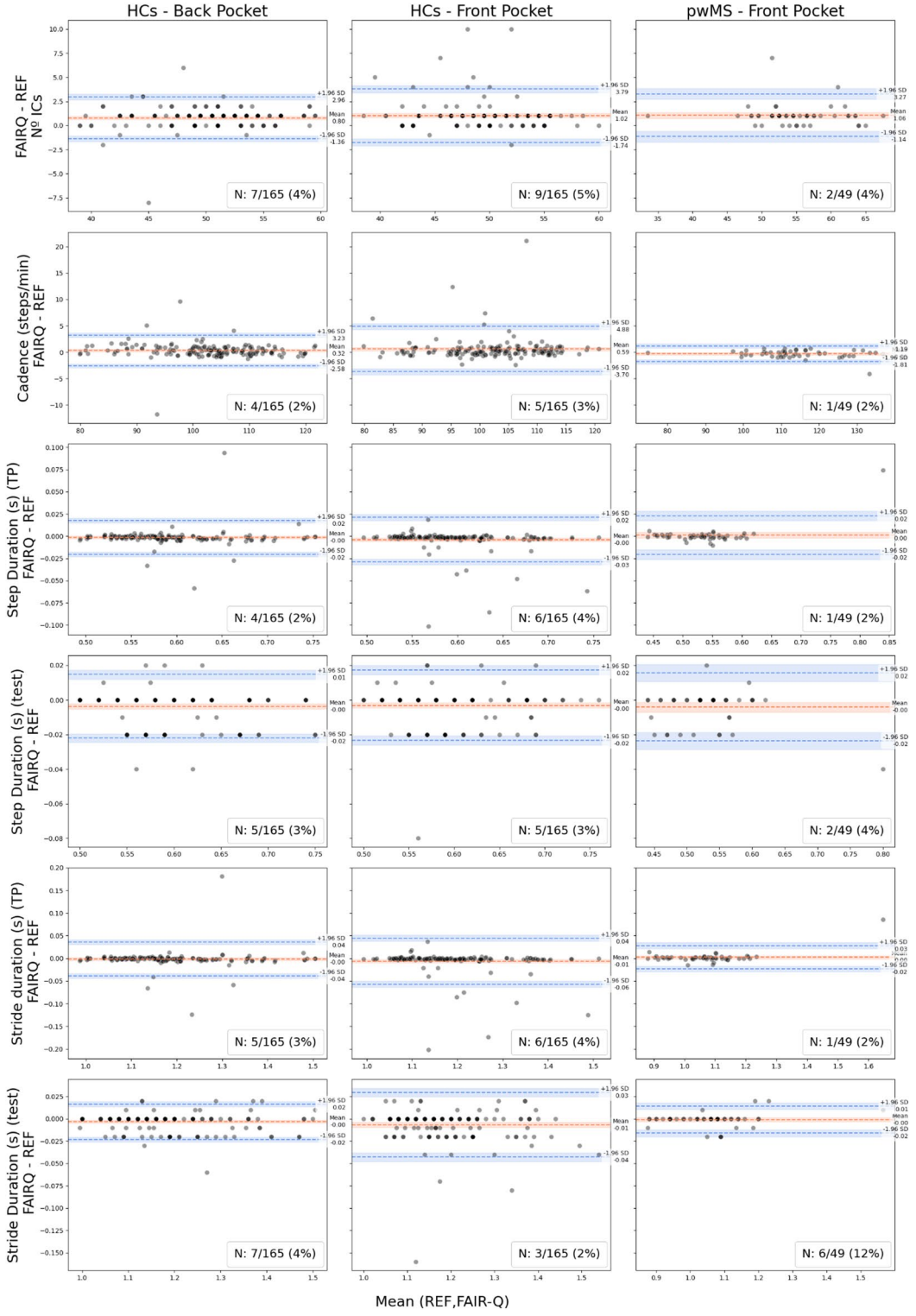

Fig. 5. Assessment of FAIR-Q performance per metric applied to data from healthy controls (HCs), Back (left column) and Front (center column) pockets, and people with multiple sclerosis (pwMS, right column). TP: true positive. N: number of points outside the limits of agreement. Blue dotted lines: Upper and Lower Limits of Agreement. Orange dotted lines: Mean of values. Visual differences in data point density across plots are due to overlapping points, with darker shading indicating higher density.

Gait Measure	Performance Measure	HCs - Back Pocket	HCs - Front Pocket	pwMS - Front Pocket
		Median (Q1-Q3)	Median (Q1-Q3)	Median (Q1-Q3)
ICs	Sensitivity	1 (0.98–1.0)	1 (0.98–1.0)	1 (0.98–1.0)
	Positive predictive value	0.98 (0.98–0.98)	0.98 (0.98–0.98)	0.98 (0.97–0.98)
	F1 score	0.99 (0.98–0.99)	0.99 (0.98–0.99)	0.99 (0.98–0.99)
	Median Absolute error in TP IC [s]	0.02 (0.02–0.04)	0.02 (0.02–0.02)	0.02 (0.02–0.04)
Step	Median absolute error in TP step duration at single step-level [s]	0.02 (0.02–0.02)	0.02 (0.02–0.02)	0.02 (0.02–0.02)
	Median relative error in TP step duration at single step level [%]	3.45 (3.33–3.70)	3.45 (3.23–3.70)	3.85 (3.57–4.35)
	Median relative error in step duration at test level [%]	0 (0.0–0.0)	0 (0.0–0.0)	0 (0.0-1.70)
Stride	Median absolute error in TP stride duration at single stride level [s]	0.02 (0.02–0.02)	0.02 (0.02–0.02)	0.02 (0.02–0.02)
	Median relative error in TP stride duration at single stride level [%]	1.79 (1.64–1.84)	1.7 (1.56–1.79)	1.82 (1.61–1.96)
	Median relative error in stride duration at test level [%]	0 (0.0–0.85)	0 (0.0–1.54)	0 (0.0–0.0)
Cadence	Median relative error in Cadence at test level [%]	0.01 (0.00–0.01)	0.01 (0.00–0.01)	0.00 (0.00–0.01)

Table 2. Performance metrics for initial contacts (ICs), Step, Stride, and Cadence measures in healthy controls (HCs) and people with multiple sclerosis (pwMS).

	FAIR-Q vs. Xsens			FAIR-Q (Android) vs. FAIR-Q (iPhone)			
	ICC [95% confidence interval]			ICC [95% confidence interval]		Bias Android - iOS /Mean (%)	
Gait measure	HCs - BP $N=165$	HCs - FrP $N=165$	pwMS - FrP $N=49$	HCs - BP $N=82$	HCs - FrP $N=82$	HCs - BP $N=82$	HCs - FrP $N=82$
Number of ICs	0.96 [0.85–0.98]	0.93 [0.78–0.97]	0.97 [0.8–0.99]	0.96 [0.93–0.97]	0.90 [0.85–0.93]	0.2/50.66 (0.39%)	−0.37/50.18 (-0.73%)
Step duration (TP at single step level)	0.98 [0.98–0.99]	0.98 [0.97–0.99]	0.98 [0.97–0.99]	0.97 [0.96–0.98]	0.91 [0.86–0.94]	−0.002/0.59 (−0.28%)	0.005/0.59 (0.85%)
Step duration (test level)	0.98 [0.97–0.99]	0.98 [0.97–0.99]	0.98 [0.97–0.99]	0.99 [0.98–0.99]	0.91 [0.86–0.94]	−0.002/0.586 (−0.27%)	0.005/0.594 (0.76%)
Stride duration (TP at single stride level)	0.99 [0.98–0.99]	0.98 [0.97–0.99]	0.99 [0.99–1]	0.98 [0.96–0.98]	0.91 [0.86–0.94]	−0.004/1.18 (−0.31%)	0.01/1.188 (0.86%)
Stride duration (test level)	1 [0.99–1]	0.99 [0.98–0.99]	1 [1–1]	0.99 [0.98–0.99]	0.92 [0.87–0.95]	−0.002/1.175 (−0.16%)	0.009/1.182 (0.76%)
Cadence	0.99 [0.98–0.99]	0.97 [0.96–0.98]	1 [1–1]	0.98 [0.96–0.98]	0.91 [0.86–0.94]	0.21/102.99 (0.21%)	−0.71/102.25 (−0.69%)

Table 3. Intraclass correlation coefficients (ICC) with 95% confidence intervals for the agreement of gait measures in HC and MS groups. The left column ("FAIR-Q vs Xsens") reports the agreement between FAIR-Q and the reference system. The right column ("FAIR-Q (Android) vs FAIR-Q (iPhone)") shows the agreement between FAIR-Q measurements obtained from different smartphones worn by HCs at the same position (FrP or BP), but recorded in separate repetitions. Also, the right column shows the bias, mean, and percentage of the bias-to-mean ratio derived from a Bland–Altman analysis. N: the number of measurement pairs used to calculate the ICC.

pwMS FrP; see Table 2). This accuracy is in line with what has been reported in our previous application of FAIR-Q to real-world lower back IMU data[19] and also with results reported by other studies for both healthy adults and pwMS in short gait laboratory assessments[13] and in continuous monitoring scenarios[13,26,27].

The next part of the analysis assessed the quality of the features derived from the ICs, namely step and stride durations (median value at the whole 30-second walking test level) and cadence. IC detection estimates showed excellent sensitivity, positive predictive value, and F1 score, with median values for all groups being 1.00, 0.98, and 0.99, respectively. This clearly shows that FAIR-Q correctly detects the vast majority of the ICs consistently with the reference system. The method clearly outperformed a previous study that evaluated IMU sensors placed in the front pocket, which achieved an F1-Score of 0.95 within-subjects, but only 0.76 between subjects[28].

The results concerning step and stride duration quantification showed very low absolute and relative errors. Absolute and relative errors were the same in all three groups, with a median absolute error for step and stride duration of 0.02s, and relative errors below 4% for steps and 2% for strides. These results are in line with literature including MS participants both when the device is placed on the lower back[13] or laterally on a non-dominant hip[29].

The reported ICC values (Table 3), when comparing FAIRQ vs. reference system, were on average in the range of 0.93–0.99. ICC values were similar between FrP and BP, but the confidence intervals were wider for FrP. This is not surprising given that IC events signals collected in a FrP position might be slightly less marked, and hence more difficult to detect, when ground contacts of the foot correspond to the leg opposite to where the

smartphone is placed. Nonetheless, ICC values were considerably better with FAIR-Q (0.98–0.99) compared to those between 0.53 and 0.69 previously reported for step duration estimates from a smartphone in a pocket[16].

FAIR-Q also demonstrated high consistency across operating systems and smartphone manufacturers (Table 3) when comparing two smartphones, with ICC values ranging from 0.90 to 0.99 for all gait metrics in both FrP and BP positions. The bias between Android and iOS devices was negligible (≤ 0.01s for step and stride duration), indicating that the algorithm's performance was robust to variations in smartphone hardware and operating system, supporting its suitability for large-scale trials.

Cadence, assessed for by both relative error and level of agreement, demonstrated that the underlying hypothesis of FAIR-Q is correct and justifies the use of synchrosqueezing wavelet transform method for signal filtering and, consequently, for IC detection. The median relative error was ≤ 0.01%, and the ICC values ranged between 0.96 and 1.00 across the entire study population. Again, this proves that the algorithm clearly outperformed previous studies that estimated cadence in similar scenarios reporting errors of 5.5% for the FrP and of 10.5% for the BP[30].

It is noteworthy that the agreement between FAIR-Q and the Xsens reference system in the pwMS group was very similar to that observed in HCs, with even narrower limits of agreement in the Bland-Altman plots. This might be due to the relatively mild disease severity in our pwMS sample and the smaller sample size. Nonetheless, it is worth highlighting that the investigated group comprised individuals with EDSS values between 0 and 6, which provided some heterogeneity in the tested pwMS population. These positive findings are encouraging, especially when considering the challenges previously identified in the literature regarding monitoring with smartphones in the pocket. A recent systematic review on the topic, in fact, highlighted that inconsistent smartphone placement can affect the accuracy of gait measurements in patients with MS[31], to the extent that smartphone data collection could become insufficiently precise for use in clinical research when the device is in the front pocket[32]. The development and validation of FAIR-Q addresses this scientific barrier and ensures a first-in-kind robustness to smartphone placement variability, which is a necessary condition for deploying smartphone-based digital outcome measures in large-scale clinical trials or for routine clinical care[33,34].

Overall, the proposed method showed strong potential for robust gait event detection. While not undermining the present findings, some limitations to their generalisability persist, which provide directions for future work. First, the patient cohort consisted mainly of individuals in earlier disease stages, and assessing performance in more advanced cases will help assess the method's resilience to more atypical gait patterns. Second, although FAIR-Q has been previously validated using real-world data from a lower-back sensor, additional datasets would be valuable to demonstrate equivalent performance with pocket-worn devices where relevant. Last but not least, extending the approach to the detection of final contact events could unlock additional clinically meaningful gait features that go beyond those derived from initial contacts alone.

Conclusions

This study proved the validity of the FAIR-Q method for detecting steps and strides from smartphone data obtained from different device placements in both healthy controls and people with MS, underlining its robustness and clinical applicability. This flexibility makes the method particularly suitable for gait monitoring in longitudinal studies performed in a non-clinical environment, where inconsistency in device location can hamper measurement accuracy. Furthermore, FAIR-Q's results were obtained from a variety of smartphone brands and models, confirming its usability also in a bring-your-own-device scenario, which ensures ecological validity and scalability of the proposed approach. In conclusion, considering also its previous validation on real-world data from different neurological populations, the proposed FAIR-Q method is an ideal solution for unsupervised smartphone-based high-frequency assessment of mobility in neurological diseases.

Data availability

The datasets generated and analysed during the current study are not publicly available due to patient privacy and ethical restrictions but could be available from the corresponding author on reasonable request.

Received: 18 September 2025; Accepted: 26 December 2025
Published online: 07 January 2026

References

1. Taylor, K. I. et al. Exploratory digital outcome measures of motor sign progression in parkinson's disease patients treated with prasinezumab. *Npj Digit. Med.* **8**, 365. https://doi.org/10.1038/s41746-025-01572-8 (2025).
2. Iodice, R. et al. A review of current rehabilitation practices and their benefits in patients with multiple sclerosis. *Mult Scler. Relat. Disord.* **69**, 104460. https://doi.org/10.1016/j.msard.2022.104460 (2023).
3. Scaramozza, M. et al. Sensor-Derived measures of motor and cognitive functions in people with multiple sclerosis using unsupervised Smartphone-Based assessments: Proof-of-Concept study. *JMIR Form. Res.* **8**, e60673. https://doi.org/10.2196/60673 (2024).
4. Karatsidis, A. et al. Characterizing gait in people with multiple sclerosis using digital data from smartphone sensors: A proposed framework. *Multiple Scler. J.* **31** (5), 512–528. https://doi.org/10.1177/13524585251316242 (2025).
5. Woelfle, T. et al. Practice effects of mobile tests of Cognition, Dexterity, and mobility on patients with multiple sclerosis: Data analysis of a Smartphone-Based observational study. *J. Med. Internet. Res.* **23** (11), e30394. https://doi.org/10.2196/30394 (2021).
6. Arteaga-Bracho, E. et al. Smartphone-Based assessment of mobility and manual dexterity in adult people with spinal muscular atrophy. *J. Neuromuscul. Dis.* **11** (5), 1049–1065. https://doi.org/10.3233/JND-240004 (2024).
7. Nguyen, L. V. & La, H. M. Real-Time human foot motion localization algorithm with dynamic speed. *IEEE Trans. Human-Machine Syst.* **46** (6), 822–833. https://doi.org/10.1109/THMS.2016.2586741 (2016).
8. Negi, S. et al. FSR and IMU sensors-based human gait phase detection and its correlation with EMG signal for different terrain walk. *Sens. Rev.* **41** (3), 235–245. https://doi.org/10.1108/SR-10-2020-0249 (2021).

9. Oudre, L. et al. Template-Based step detection with inertial measurement units. *Sensors* **18** (11), 4033. https://doi.org/10.3390/s18114033 (2018).
10. McCamley, J. et al. An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. *Gait Posture*. **36** (2), 316–318. https://doi.org/10.1016/j.gaitpost.2012.02.019 (2012).
11. Soltani, A. et al. Algorithms for walking speed Estimation using a Lower-Back-Worn inertial sensor: A Cross-Validation on speed ranges. *IEEE Trans. Neural Syst. Rehabil. Eng.* **29**, 1955–1964. https://doi.org/10.1109/TNSRE.2021.3111681 (2021).
12. Romijnders, R. et al. A deep learning approach for gait event detection from a single Shank-Worn IMU: validation in healthy and neurological cohorts. *Sensors* **22** (10), 3859. https://doi.org/10.3390/s22103859 (2022).
13. Micó-Amigo, M. E. et al. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J. Neuroeng. Rehabil.* **20** (1), 78. https://doi.org/10.1186/s12984-023-01198-5 (2023).
14. Di Bacco, V. E. & Gage, W. H. Evaluation of a smartphone accelerometer system for measuring nonlinear dynamics during treadmill walking: Concurrent validity and test-retest reliability. *J. Biomech.* **151**, 111527. https://doi.org/10.1016/j.jbiomech.2023.111527 (2023).
15. Shahar, R. T. & Agmon, M. Gait analysis using accelerometry data from a single smartphone: agreement and consistency between a smartphone application and Gold-Standard gait analysis system. *Sensors* **21** (22), 7497. https://doi.org/10.3390/s21227497 (2021).
16. Silsupadol, P. et al. Reliability and validity of a smartphone-based assessment of gait parameters across walking speed and smartphone locations: Body, bag, belt, hand, and pocket. *Gait Posture*. **58**, 516–522. https://doi.org/10.1016/j.gaitpost.2017.09.030 (2017).
17. Shirley, S. et al. Non-Invasive Biomechanical intervention leads to low rates of total knee replacement and reduced utilization of healthcare resources among older adults with chronic knee pain: A 5-Year Follow-Up study. *J. Musculoskelet. Disorders Treat.* https://doi.org/10.23937/2572-3243.1510121 (2023).
18. Godfrey, A. et al. Beyond the front end: investigating a thigh worn accelerometer device for step count and bout detection in parkinson's disease. *Med. Eng. Phys.* **38** (12), 1524–1529. https://doi.org/10.1016/j.medengphy.2016.09.023 (2016).
19. Ena, A. et al. FAIR-Q: A Frequency-Adaptive Iterative and Robust Quantification Method for Step Detection. *IEEE International Conference on Digital Health (ICDH)*, Helsinki, Finland, pp. 196–198 (2025). https://doi.org/10.1109/ICDH67620.2025.00035
20. D'Souza, M. et al. Neurostatus e-Scoring improves consistency of expanded disability status scale assessments: A proof of concept study. *Mult Scler. Houndmills Basingstoke Engl.* **23**, 597–603. https://doi.org/10.1177/1352458516657439 (2017).
21. Konrath, J. et al. Xsens MVN gait report: The use of inertial motion capture for cloud based reporting of gait parameters [White paper]. *Xsens Technol. B V* (2021).
22. Ena, A. et al. A novel methodology for developing Smartphone-instrumented tests for assessing Movement, Dexterity, and balance in neurological patients: Technical verification of ground truth Datasets, and analytical and clinical validation of digital biomarkers. *Neurology* **102** (22_supplement_1). https://doi.org/10.1212/WNL.0000000000205654 (2024).
23. Daubechies, I. et al. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* **30** (2), 243–261. https://doi.org/10.1016/j.acha.2010.08.002 (2011).
24. Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8** (2), 135–160. https://doi.org/10.1177/096228029900800204 (1999).
25. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86** (2), 420–428.https://doi.org/10.1037/0033-2909.86.2.420 (1979).
26. González, L. et al. Estimation of ground contact time with inertial sensors from the upper arm and the upper back. *Sensors* **23** (5), 2523. https://doi.org/10.3390/s23052523 (2023).
27. Storm, F. A. et al. Free-living and laboratory gait characteristics in patients with multiple sclerosis. *PloS one*. **13** (5), e0196463 (2018). https://doi.org/10.1371/journal.pone.0196463
28. Garcia, F. A. et al. Adaptive algorithm for gait segmentation using a single IMU in the thigh pocket. *IEEE Sens. J.* **22** (13), 13251–13261. https://doi.org/10.1109/JSEN.2022.3177951 (2022).
29. Moon, Y. et al. Monitoring gait in multiple sclerosis with novel wearable motion sensors. PLOS ONE 12(2), e0171346 (2017). https://doi.org/10.1371/journal.pone.0171346 (2017).
30. Karuei, I. et al. Robust realtime algorithm for Cadence Estimation. *Pervasive Mob. Comput.* **13**, 52–66. https://doi.org/10.1016/j.pmcj.2013.09.006 (2014).
31. Abou, L. et al. Smartphone applications to assess gait and postural control in people with multiple sclerosis: A systematic review. *Multiple Scler. Relat. Disorders* https://doi.org/10.1016/j.msard.2021.102943 (2021).
32. Balto, J. M. et al. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. *Multiple Scler. J. - Exp. Transac. Clin.* https://doi.org/10.1177/2055217316634754 (2016).
33. Demanuele, C. et al. Considerations for conducting bring your own device (BYOD) clinical studies. *Digit. Biomarkers*. **6** (2), 47–60. https://doi.org/10.1159/000525080 (2022).
34. Gwaltney, C. et al. Bring your own device (BYOD): the future of Field-Based Patient-Reported outcome data collection in clinical trials? *Therapeutic Innov. Regul. Sci.* **49** (6), 783–791. https://doi.org/10.1177/2168479015609104 (2015).

Author contributions

A. Ena: Software, Conceptualization, Methodology, Statistical analysis & Visualization, Writing, review & editingC. Mazzà: Conceptualization, Methodology, Writing, review & editingA. Rodríguez-Romero: Data Collection, Data curation, Statistical analysis & Visualization, Writing, review & editingT. Woelfle: Data collection, Writing, review & editingJ. Mueller: Writing, review & editingC. Granziera: Writing, review & editingL. Kappos: Writing, review & editingE. Bartholomé: Writing, review & editingC. Bernasconi: Writing, review & editingS. Belachew: Methodology, Writing, review & editingÓ. Reyes: Software, Conceptualization, Methodology, Writing, review & editing.

Funding

This research received no external funding.

Declarations

Competing interests

A. Ena, C. Mazzà, A. Rodríguez-Romero, E. Bartholomé, C. Bernasconi, S. Belachew and Ó. Reyes are employees of Indivi AG, which partially supported this research. T. Woelfle declares no competing interests. J. Mueller received funding from the Swiss National Science Foundation (grant no. P500PM214230 and P5R5PM225288) and the „Pool für Innere Medizin" of the University Hospital Basel, outside the presented work. The University Hospital Basel (USB) and the Research Center for Clinical Neuroimmunology and

Neuroscience (RC2NB), as the employers of C. Granziera and J. Mueller, have received the following fees which were used exclusively for research support: (i) advisory board and consultancy fees from Actelion, Genzyme-Sanofi, Novartis, Merck, GeNeuro and Hoffmann La Roche; (ii) speaker fees from Genzyme-Sanofi, Novartis, Merck, Jannsen, GeNeuro and Hoffmann La Roche; (iii) research support from Siemens Healthineers, GeNeuro, Hoffmann La Roche, Sanofi-Genzyme and Novartis. L. Kappos's institution (University Hospital Basel) has received the following exclusively for research support: Steering committee, advisory board and consultancy fees (Bayer, Biogen, Biohaven, Bristol Myers Squibb, Celltrion Inc, Clene Nanomedicine Inc., Eli Lilly, EMD Serono, Genentech, GlaxoSmithKline, Galapagos NV, Janssen, Japan Tobacco, Kiniksa Pharmaceuticals, Merck Healthcare AG, Minoryx, Santhera, Neurostatus UHB AG, Novartis, Roche, Sanofi, Shionogi, Wellmera, Zai Lab); speaker fees (Bristol Myers Squibb, Janssen, Novartis, Sanofi and Roche); support for educational activities (Merck, Novartis, Roche, Sanofi); testimony for Df-mp Molina & Pohlman; and grants (European Union, Innosuisse, Merck Healthcare AG, Novartis, Neurostatus UHB AG, Sanofi, Roche). External funders did not have involvement in the study design, data collection, data analysis, interpretation, or writing of the report.

Ethical approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki. Ethics approval for the healthy controls (HCs) was obtained from the University of Sheffield, UK (Approval 064650), and for people with multiple sclerosis (pwMS) from the Ethikkommission Nordwest- und Zentralschweiz, Basel, Switzerland (BASEC ID 2021 D0040, NCT05009160). All participants provided their informed consent prior to inclusion in the study.

Additional information

Correspondence and requests for materials should be addressed to Ó.R.