Quantification of Motor Impairment in Parkinson's Disease Using an Instrumented Timed Up and Go Test

Luca Palmerini, Sabato Mellone, Guido Avanzolini, Franco Valzania, and Lorenzo Chiari

***Abstract*—The Timed Up and Go (TUG) test is a clinical test to assess mobility in Parkinson's disease (PD). It consists of rising from a chair, walking, turning, and sitting. Its total duration is the traditional clinical outcome. In this study an instrumented TUG (iTUG) was used to supplement the quantitative information about the TUG performance of PD subjects: a single accelerometer, worn at the lower back, was used to record the acceleration signals during the test and acceleration-derived measures were extracted from the recorded signals. The aim was to select reliable measures to identify and quantify the differences between the motor patterns of healthy and PD subjects; in order to do so, besides comparing each measure individually to find significant group differences, feature selection and classification were used to identify the distinctive motor pattern of PD subjects. A subset of three features (two from Turning, one from the Sit-to-Walk component), combined with an easily-interpretable classifier (Linear Discriminant Analysis), was found to have the best accuracy in discriminating between healthy and early-mild PD subjects. These results suggest that the proposed iTUG can characterize PD motor impairment and, hence, may be used for evaluation, and, prospectively, follow-up, and monitoring of disease progression.**

***Index Terms*—Accelerometer, classification, feature selection, Parkinson's disease, Timed Up and Go (TUG).**

I. Introduction

THE Timed Up and Go (TUG) test is a widely used clinical test [1] to assess balance, mobility and fall risk in the elderly [2]–[4] and in patients with Parkinson's disease (PD) [5], [6]. It consists of rising from a chair, walking 3 m at preferred speed, turning around, returning and sitting. It is simple and easy to perform in the clinic. The traditional clinical outcome of this test is its total duration [1], which is usually measured by a stopwatch. This single measure evaluates the locomotor performance as a whole, providing no information regarding the performance in specific components of the test. This is why, recently, instrumented versions of the test [instrumented Timed Up and Go (iTUG)] have been proposed [7]–[12], in which inertial sensors were used as the measurement system. These studies showed the potential of using inertial sensors to boost the quantitative information about the TUG performance for fall risk estimation in the elderly [10], [11], assessment of cognitive decline [12], and motor function evaluation of PD subjects [7]–[9].

In this latter group, in particular, the total test duration measured by a stopwatch was not able to differentiate either between moderate PD and age-matched healthy subjects [8] or between early-mild PD and age-matched healthy subjects [7]. It is clear, then, that the traditional clinical outcome is not appropriate for assessing the distinctive locomotor signs of PD, especially in an early stage when the locomotor function is not highly impaired yet.

Although the above-mentioned studies supplement the information available about TUG performance in PD, they share a limitation: they focus on comparing individual measures between healthy and PD subjects. Instead, the present study focuses on using combinations of different measures to accurately discriminate between the motor patterns of the two populations by using a classifier. To test the accuracy of the classification rule, we planned an iTUG evaluation on a group of early-mild PD subjects and age-matched healthy controls.

In a previous study [13], classification with subsets of features from an accelerometer-based posture analysis led to highly accurate discrimination between early-mild PD and healthy subjects. Several feature subsets achieved misclassification rates as low as 5%. Most of them included a tremor-related measure, a postural measure in the frequency domain, and a postural displacement measure. This result suggested that quantitative posture analysis with a single accelerometer can provide information for characterizing early postural decline in PD subjects. In this study the same group of subjects were also assessed on their locomotor performance. We hypothesized that the locomotor evaluation using iTUG, followed by a feature selection process, would be able to discriminate between early-mild PD subjects, whose locomotor function is only slightly impaired, and healthy age-matched control subjects.

The specific aims of the present study were: 1) to analyze the reliability of iTUG measures extracted from acceleration signals; 2) to identify the acceleration-derived measures which best characterize early locomotor impairment in PD; 3) to test correlations between acceleration-derived measures and clinical data; and 4) to select a reduced subset of informative and reliable features which can be used to classify the motor patterns of the two populations.

II. Methods

A. Subjects and Experimental Setup

1) *Subjects*: Twenty early-mild PD subjects (62 ± 7 years old, eight females) and twenty healthy age-matched control

Manuscript received May 31, 2012; revised November 01, 2012; accepted December 03, 2012. Date of publication January 01, 2013; date of current version July 02, 2013. The research leading to these results was supported by the European Union— Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n°288516 (CuPiD Project).

L. Palmerini, S. Mellone, G. Avanzolini, and L. Chiari are with the Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi", University of Bologna, Bologna 40136, Italy (e-mail: luca.palmerini@unibo.it; sabato.mellone@unibo.it; guido.avanzolini@unibo.it; lorenzo.chiari@unibo.it).

F. Valzania is with the Department of Neuroscience, University of Modena, and Reggio Emilia, Modena 41100, Italy (e-mail: f.valzania@ausl.mo.it).

Digital Object Identifier 10.1109/TNSRE.2012.2236577

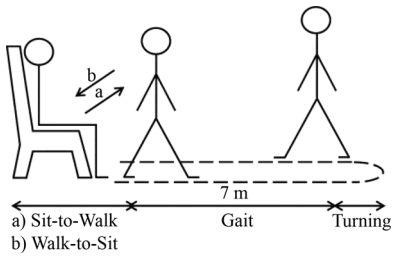

Fig. 1. Components of iTUG.

(CTRL) subjects (64 ± 6 years old, 14 females[1]) were examined in this study. Disease severity of the PD group was assessed by qualified medical staff at the Department of Neuroscience, University of Modena and Reggio Emilia with the Unified Parkinson's Disease Rating Scale (UPDRS). The experimental protocol was approved by the local ethics committee; all subjects gave their written informed consent. The PD subjects were tested OFF medication (both for the UPDRS evaluation and for the experimental session, which were done the same day) obtained by a levodopa washout of at least 18 h and a dopamine-agonists washout of at least 36 h. Eighteen PD subjects were in Hoehn & Yahr stage 2.5, one in stage 2, and one in stage 1.5; the motor-UPDRS score (Section III of UPDRS) was 26.6 ± 7.1 (range: 13-41); the mean disease duration was 5.2 ± 4.1 years (range: 9 months –14.2 years).

2) *iTUG*: As shown in Fig. 1, subjects were instructed to stand up from a chair without armrests, walk at their preferred speed on level ground, for 7 m (instead of the 3 m of the traditional TUG) from the chair to the turning point, walk back and sit; turns were performed around a traffic cone. The measurement session for each subject consisted of three consecutive trials (repetitions). Between consecutive trials subjects took a rest between 30 s and 2 min long, depending on whether and how long they needed to recover.
3) *Measurement System*: The test was instrumented by a tri-axial accelerometer (McRoberts Dynaport Micromod), with a sample rate of 100 Hz and a range of ±2 g. The same experimenter mounted the system at the beginning of each measurement session; the system was not removed until the end of the three trials. The accelerometer was worn on the lower back, by means of an elastic waist belt, at the level of the fifth lumbar vertebra. Acceleration signals from the antero-posterior (AP), medio-lateral (ML), and vertical (V) directions were recorded.

B. Clinical Data

The following clinical data were collected from the PD subjects: age, disease duration, UPDRS scores. The following UPDRS subscores were computed by summing the scores of selected items of the UPDRS: 1) postural instability and gait difficulties (PIGD) subscore [14]; 2) gait and posture subscore [7]; 3) rigidity subscore [7]; 4) bradykinesia subscore [7]; and 5) motor-UPDRS subscore.

Only age was collected from the CTRL subjects.

[1]In the previous study [13], 13 females were incorrectly reported.

C. Processing and iTUG Measures

The iTUG was segmented into four components: Sit-to-Walk (STW), Gait, Turning, and Walk-to-Sit (WTS), as shown in Fig. 1. The AP acceleration signal was used to identify postural transitions [8] and heel strikes [15] (see Fig. 2). Turning was marked manually during the trial by means of a remote control. Unlike the study by Weiss *et al.* [8], the time interval following the acceleration peak was included as a part of the STW and the end of the STW was defined as the moment when the first identified heel strike occurs (Fig. 2). The Gait component was defined as the period of straight-trajectory walking in a steady state. The Gait component starts after the end of the STW (the first identified heel strike is not part of the Gait component), does not include Turning and does not include the last two identified heel strikes. These last two steps were considered part of the following WTS component, assuming they are the ones that the subject does to prepare for sitting. This latter assumption looked reasonable after direct observation of movement patterns in both PD and CTRL subjects. The complete set of measures that were considered for each iTUG component is reported and described in Table I.

Duration was considered for each component; summary measures of the signal (root mean square, normalized jerk score) were computed for each transition (STW, Turning, and WTS). Acceleration was filtered at 20 Hz before the computation of the root mean square (RMS) of the acceleration.

Normalized jerk score (NJS) was defined by analogy with Caby *et al.* [16]: the acceleration was bandpass filtered between 0.15 and 5 Hz (zero lag fourth-order Butterworth filter) to limit the effect of very slow or abrupt variations on the derivative of the acceleration. NJS was normalized with respect to the total movement time.

With regard to Gait measures, step time (Tstep) was computed by identifying heel strikes as described in Zijlstra *et al.* [15]. Tstep was also computed for the Turning component. Cadence (number of steps/min) was not considered as a distinct feature since it can be obtained by dividing 60 s by Tstep. The measures related to the Gait phase [mean phase, standard deviation -STD- of the phase, coefficient of variation -CV- of the phase, and phase coordination index (PCI)] were computed according to Plotnik *et al.* [17]. Harmonic ratio (HR) was computed as described in Menz *et al.* [18]: acceleration was lowpass filtered at 20 Hz (zero lag fourth-order Butterworth filter). NJS was also computed for the Gait component (see Table I).

D. Feature Selection Procedure

For each subject the mean values of each measure across the three consecutive trials were considered in the subsequent feature selection procedure.

One of the aims of the present study is to select a reduced subset of informative and reliable features which can discriminate between the two populations by using a classifier. To this aim, we evaluated classifiers which are easy to interpret and visualize (at least for low dimensions) and that could be

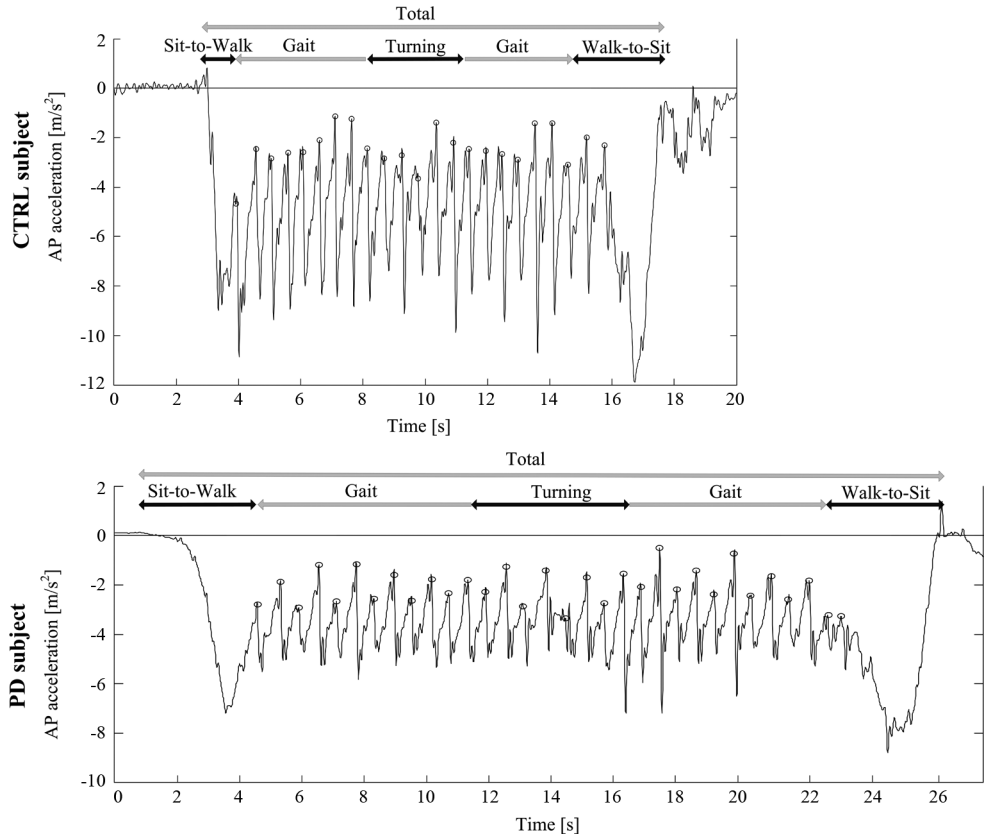

Fig. 2. AP acceleration signal of a CTRL subject and of a PD subject recorded during the iTUG. Signal was segmented into the different components: Identified heel strikes are marked by circles.

easily integrated into the feature selection procedure that we designed. Chosen classifiers included: the linear and quadratic discriminant analysis (LDA and QDA, respectively) and the Mahalanobis classifier (MC). The feature selection procedure we designed (see Fig. 3) has two steps: first, features which were found to be unreliable (according to the criteria defined in Section II-E) were discarded; second, an algorithm was designed to find the best subset among the reliable features. The second step of such a procedure is a modified and an improved version (in terms of generalizability of the accuracy estimates) of the one we designed in a previous paper [13]. It is characterized by a nested cross validation (see Fig. 3): the external Leave-One-Out cross-validation (LOO CV) is used to evaluate the final discrimination ability of the classifier by misclassification rate (MR); it was chosen because of the small sample size. This external LOO CV divides the database into 40 (the total number of subjects) different combinations of training and testing sets ($\mathrm{TR_i}$, $\mathrm{TS_i}$ with $1 \leq \mathrm{i} \leq 40$ in Fig. 3). An internal 10-fold CV was performed on each resulting $\mathrm{TR_i}$ in order to find the best subset related to that specific training set. In each internal 10-fold CV an exhaustive search was performed (all subsets of all cardinalities) to find the subset that optimizes the objective function. Following the *wrapper* approach [19], the penalized MR of the considered classifier in $\mathrm{TR_i}$, obtained by the internal 10-fold CV, was the objective function to be optimized; the 10-fold CV was repeated ten times to limit the variance of the results, as in [19]. The penalty which was added to the obtained MR penalizes feature subsets with many features (as in [19]). The penalty was set to 0% for subsets of a single feature and 1% for each extra feature. The cross-validated penalized MR (mean and standard deviation across the ten repetitions) was computed for all the subsets of all cardinalities to find the subset with the minimum value (see Fig. 4, where it corresponds to a subset of four features). The best subset for $\mathrm{TR_i}$ was then selected as the smallest one which had an objective function lying within one standard deviation from the minimum (in Fig. 4 it corresponds to a subset of three features). The complexity penalty and the selection within one standard deviation were both performed in order to select the smaller of two feature subsets that showed a comparable estimated misclassification rate. This was done in order to avoid overfitting, because in high-dimensionality settings (few samples, many features) smaller subsets of features are preferred [19].

Finally, the accuracy of the classifier (MR) built with the final subset was evaluated by the external LOO CV: for each combination of $\mathrm{TR_i}$ and $\mathrm{TS_i}$, the classifier is trained in $\mathrm{TR_i}$ with the best subset specific to $\mathrm{TR_i}$ and then it is tested in $\mathrm{TS_i}$. It is possible that different subsets (possibly of different cardinalities) are selected in different training sets; therefore, the final selected cardinality is the cardinality most frequently chosen among the 40 training sets; the final selected subset is the one chosen most frequently among the best subsets of the selected cardinality. The 95% confidence intervals (CI) of the final MR estimate, which are reported in Table III, were computed according to Witten *et al.* [20].

TABLE I
INSTRUMENTED TIMED UP AND GO MEASURES: COMPONENTS THEY WERE EXTRACTED FROM, DESCRIPTIONS, MEASUREMENT UNITS (M.U.), AND CONSIDERED DIRECTIONS

MEASURE	COMPONENTS	DESCRIPTION	M.U.	DIRECTIONS
Duration	Total, Sit-to-Walk, Gait, Turning, Walk-to-Sit	Duration of each iTUG component.	[s]	NA
RMS	Sit-to-Walk, Gait, Turning, Walk-to-Sit	Root mean square of the acceleration (a) during the considered component: $RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(a(i) - mean(a))^2}$	[m/s^2]	AP, ML, V
NJS	Sit-to-Walk, Gait, Turning, Walk-to-Sit	Normalized jerk score of the acceleration: $NJS = \sqrt{\frac{T^5}{2}\int_{Tstart}^{Tend}(\dot{a})^2 dt}$ where T is the duration (*Tend-Tstart*) of the considered component and a is the acceleration measured in m/s^2. The NJS during Gait is computed for each step (i.e., between two consecutive heel strikes), then normalized to the step duration, and then averaged across all steps: $Gait\ NJS = \frac{1}{N}\sum_{i=1}^{N}\left(\sqrt{\frac{(hs_{i+1} - hs_i)^5}{2}\int_{hs_i}^{hs_{i+1}}(\dot{a})^2 dt}\right)$ where hs_i denotes the time of the i^{th} heel strike and a is the acceleration (m/s^2).	[m]	AP, ML, V
Tstep	Gait, Turning	Mean value of the step duration, computed as the time distance between two consecutive heel strikes.	[s]	NA
Tstep STD	Gait	Standard deviation (STD) of the step duration, computed as the time distance between two consecutive heel strikes.	[s]	NA
Tstep CV	Gait	Coefficient of variation (CV) of the step duration, computed as the time distance between two consecutive heel strikes. $Tstep\ CV = 100 \cdot \frac{Tstep\ STD}{Tstep}$	[%]	NA
Phase	Gait	Mean value of the phase denoted in degrees. The i^{th} phase (φ_i), measures the step time with respect to the stride time assigning 360° to each stride (gait cycle): $\varphi_i = 360^\circ \frac{hs_{Si} - hs_{Li}}{hs_{L(i+1)} - hs_{Li}}$ Phase is the average of φ_i, where hs_{Li} and hs_{Si} denote the time of the i^{th} heel strike of the legs with the long and short step times, respectively.	[deg]	NA
Phase STD	Gait	Standard deviation (STD) of the phase.	[deg]	NA
Phase CV	Gait	Coefficient of variation (CV) of the phase.	[%]	NA
PCI	Gait	Phase coordination index (PCI). PCI measures gait coordination (i.e., the accuracy and consistency of the phase generation). $PCI = Phase\ CV + 100 \cdot \frac{\frac{1}{N}\sum_{i=1}^{N}\lvert\varphi_i - 180^\circ\rvert}{180^\circ}$	[%]	NA
HR	Gait	Stride frequency is used as the fundamental frequency of the periodic acceleration signals during steady state walking; the fundamental period of such signals is a multiple of the stride duration. The coefficients of the first 10 even harmonics and the first 10 odd harmonics are computed by using a finite Fourier series; then the harmonic ratio (HR) is calculated by dividing the sum of the amplitudes of the in phase harmonics by the sum of the amplitudes of the out of phase harmonics: $HR = \frac{\sum_{i=1}^{10} eh_i}{\sum_{i=1}^{10} oh_i}$ for AP and V directions; $HR = \frac{\sum_{i=1}^{10} oh_i}{\sum_{i=1}^{10} eh_i}$ for ML direction where eh_i and oh_i denote the coefficient of the i^{th} even and odd harmonic.	[-]	AP, ML, V

NA = NOT APPLICABLE; AP = ANTERO-POSTERIOR; ML = MEDIO-LATERAL; V = VERTICAL, [-] = UNITLESS

E. Statistical Analysis

A repeated-measures ANOVA was used to identify differences between PD and CTRL groups, with the three consecutive trials as the *within* factor and group as the *between* factor. Since a high number of features were tested on the same dataset, we applied the correction of Benjamini & Yekutieli for multiple testing procedures [21].

Test-retest reliability was assessed separately for the CTRL and PD groups by means of Intraclass Correlation Coefficient

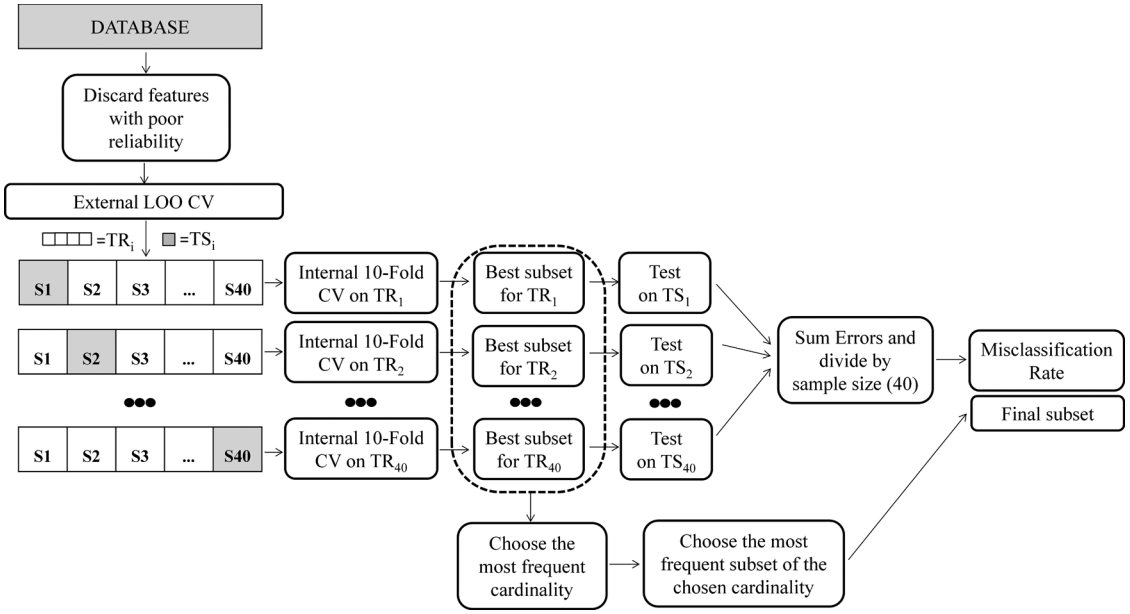

Fig. 3. Explanation of the feature selection procedure; $\mathrm{TR_i} = \mathrm{i^{th}}$ training set, $\mathrm{TS_i} = \mathrm{i}$th testing set.

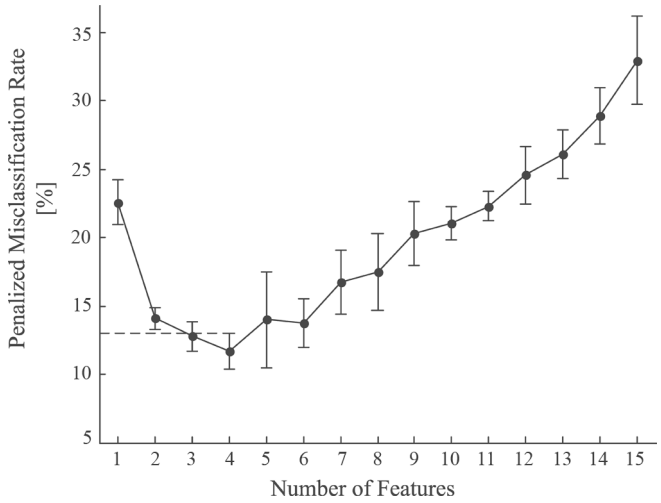

Fig. 4. Representative plot of the values of the objective function (penalized misclassification rate) of the internal CV for subsets of different cardinalities. Mean and standard deviation of the penalized misclassification rate for subsets of different cardinalities are reported.

(ICC) [22]. According to the guidelines provided in Shrout *et al.* [22], the ICC (3,1) was chosen for the analysis of reliability, because the same device (accelerometer) was used for each trial of each subject. According to [23], an ICC value lower than 0.4 is considered as an indicator of poor reliability, an ICC value between 0.4 and 0.75 as a fair to good reliability and an ICC value higher than 0.75 indicates excellent reliability [23]. On this basis, the criteria to discard unreliable features for the first step of the feature selection procedure was defined as the following: a feature must have the lower bound of the confidence interval of its ICC greater than or equal to 0.4 for both CTRL and PD groups in order to be considered reliable (see Table II); otherwise, the feature is discarded.

A Pearson's correlation analysis was performed between the acceleration-derived measures which were thus found to be reliable and the available clinical data (age, disease duration and UPDRS subscores). To this end, the average of the acceleration-derived measures across the three trials was considered for each subject. Correlation was performed separately for the PD and CTRL groups. MATLAB R2011a was used for the signal processing, feature selection, and correlation analysis. NCSS version 2007 was used for repeated-measures ANOVA and IBM SPSS version 19 was used to compute intraclass correlation coefficients.

III. Results

A. Statistical Properties of Acceleration-Derived Measures

Mean values and standard deviations of each measure (for the three consecutive trials for each subject) are reported in Table II for both CTRL and PD groups. Table II also shows statistical significances of the repeated-measures ANOVA for the *within* (consecutive trials) and the *between* (group) factors, and test-retest reliability. Significant group differences were found for six measures; however, no differences for the *within* factor were found. All ICCs are reported along with their 95% confidence interval; negative values, due to a negative average covariance among items which violates reliability model assumptions, are replaced by zeros. Fifteen measures out of 37 were found to be reliable according to the adopted criteria (see Section II-E and Table II).

B. Feature Selection and Classifiers Performance

In Table III, the results of the feature selection procedure are reported: the best classifier (the one with the lowest MR, 22.5%)

TABLE II
VALUES OF ITUG MEASURES FOR CTRL AND PD GROUPS, RESULTS OF REPEATED MEASURES ANOVA, AND RESULTS OF TEST-RETEST RELIABILITY ANALYSIS

MEASURE	CONTROL SUBJECTS		PARKINSONIAN SUBJECTS		SIGNIFICANT FACTORS
	MEAN (STD)	ICC(3,1) [LB-UB]	MEAN (STD)	ICC(3,1) [LB-UB]	
Total Duration	16.08 (2.42)	0.89 [0.74-0.95]	17.54 (2.95)	0.81 [0.65-0.91]	
STW Duration	1.24 (0.23)	0.47 [0.2-0.71]	1.3 (0.46)	0.75 [0.55-0.88]	
STW RMS AP	2.67 (0.53)	0.74 [0.53-0.87]	2.43 (0.53)	0.63 [0.4-0.82]	
STW RMS ML	0.68 (0.3)	0.38 [0.11-0.65]	0.59 (0.22)	0.55 [0.29-0.77]	
STW RMS V	1.77 (0.46)	0.61 [0.36-0.8]	1.51 (0.4)	0.78 [0.6-0.9]	
STW NJS AP	11.52 (6.71)	0.39 [0.11-0.66]	10.74 (10.42)	0.66 [0.44-0.83]	
STW NJS ML	5.9 (7.18)	0.39 [0.12-0.66]	5.02 (6.09)	0.46 [0.2-0.71]	
STW NJS V	11.23 (7.32)	0.43 [0.15-0.69]	9.88 (9.78)	0.54 [0.29-0.76]	
Gait Duration	7.79 (1.59)	0.84 [0.59-0.94]	8.72 (2.18)	0.63 [0.37-0.81]	
Gait Tstep	0.52 (0.05)	0.94 [0.83-0.98]	0.52 (0.04)	0.86 [0.74-0.94]	
Tstep STD	0.02 (0.01)	0.37 [0.11-0.64]	0.03 (0.02)	0.1 [0-0.42]	
Tstep CV	4.6 (2.09)	0.4 [0.13-0.66]	6.51 (4.3)	0.12 [0-0.44]	
Phase	179.93 (4.46)	0 [0-0]	178.72 (6.29)	0.15 [0-0.46]	
Phase STD	5.57 (2.79)	0.25 [0.01-0.54]	7.17 (5.79)	0.33 [0.05-0.61]	
Phase CV	3.1 (1.56)	0.24 [0-0.53]	4.06 (3.41)	0.31 [0.03-0.6]	
PCI	5.98 (2.96)	0.37 [0.11-0.64]	7.97 (5.24)	0.33 [0.06-0.62]	
Gait NJS AP	1.15 (0.14)	0.81 [0.65-0.91]	0.92 (0.2)	0.96 [0.92-0.98]	Group (p = 0.01)
Gait NJS ML	1.11 (0.26)	0.93 [0.86-0.97]	0.75 (0.33)	0.95 [0.9-0.98]	Group (p = 0.01)
Gait NJS V	1.1 (0.24)	0.9 [0.8-0.95]	1.08 (0.25)	0.9 [0.79-0.95]	
HR AP	1.81 (0.28)	0.72 [0.51-0.86]	1.48 (0.32)	0.92 [0.85-0.97]	Group (p = 0.03)
HR ML	2.02 (0.47)	0.71 [0.5-0.86]	1.59 (0.55)	0.83 [0.67-0.92]	
HR V	2.33 (0.43)	0.69 [0.47-0.85]	1.82 (0.41)	0.79 [0.63-0.9]	Group (p = 0.01)
Turning Duration	3.26 (0.6)	0.72 [0.52-0.87]	3.68 (0.61)	0.57 [0.31-0.78]]	
Turning Tstep	0.56 (0.05)	0.86 [0.73-0.94]	0.56 (0.07)	0.87 [0.75-0.94]	
Turning RMS AP	1.58 (0.34)	0.73 [0.52-0.87]	1.43 (0.42)	0.92 [0.84-0.96]	
Turning RMS ML	1.61 (0.44)	0.86[0.74-0.94]	1.44 (0.37)	0.91 [0.82-0.96]	
Turning RMS V	1.96 (0.45)	0.79 [0.61-0.9]	1.87 (0.51)	0.93 [0.85-0.97]	
Turning NJS AP	1.11 (0.2)	0.8 [0.63-0.91]	0.87 (0.21)	0.61 [0.36-0.81]	Group (p = 0.01)
Turning NJS ML	1.13 (0.28)	0.81 [0.64-0.91]	0.78 (0.27)	0.84 [0.71-0.93]	Group (p = 0.01)
Turning NJS V	1.08 (0.26)	0.84 [0.7-0.93]	1.12 (0.35)	0.78 [0.61-0.9]	
WTS Duration	3.46 (0.53)	0.52 [0.25-0.75]	3.6 (0.59)	0.52 [0.26-0.75]	
WTS RMS AP	2.53 (0.43)	0.68 [0.46-0.85]	2.32 (0.42)	0.55 [0.28-0.77]	
WTS RMS ML	1.41 (0.44)	0.85 [0.71-0.93]	1.15 (0.27)	0.46 [0.19-0.71]	
WTS RMS V	1.78 (0.47)	0.74[0.54-0.88]	1.48 (0.35)	0.63 [0.39-0.82]	
WTS NJS AP	224 (87.8)	0.08 [0-0.39]	193,8 (102,5)	0.45 [0.18-0.7]	
WTS NJS ML	230.5 (90.2)	0.28 [0.01-0.57]	180.2 (93.4)	0.37 [0.1-0.65]	
WTS NJS V	267.2 (92.2)	0.21 [0-0.51]	249.5 (101.6)	0.27 [0.02-0.56]	

THE MEASURES WHICH WERE CONSIDERED TO BE RELIABLE (WITH AN ICC LOWER BOUND GREATER THAN OR EQUAL TO 0.4 FOR BOTH THE CTRL AND PD GROUPS), ARE HIGHLIGHTED IN GRAY.
STD = STANDARD DEVIATION; ICC = INTRACLASS CORRELATION COEFFICIENT; LB = LOWER BOUND OF 95% CONFIDENCE INTERVAL; UB = UPPER BOUND OF 95% CONFIDENCE INTERVAL

is the LDA classifier; the corresponding best subset is made of one measure from STW and two measures from Turning. Two measures from Gait are selected by the second-best classifier (QDA, 27.5%). The MC shows the worst MR.

C. Correlation Between Acceleration-Derived Measures and Clinical Data

In the PD group: i) Total Duration positively correlates with the gait and posture subscore $(\mathrm{r} = 0.6, \mathrm{p} = 0.005)$, with the

TABLE III
FOR EACH CLASSIFIER THE FINAL SELECTED SUBSET IS REPORTED WITH CORRESPONDING MISCLASSIFICATION RATE

CLASS	FINAL SUBSETS Measure			MISCLASSIFICATION RATE % [CI]
LDA	STW RMS AP	Turning NJS ML	Turning NJS V	22.5 [12.3-37.5]
QDA	Gait NJS AP	Gait HR V		27.5 [16.1-42.8]
MC	Gait NJS AP	Gait HR V	Turning RMS AP	37.5 [24.2-53]

CI = 95% CONFIDENCE INTERVAL; LDA = LINEAR DISCRIMINANT ANALYSIS; QDA = QUADRATIC DISCRIMINANT ANALYSIS; MC = MAHALANOBIS CLASSIFIER

rigidity subscore ($\mathrm{r} = 0.46, \mathrm{p} = 0.04$), and with the PIGD subscore ($\mathrm{r} = 0.51, \mathrm{p} = 0.02$); ii) HR, in both the ML and V directions, negatively correlates with the gait and posture subscore ($r = -0.47, p = 0.037$ and $r = -0.56, p = 0.01$ respectively); iii) Turning RMS in the V direction negatively correlates with the gait and posture subscore ($\mathrm{r} = -0.58, \mathrm{p} = 0.007$) and with the rigidity subscore ($\mathrm{r} = -0.47, \mathrm{p} = 0.038$).

In the CTRL group both Gait NJS and Turning NJS in the ML direction are negatively correlated with age ($r = -0.55$, $p = 0.014$ and $r = -0.54, p = 0.018$).

IV. DISCUSSION

A. Instrumented Timed Up and Go

The iTUG used in the present study is a modified version of the traditional TUG with an increase in walking distance from 3 to 7 m [7], [9] and a single accelerometer worn on the lower back [8], [11]. The increased walking distance was chosen to obtain a better characterization of the steps from the acceleration signals [9], and the setup with a single accelerometer was chosen to obtain a simple and low-cost protocol [8].

B. Characteristics of Acceleration-Derived Measures

As expected [5], the total duration of the iTUG is characterized by an excellent test-retest reliability ($\mathrm{ICC} > 0.8$, see Table II) but it is not sensitive to group differences for early stages of the disease (as in Zampieri *et al.* [7]). However, it positively correlates with the severity of clinical subscores of gait and posture impairment and rigidity in PD subjects. As shown in Table II, the average total trial duration of a CTRL subject (16.08 ± 2.42 s) does not significantly differ from that of a PD subject (17.54 ± 2.95 s). It could be argued that the lack of sensitivity observed in the current study is due to the differences in walking distances, since a 7-m iTUG was used instead of the traditional 3-m TUG; however, Zampieri *et al.* [7] already showed that in a comparable population the traditional 3-m TUG did not show any significant group differences.

In the Sit-to-Walk component all the measures were discarded because of low reliability with the exception of the RMS in the AP direction.

In the Gait component, all traditional measures related to gait variability (variability of Tstep, variability of Phase, PCI) show poor reliability and no group differences. A possible explanation of this result is that the 7-m distance (even if was increased with respect to the traditional 3-m distance) was not enough to obtain an adequate number of steps (or strides).

On the other hand, Gait Tstep along with most of the normalized jerk scores and harmonic ratios showed good to excellent reliability. Interestingly, Gait Tstep showed no group differences, in contrast with Salarian *et al.* [9] where PD subjects of comparable age and disease severity (comparable Hoehn & Yahr stages and motor-UPDRS scores) and healthy age-matched subjects were considered. The three main differences between the two data sets that could explain the contrasting results are that in Salarian *et al.* [9]: i) the PD subjects had never taken anti-Parkinson medications; ii) no corrections for multiple testing procedures were applied; and iii) the sample size was smaller (12 PD, 12 CTRL).

Unlike Tstep, normalized jerk scores and harmonic ratios showed significant group differences. PD subjects are characterized by lower values of harmonic ratios both in AP and V directions (see Table II); low values of harmonic ratios may be due to low smoothness and rhythmicity of trunk motion while walking [24]. A possible explanation is that in healthy subjects the foot contact is characterized by a smooth heel-to-toe pattern, while this may already be degraded in early-mild PD, thus leading to a prominent flat-footed gait and a reduced rolloff [25], [26]. The observed reduction of HR values in PD subjects is consistent with previous literature results, which support the interpretation that a reduction in HR values reflects degradation in walking stability and motor control [24], [27]–[31]. Further supporting this interpretation, we found that PD subjects with increasing gait and posture impairments have decreasing HR values, as reflected by their negative correlation with the gait and posture subscore (Section III-C). It has to be noted that Lowry *et al.* [24] found speed-related changes in HR values; in the present study this was not an issue because the gait velocity (which is proportional to gait time, since the distance is fixed) did not differ between CTRL and PD (Table II).

The normalized jerk score during Gait measures the time-normalized dynamic of the acceleration signal during a step: fast and large variations in the signal lead to high values of this measure. Lower NJS values were found in PD subjects both for the AP and ML direction. This reduced dynamic of the acceleration signals (low and small variations) could be related to a general loss of complexity of the motor control system, leading to a loss of adaptivity in motor strategies. Besides being lower in PD with respect to CTRL, Gait NJS in the ML direction tends to decrease as age increases in CTRL subjects.

In the Turning component, most of the measures were sufficiently reliable (Table II) but only one was both reliable and significantly different between groups: the NJS in the ML direction. This measure was found to be lower in PD subjects. As in

the Gait component, the dynamic of the ML acceleration during Turning was found to be reduced in PD with respect to CTRL and to decrease as age increases in CTRL subjects.

In the Walk-to-Sit component none of the measures was sufficiently reliable; this is probably because a single accelerometer was not enough to accurately segment this component. It is possible that the addition of a gyroscope would improve the quality of the segmentation and also permit the extraction of new, reliable gyroscope-based measures.

C. Accuracy in Discriminating Between CTRL and PD Subjects

A subset of three reliable features from the Sit-To-Walk and Turning components (STW RMS AP, Turning NJS ML, and Turning NJS V) could obtain the best misclassification rate (22.5%), in combination with the LDA classifier. In order to have a generalizable result, the final misclassification rate estimate was obtained using an independent external cross validation (not the same as the internal cross validations that are used only to select the best features). This allowed us to avoid an optimistic estimate [19]. An example of this optimistic bias can be seen in Fig. 4; the misclassification rate in one of the internal cross validations, even if it is penalized, would show values as low as 12.5% (10% lower than the final MR estimate of the LDA classifier).

The LDA classifier, which performed the best among the tested classifiers, has the additional advantage of being easy to interpret and visualize, especially for subsets made of few features: with the three selected features the classifier is a discriminating plane that divides the subjects' values in two parts, as shown in Fig. 5.

The second-best subset is made of two Gait measures selected by the quadratic classifier, which can discriminate between the two populations with a misclassification rate of 27.5%.

In contrast, the Mahalanobis classifier has the worst accuracy, with a higher bound of the confidence interval of misclassification rate over 50% (see Table III); it is therefore not suitable for discriminating between the two considered populations.

D. Limitations and Future Developments

The relatively small sample size is a limitation, reflected by the large confidence intervals of intraclass correlation coefficients in Table II and of misclassification rates in Table III. Although a nested cross validation was used to obtain a fair (rather than optimistic) estimate of the final accuracy, the small sample size reduces the generalizability of the results which should be verified in a larger sample. Correlation results could also be biased by the small sample size. This issue could be overcome by future studies implementing this setup and protocol in additional PD subjects. In this regard, it is suggested for future studies using more sensors to position one of the sensors on the lower back as in the present study. In this way the future studies will be directly comparable with this one and with similar single-sensor studies.

As already mentioned in Section IV-B, a future development could be to consider inertial sensors including a gyroscope and an accelerometer. This could improve the quality of the segmentation of the different components of the tests and the reliability

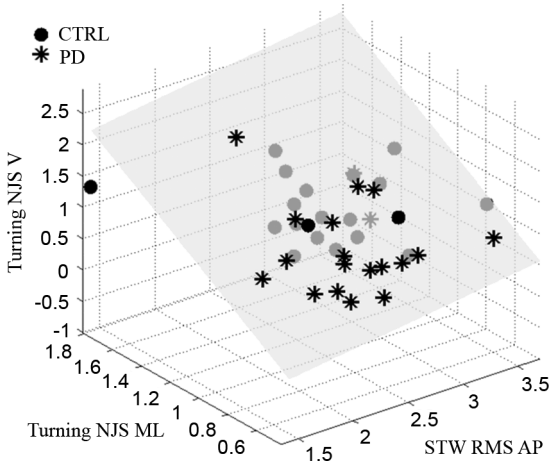

Fig. 5. Scatter plot of the values of the feature subset selected by the LDA classifier: STW RMS AP, Turning NJS ML, and Turning NJS V. The discrimination plane resulting from the LDA classifier is shown.

of the existing measures. In addition, it could allow the extraction of new gyroscope-based measures. It is possible that these new features could improve the best misclassification rate that was achieved with this study.

In this study some of the acceleration-derived measures were found to be associated with clinical measures of disease severity. Future studies should consider different stages of the disease along with longitudinal monitoring to assess the sensitivity of acceleration-derived measures to disease severity and progression. In addition, the proposed method (single accelerometer-based iTUG and feature selection) might prove useful for quantifying the locomotor function of populations characterized by different motor impairments.

Finally, instrumented TUG, given its simplicity and reproducibility, can be used outside the clinical setting; in a pilot study, iTUG has already been used for remote home monitoring [32]. For this kind of application a smartphone-based version could also be considered, which uses the smartphone's accelerometer as the measurement device [33]. This would further improve the ubiquity, ease of use, and cost of the protocol.

V. Conclusion

In this study, the locomotor function of early-mild PD subjects was quantitatively evaluated using a minimal setup (a single accelerometer) to measure the performance of a simple clinical test (TUG). Using the reliable measures that were extracted from the acceleration signals in different components of the test (Total Test, Sit-to-Walk, Gait, Turning, Walk-to-Sit), we made the following observations: the performance of early-mild PD subjects OFF-medication is characterized by normal temporal measures (duration of the test, average step duration), a reduced smoothness and dynamics in trunk movement during gait, and reduced lateral dynamics in trunk movement during turning.

In addition to these results, a combination of three measures (one from Sit-to-Walk, two from Turning) was found to be able to discriminate, with the aid of a simple and easily interpretable classification algorithm, between the locomotor pattern of CTRL and PD subjects. This discrimination is achieved with

a misclassification rate of 22.5% in the early-mild stage of the disease, when the temporal outcome of the traditional TUG test would not be able to discriminate between the healthy and the pathological performance.

Finally, aside from being able to discriminate between healthy and PD subjects, these results suggest that the proposed iTUG plausibly characterizes PD locomotor impairment and, hence, may be useful for evaluation, follow-up, and monitoring of therapies and disease progression.

ACKNOWLEDGMENT

The authors would like to thank L. Codeluppi and V. Fioravanti from the Department of Neuroscience, University of Modena and Reggio Emilia, Modena, Italy, for clinical supervision and assistance in data acquisition. They would also like to thank K. Mayberry for the linguistic revision of the manuscript and D. G. Calò for the assistance with statistical procedures.

REFERENCES

[1] D. Podsiadlo and S. Richardson, "The timed 'Up & Go': A test of basic functional mobility for frail elderly persons," *J. Amer. Geriatrics Soc.*, vol. 39, no. 2, pp. 142–148, Feb. 1991.

[2] A. Shumway-Cook, S. Brauer, and M. Woollacott, "Predicting the probability for falls in community-dwelling older adults using the timed Up & Go test," *Phys. Therapy*, vol. 80, no. 9, pp. 896–903, Sep. 2000.

[3] L. K. Boulgarides, S. M. Mcginty, J. A. Willett, and C. W. Barnes, "Research report use of clinical and impairment-based tests to predict falls by community-dwelling older adults," *Phys. Therapy*, pp. 328–339, 2003.

[4] T. Herman, N. Giladi, and J. M. Hausdorff, "Properties of the 'Timed Up and Go' test: More than meets the eye," *Gerontology*, vol. 57, no. 3, pp. 203–210, Jan. 2011.

[5] S. Morris, M. E. Morris, and R. Iansek, "Reliability of measurements obtained with the timed 'Up & Go' test in people with Parkinson's disease," *Phys. Therapy*, vol. 81, no. 2, pp. 810–818, Feb. 2001.

[6] Y. Balash, C. Peretz, G. Leibovich, T. Herman, J. M. Hausdorff, and N. Giladi, "Falls in outpatients with Parkinson's disease: Frequency, impact and identifying factors," *J. Neurol.*, vol. 252, no. 11, pp. 1310–1315, Nov. 2005.

[7] C. Zampieri, A. Salarian, P. Carlson-Kuhta, K. Aminian, J. G. Nutt, and F. B. Horak, "The instrumented timed Up and Go test: Potential outcome measure for disease modifying therapies in Parkinson's disease," *J. Neurology, Neurosurgery, Psychiatry*, vol. 81, no. 2, pp. 171–176, Feb. 2010.

[8] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, I. Maidan, N. Giladi, T. Gurevich, and J. M. Hausdorff, "Can an accelerometer enhance the utility of the timed Up & Go test when evaluating patients with Parkinson's disease," *Medical Eng. Phys.*, vol. 32, no. 2, pp. 119–125, Mar. 2010.

[9] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, "ITUG, a sensitive and reliable measure of mobility," *IEEE Trans. Neural Syst. Rehabi. Eng.*, vol. 18, no. 3, pp. 303–310, Jun. 2010.

[10] B. R. Greene, A. O'Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R. A. Kenny, "Quantitative falls risk assessment using the timed Up and Go test," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2918–2926, Dec. 2010.

[11] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, N. Giladi, and J. M. Hausdorff, "An instrumented timed Up and Go: The added value of an accelerometer for identifying fall risk in idiopathic fallers," *Physiological Measurement*, vol. 32, no. 12, pp. 2003–2018, Dec. 2011.

[12] B. R. Greene and R. A. Kenny, "Assessment of cognitive decline through quantitative analysis of the timed Up and Go test," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 988–995, Apr. 2012.

[13] L. Palmerini, L. Rocchi, S. Mellone, F. Valzania, and L. Chiari, "Feature selection for accelerometer-based posture analysis in Parkinson's disease," *IEEE Trans. Inform. Technol. Biomedicine*, vol. 15, no. 3, pp. 481–490, May 2011.

[14] J. Jankovic, M. McDermott, J. Carter, S. Gauthier, C. Goetz, L. Golbe, S. Huber, W. Koller, C. Olanow, and I. Shoulson, "Variable expression of Parkinson's disease: A base-line analysis of the DATATOP cohort. The Parkinson study group," *Neurology*, vol. 40, no. 10, pp. 1529–1534, Oct. 1990.

[15] W. Zijlstra, "Assessment of spatio-temporal parameters during unconstrained walking," *Eur. J. Appl. Physiol.*, vol. 92, no. 1–2, pp. 39–44, Jun. 2004.

[16] B. Caby, S. Kieffer, M. de Saint Hubert, G. Cremer, and B. Macq, "Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry," *Biomed. Eng. Online*, vol. 10, no. 1, pp. 1–19, Jan. 2011.

[17] M. Plotnik, N. Giladi, and J. M. Hausdorff, "A new measure for quantifying the bilateral coordination of human gait: Effects of aging and Parkinson's disease," *Experimental Brain Res.*, vol. 181, no. 4, pp. 561–70, Aug. 2007.

[18] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, "Acceleration patterns of the head and pelvis when walking on level and irregular surfaces," *Gait Posture*, vol. 18, no. 1, pp. 35–46, Aug. 2003.

[19] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intell.*, vol. 97, no. 1, pp. 273–324, Dec. 1997.

[20] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. New York, NY, USA: Morgan Kaufmann, 2005, p. 560.

[21] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, Aug. 2001.

[22] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychol. Bull.*, vol. 86, no. 2, pp. 420–428, 1979.

[23] B. Rosner, *Fundamentals of Biostatistics*, 7th ed. Belmont, CA, USA: Duxbury Press, 2010, p. 569.

[24] K. A. Lowry, N. Lokenvitz, and A. L. Smiley-Oyen, "Age- and speed-related differences in harmonic ratios during walking," *Gait Posture*, vol. 35, no. 2, pp. 272–276, Feb. 2012.

[25] R. Baltadjieva, N. Giladi, L. Gruendlinger, C. Peretz, and J. M. Hausdorff, "Marked alterations in the gait timing and rhythmicity of patients with de novo Parkinson's disease," *Eur. J. Neurosci.*, vol. 24, no. 6, pp. 1815–1820, Sep. 2006.

[26] J. R. Hughes, S. G. Bowes, A. L. Leeman, C. J. O'Neill, A. A. Deshmukh, P. W. Nicholson, S. M. Dobbs, and R. J. Dobbs, "Parkinsonian abnormality of foot strike: A phenomenon of ageing and/or one responsive to levodopa therapy," *Brit. J. Clin. Pharmacol.*, vol. 29, no. 2, pp. 179–186, Feb. 1990.

[27] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, "Acceleration patterns of the head and pelvis when walking are associated with risk of falling in community-dwelling older people," *J. Gerontology Series A: Biological Sci. Medical Sci.*, vol. 58, no. 5, pp. M446–M452, May 2003.

[28] H. B. Menz, S. R. Lord, R. S. George, and R. C. Fitzpatrick, "Walking stability and sensorimotor function in older people with diabetic peripheral neuropathy," *Arch. Phys. Medicine Rehabil.*, vol. 85, no. 2, pp. 245–252, Feb. 2004.

[29] M. D. Latt, H. B. Menz, V. S. Fung, and S. R. Lord, "Acceleration patterns of the head and pelvis during gait in older people with Parkinson's disease: A comparison of fallers and nonfallers," *J. Gerontology, Series A, Biological Sci. Medical Sci.*, vol. 64, no. 6, pp. 700–706, Jun. 2009.

[30] K. A. Lowry, A. L. Smiley-Oyen, A. J. Carrel, and J. P. Kerr, "Walking stability using harmonic ratios in Parkinson's disease," *Movement Disorders*, vol. 24, no. 2, pp. 261–7, Jan. 2009.

[31] H. J. Yack and R. C. Berger, "Dynamic stability in the elderly: Identifying a possible measure," *J. Gerontol.*, vol. 48, no. 5, pp. M225–230, Sep. 1993.

[32] C. Zampieri, A. Salarian, P. Carlson-Kuhta, J. G. Nutt, and F. B. Horak, "Assessing mobility at home in people with early Parkinson's disease using an instrumented timed Up and Go test," *Parkinsonism Related Disorders*, vol. 17, no. 4, pp. 277–280, May 2011.

[33] S. Mellone, C. Tacconi, and L. Chiari, "Validity of a smartphone-based instrumented timed Up and Go," *Gait Posture*, vol. 36, no. 1, pp. 163–165, May 2012.

Luca Palmerini was born in Verona, Italy, in 1981. He received the Laurea degree in computer science engineering from the University of Padua, Padua, Italy, in 2007, and the Ph.D. degree in biomedical engineering from the University of Bologna, Bologna, Italy, in 2012.

He is presently a Postdoctoral Researcher in biomedical engineering in the Department of Electrical, Electronic and Information Engineering, University of Bologna. His main research interests include data mining, statistics, and their applications to human movement analysis.

Sabato Mellone received the Laurea degree in electronic engineering, in 2008. He is currently working toward the Ph.D. degree in biomedical engineering in the Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy.

His main research activity is in the area of human movement analysis, signal processing, wearable sensors, ICT in clinical practice, Personal Health Systems, and mobile health.

Guido Avanzolini was born in Forlí, Italy, in 1942. He received the Dr. Ing. degree in electrical engineering from the University of Bologna, Bologna, Italy, in 1967.

He is presently with the Department of Electrical, Electronic and Information Engineering, University of Bologna, where he is a Full Professor of biomedical instrumentation, past Coordinator of the Ph.D. program in biomedical engineering and past Chairman of the Department. His main research interests include measurement, modeling and control of physiological systems, in particular cardiorespiratory and renal.

Franco Valzania received the Laurea degree in medicine and the specialist degree in neurology from the University of Bologna, Bologna, Italy, in 1987 and 1991, respectively.

He is Director of the Neurophysiologic Unit of NOCSAE Hospital, University of Modena and Reggio Emilia, Modena, Italy. His main research activity is in the area of movement disorders, deep brain stimulation, neurophysiology and neurobiomechanics of human posture and movement, intraoperatory monitoring. He was a Reviewer for *EEG* and *Clinical Neurophysiology Journal*. He is author of about 150 papers published on international journals, conference proceedings, and book chapters.

Lorenzo Chiari received the Laurea degree in electronic engineering and the Ph.D. degree in biomedical engineering from the University of Bologna, Bologna, Italy, in 1993 and 1997, respectively.

He is an Associate Professor of biomedical engineering in the Department of Electrical, Electronic and Information Engineering, University of Bologna. His main research activity is in the area of neurobiomechanics of human posture and movement, multisensory integration and sensory augmentation, neuromuscular control of posture in health and disease, rehabilitation engineering. He is a member of the editorial board of the journal *Gait & Posture* and regular Reviewer for several peer-reviewed international journals. He served as external reviewer for NASA, ESA, Health Research Board Ireland, Ataxia U.K., MIUR, and other International Health Organizations. He is author of about 200 papers published on international journals, conference proceedings, and book chapters.