Towards fully instrumented and automated assessment of motor function tests

Valeria De Luca*[1], Amir Muaremi[1], Oonagh M. Giggins[2], Lorcan Walsh[3] and Ieuan Clay[1]

***Abstract*— Quantitative assessment of mobility and motor function is critical to our understanding and treatment of musculoskeletal and neurological diseases. Instrumented tests augment traditional approaches by moving from a single, often subjective, performance metric to multiple objective measures. In this study, we investigated ways of automatically capturing motor performance by leveraging data from a network of six wearable sensors worn at five different locations by 17 healthy volunteers while performing a battery of motor function tests. We developed a framework to segment motor tasks, e.g. walking and standing up, from 3D acceleration and angular velocity data, and extracted features. Results were compared to clinical test scores and manual annotations of the data. For the best performing sensors, we achieved a rate of correct classification of 82 to 100% and mean temporal accuracy of 0.1 to 0.6 s. We provided guidelines on sensor placement to maximize accuracy of the motor assessment, and a better interpretation of the data using our unsupervised subject-specific approach.**

***Index Terms*— Mobility, physical activity, activity monitoring, wearables.**

I. INTRODUCTION

Mobility limitations and declining motor function are associated with reduced independence and disability [17], longer hospital stays [2], nursing home placement [18], and mortality [17]. Measuring physical activity, and particularly mobility, allows clinicians to understand a patient's functional ability in order to decide upon treatment or prognosis. Clinical practice for assessing motor function has traditionally relied on subjective questionnaires, non-granular surveys, non-quantitative activity diaries or performance testing with observational scoring. More precise and fine grained instrumented measures are poised to enable better characterization of motor function. Wearable inertial measurement unit (IMU) sensors offer an objective, quantitative, portable, flexible and moderately-priced alternative to expensive and large laboratory systems such as walkway (e.g. GAITRite, CIR Systems, Inc., USA [6]) or vision-based systems (e.g. Vicon motion capture system, Vicon Industries, Inc., UK [21]). Significant work has been undertaken to instrument standard physical tests (gait [14], [16], timed up and go (TUG) [15], balance and sit to stand (STS) [8]) using specifically positioned sensors and associated algorithms to quantify motor function through summary statistics or quantifying performance in test subcomponents. Such approaches represent an entry point into deploying similar instrumented assessments outside of the clinic, in the patients everyday life, further increasing our ability to characterize how disease state may fluctuate longitudinally or respond to therapy.

Several reviews have investigated the validity and reliability of physical activity sensors across various conditions including neurological disorders [12], chronic disease [20], chronic obstructive pulmonary disease [5], chronic lung disease [7], stroke [11] and Parkinson's disease [13]. Broadly, these reviews found heterogenous study designs and methods, inconsistent outcome measures and technologies (e.g. sensor types, configuration, placement and orientation), custom algorithms and inconsistent reporting methods, undermining the optimal selection, configuration and placement of sensors for a given instrumented physical activity test.

In this work, we provide guidelines on sensor placement to maximize the accuracy and robustness of motor assessment, based on data simultaneously acquired from six IMU sensors worn at five different anatomical positions while performing standardized motor function tests. We propose methods to segment subtasks in each test, i.e. walking, turning, standing up and sitting down, and derive clinically-relevant metrics for motor function assessment. Our approach is unsupervised and automatically adapts to each subject in order to be easily extended to different patient cohorts. We will support further development of research in the field by making our data and code publicly available at `https://github.com/Novartis`.

II. DATA

Twenty healthy subjects (11 male, 9 female, 28.6 $\pm$ 4.3 years) were recruited by the University College Dublin, Ireland. Ethical approval was granted by the Human Research and Ethics Committee, University College Dublin and written consent given by all participants. Each participant was asked to wear six IMU sensors:

- One actibelt (Trium Analysis Online GmbH, Germany) centered on the waist (act) (tri-axial accelerometer of range $\pm$6 g and sample rate of 100 Hz)
- One BioStampRC (MC10, Inc., USA) on the chest (bst) (tri-axial accelerometer of range $\pm$4 g, tri-axial gyroscope of $\pm$2000 deg/s and sample rate of 125 Hz)
- Four Shimmer3 (Shimmer, Ireland), each on the waist (shA), lower back (shP), left (shL) and right ankle (shR) (tri-axial low-noise accelerometer of range $\pm$2 g, wide-range accelerometer of range $\pm$8 g and tri-axial gyroscope of $\pm$2000 deg/s; sample rate of 102.4 Hz)

[1]Valeria De Luca, Amir Muaremi and Ieuan Clay are with the Novartis Institutes for Biomedical Research, Basel, Switzerland.

[2]Oonagh M. Giggins is with the Insight Centre for Data Analytics, University College Dublin, Ireland.

[3]Lorcan Walsh is with Novartis Business Services, Dublin, Ireland.

*Corresponding author. `valeria.de_luca@novartis.com`

Under supervision of a clinician and with short breaks in between each test, the participants performed a sequence of standardized physical assessment tests, including:

- Short Physical Performance Battery (SPPB) [10] gait speed, chair stand and repeated chair stand tests (5 times sit to stand; 5STS)
- Timed Up and Go (TUG) test [3]
- Two minute walking test (2MWT)

Data from three of the 20 subjects were discarded in this work, due to partial missing data (e.g. data from a subset of sensors was not recorded).

An Android tablet application time-synced with the actibelt was used to define start and end times of each test and synced to the other sensor data post-data collection, see Sec. III-A. Standard clinical practice measures of time and distance were also captured using a stopwatch and measuring tape. Before fitting the subject with IMUs for each data acquisition, the clinician simultaneously shook all the IMUs.

III. METHOD

For each sensor $s \in \{act; bst; shA; shP; shL; shR\}$, subject $i \in [1, \ldots, 17]$ and time $t_{i,s} \in [t_{0,i,s}, \ldots, T_{i,s}]$, let us denote 3D accelerations as $\mathbf{a}_{i,s}(t_{i,s}) = [a_{x,i,s}(t_{i,s}); a_{y,i,s}(t_{i,s}); a_{z,i,s}(t_{i,s})] \in \mathbb{R}^3$ and angular velocities as $\mathbf{g}_{i,s}(t_{i,s}) = [g_{x,i,s}(t_{i,s}); g_{y,i,s}(t_{i,s}); g_{z,i,s}(t_{i,s})] \in \mathbb{R}^3$, with $T_{i,s}$ being the duration of each dataset.

A. Pre-processing

Data was temporally aligned in order to account for varying sampling rate and internal clocks across the sensors using the following steps. Firstly, accelerations $\mathbf{a}_{i,s}$ and angular velocities $\mathbf{g}_{i,s}$ were linearly resampled to 100 Hz. For simplicity, we will use the same notation for $\mathbf{a}_{i,s}(t_{i,s})$ and $\mathbf{g}_{i,s}(t_{i,s})$ after resampling. Then, at each time t_s, we computed the acceleration magnitude as $\bar{a}_{i,s}(t_{i,s}) = \sqrt{a^2_{x,i,s}(t_{i,s}) + a^2_{y,i,s}(t_{i,s}) + a^2_{z,i,s}(t_{i,s})}$. For each i, we manually selected start and end times of shaking patterns prior to the assessments from $\bar{\mathbf{a}}_{i,act}$, $t^{start}_{i,act}$ and $t^{end}_{i,act}$, respectively. Alignment was obtained by finding $\tau_{i,s} \in \{t_{i,s}\}$ that maximized the correlation coefficient between the reference actibelt window $\{\bar{a}_{i,act}(t^*_i)\}$, with $t^*_i \in [t^{start}_{i,act}, \ldots, t^{end}_{i,act}]$, and a rolling window of the remaining sensors' $\bar{\mathbf{a}}_{i,s}$ of step 1 and same length. Finally, we used annotations of start and end time of each test captured from the actibelt system to exclude data acquired outside the tests.

B. Euler-Angles calculation

We derived another set of data channels corresponding to the three Euler angles; roll (ϕ), pitch (θ) and yaw (ψ), which describe the orientation of the inertial sensor: $\mathbf{r}_{i,s}(t_i) = [\phi_{i,s}(t_i); \theta_{i,s}(t_i); \psi_{i,s}(t_i)] \in \mathbb{R}^3$. We based our method on the open source code of the OpenShoe project available at http://www.openshoe.org. Starting from an arbitrary position and heading $\psi = 0$, the initial ϕ and θ are calculated from the acceleration data of the first samples where we assume that the sensor is stationary (not moving). Then, for each t_i, $\mathbf{a}_{i,s}(t_i)$ and $\mathbf{g}_{i,s}(t_i)$ are read and the inertial navigation system: (i) estimates orientation through gyro integration; (ii) transforms accelerations from sensor frame into navigation frame; (iii) removes gravity to get linear acceleration and integrates it to get the velocity; and (iv) integrates velocity to get the position. We addressed accumulation of position estimate error due to double integration by applying the Kalman filter, which is based on the assumption that during the stance phase the foot is not moving and hence the sensor velocity is zero [9]. We declare a time segment to be a stance when the angular rate energy falls below a threshold [19]. An example of the nine data channels $\mathbf{C}_{i,s} = \{c_{i,s}(t_i)\} = \{\mathbf{a}_{i,s}(t_i); \mathbf{g}_{i,s}(t_i); \mathbf{r}_{i,g}(t_i)\}$ of a TUG test is shown in Fig. 1.

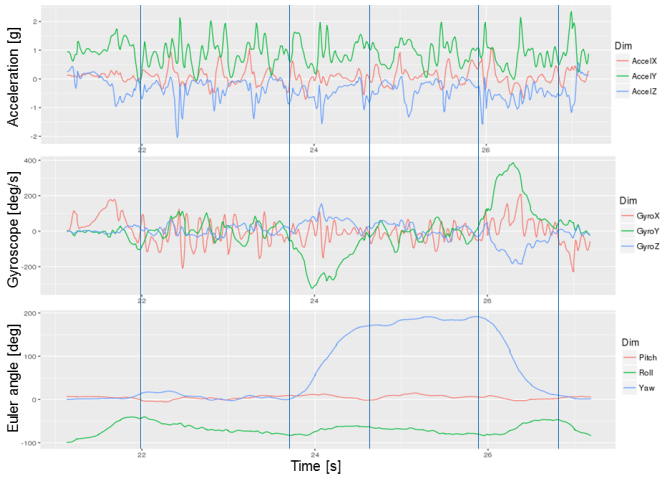

Fig. 1: Example of data channels $\mathbf{C}^{TUG}_{17,shA}$ of TUG test from subject 17 and Shimmer3 sensor worn on the waist (shA). Vertical lines indicate the different tasks (standing, walking, turning, walking back and turning to sit).

C. Unsupervised classification of subtasks

The aim of the proposed method is to segment subtasks performed during motor function assessment tests, i.e. walking, turning, standing up and sitting down. We used k-means clustering [4] on IMU data $\mathbf{C}_{i,s}$ to classify the aforementioned subtasks implementing two independently trained two-class k-means clustering representations. 2MWT data was used to train one k-means clustering representation for walking and turning, while STS data to train the second representation for sitting and standing. These two models were then tested on TUG data for classifying the four tasks of walking, turning, sitting and standing. Details of our approach are described below.

2MWT: For each subject i and sensor s independently, we used two-class k-means clustering to classify each 2MWT data point $c_{i,s}(t)$ as walking or turning. We extract a set of features from a sliding window $\hat{\mathbf{c}}_{i,s}(t) = [c_{i,s}(t - \Delta t); \ldots; c_{i,s}(t)]$ of length L_s and step size S_s. Features were mean, standard deviation (std), 5th and 95th percentile of $\hat{\mathbf{c}}_{i,s}(t)$. Only std was considered for euler angles. The final class assignment (walking vs. turning) was based on the assumption that walking occurs for a longer time than turning. As a result, the test data $\mathbf{C}^{2MWT}_{i,s} = \{\mathbf{c}^{2MWT}_{i,s}\}$ was

divided into $J_{i,s}^{2MWT} = J_{i,s}^{walk} + J_{i,s}^{turn}$ segments.

STS: The SPPB chair test is divided into two tests, a single sit-to-stand (STS) and a sequence of five repeated sit-to-stand transitions (5STS). We considered the single STS as a template to classify the repeated 5STS. Single tasks (i.e. standing up in the single STS and standing up and sitting down in 5STS) were segmented for each i and $s \in \{bst, shA, shP\}$ independently, by detecting local minima from the roll channel (ϕ_s). We excluded act as no gyroscope data was available, and shL and shR as no motion was detected from these sensors in this motor test as the participants feet are static. We then classified each segment by using two-class k-means clustering. In this case, the input of k-means was the Dynamic Time Warping [1] between the reference stand-up segment from the single STS test and each segment of all data channels $\hat{c}_{i,s}$. The class characterized by the smallest magnitude of its centroid (i.e. closest to the reference) was labeled as stand-up, sit-down otherwise. Similarly to 2MWT, $\mathbf{C}_{i,s}^{5STS}$ was finally divided into $J_{i,s}^{5STS} = J_{i,s}^{stand} + J_{i,s}^{sit}$ segments.

TUG: We tested the clustering approaches on the TUG test, as it is a combination of the aforementioned tasks. For each i and s we considered the class centroid $\mathbf{\Gamma}_{i,s}^{class}$ for all subtasks. Centroids of classes standing-up and sitting-down were learned from clustering of 5STS, while the ones for walking and turning from clustering of 2MWT. Similarly to 2MWT, we applied a rolling window $\hat{\mathbf{c}}_{i,s}(t)$ of length L_s and step size S_s to the test data. For each window n, we projected the data onto the feature space of the two clustering models and computed the Euclidean distances to all centroids $\mathbf{\Gamma}_{i,s}^{class}$ ($\delta\Gamma_{i,s,n}^{class}$). For each n, its class was predicted based on $\delta\Gamma_{i,s,n}^{class}$ being below a thresholds θ_s^{class}, derived from statistics of the distances of all subjects i in the training data (2MWT and 5STS).

D. Sensor-derived clinical metrics

For tests involving walking, i.e. 2MWT, TUG and SPPB 3-meter gait (SPPB 3m) tests, we calculated the mean gait speed ($v_{i,s}$) and distance walked ($d_{i,s}$). We used methods provided by the actibelt system for act data, and the inertial pedestrian tracking approach as described in III-B for the shoe-mounted IMUs.

E. Evaluation Criteria

Subtask segmentation results (see Sec. III-C) were quantitatively evaluated for all subjects with respect to the mean manual annotation of raw data segments of walking shuttles and standing-up subtasks by three expert raters. For all segmented motor assessment tests, we calculated three error metrics, namely:

- Overall classification accuracy (*ACC*$_s$)
- The number of subjects for which the number of segments $J_{i,s}$ matched the ones form the mean annotations $\hat{J}_i$ (Ω_s)
- Duration error $DE_s = |\Delta t_{j,i,s} - \Delta\hat{t}_{j,i}|$ for these Ω_s subjects, with $\Delta t = t_j^{end} - t_j^{start}$ of segment j with respect to corresponding annotated segments ($\hat{}$)

For each sensor s, DE_s was summarized by mean (*MDE*$_s$) and std, considering Ω_s subjects as a single distribution. Mean errors were also computed for each i and the range for all subjects reported.

Sensor-derived clinical metrics (Sec. III-D) were compared to metrics captured by standard clinical practice ($v_{i,clin}$ and $d_{i,clin}$) for each gait assessment. Results are summarized by mean and std of $\Delta v_{i,s} = |v_{i,s} - v_{i,clin}|$ and $\Delta d_{i,s} = |d_{i,s} - d_{i,clin}|$ for all i.

IV. RESULTS

A. Subtask segmentation

In this study, we are deliberately "oversensing" and our first key observation was to define the optimal parameter sets and input channels in order to inform future study designs.

For the segmentation of the 2MWT, we selected optimal methods parameters (window length L_s and step size S_s) and combinations of input data channels $\mathbf{C}_{i,s}^{input}$ per sensor s using a grid search on

- $L_s \in [50, 60, \ldots, 300]$ samples (i.e. $[0.5, 0.6, \ldots, 3]$ s),
- $S_s \in [1, 10, \ldots, L_s]$ samples and
- $\mathbf{C}_{i,s}^{input} \in [\mathbf{C}_{i,s}, \mathbf{a}_{i,s}, \mathbf{g}_{i,s}, \mathbf{r}_{i,s}, \{\mathbf{a}_{i,s}; \mathbf{g}_{i,s}\}, \{\mathbf{g}_{i,s}; \mathbf{r}_{i,s}\}, \{\mathbf{a}_{i,s}; \mathbf{r}_{i,s}\}]$

to simultaneously maximize Ω_s and minimize *MDE*$_s$. Fig. 2 shows the error metrics per sensor for all considered grid points. For all sensors, good results were obtained for window length $L_s \in [150, \ldots, 250]$ samples and small $S_s \in [1, \ldots, 40]$. Using only acceleration data as model input led to poor results, especially when considering Ω_s. The selection of L_s^{opt} and S_s^{opt} parameters and optimal input data is illustrated in Fig. 3 and related results are summarized in Tab. I. The best results were achieved when considering all data channels $\mathbf{C}_{i,s}^{2MWT}$ or angular velocities $\mathbf{g}_{i,s}^{2MWT}$. The worst performance was obtained for results based on only $\mathbf{a}_{i,s}^{2MWT}$, with reduction in Ω_s (*ACC*$_s$) from 66.7 to 88.2% (7.9 to 21.6%) for all sensors. The best performing sensor is the one placed on the waist (shA) with all subjects correctly classified and an overall mean error of the segments' duration $MDE_{shA} = 0.5$ s, while the worst sensor was act, mainly due to this sensor collecting only acceleration data.

Results of 5STS segmentation are summarized in Tab. II, with shA being the sensor with highest accuracy $ACC_{shA} = 98.01\%$ and $MDE_{shA} = 0.13$ s over all 17 subjects.

Additionally, we compared k-means to Gaussian Mixture Models (GMM) clustering. Following the same parame-

TABLE I: Results of 2MWT subtask segmentation for all subjects. The best results are highlighted in bold font.

	Parameters			***ACC***	Ω	***DE* [s]**		
Sensor	L^{opt}	S^{opt}	C^{input}	**[%]**	**[%]**	**Mean**	**Std**	**Range**
act	260	140	**a**	76.20	17.65	1.36	1.15	[1.05, 1.51]
bst	160	10	**g**	91.11	94.12	0.50	0.58	[**0.31**, 0.80]
shA	190	1	**g**	**91.25**	**100**	**0.50**	**0.54**	[0.34, 0.71]
shP	190	10	**C**	91.17	**100**	0.52	0.58	[0.33, **0.68**]
shL	210	40	**g**	89.17	94.12	0.62	0.68	[0.42, 1.00]
shR	190	1	**C**	90.33	88.24	0.60	0.61	[0.40, 0.82]

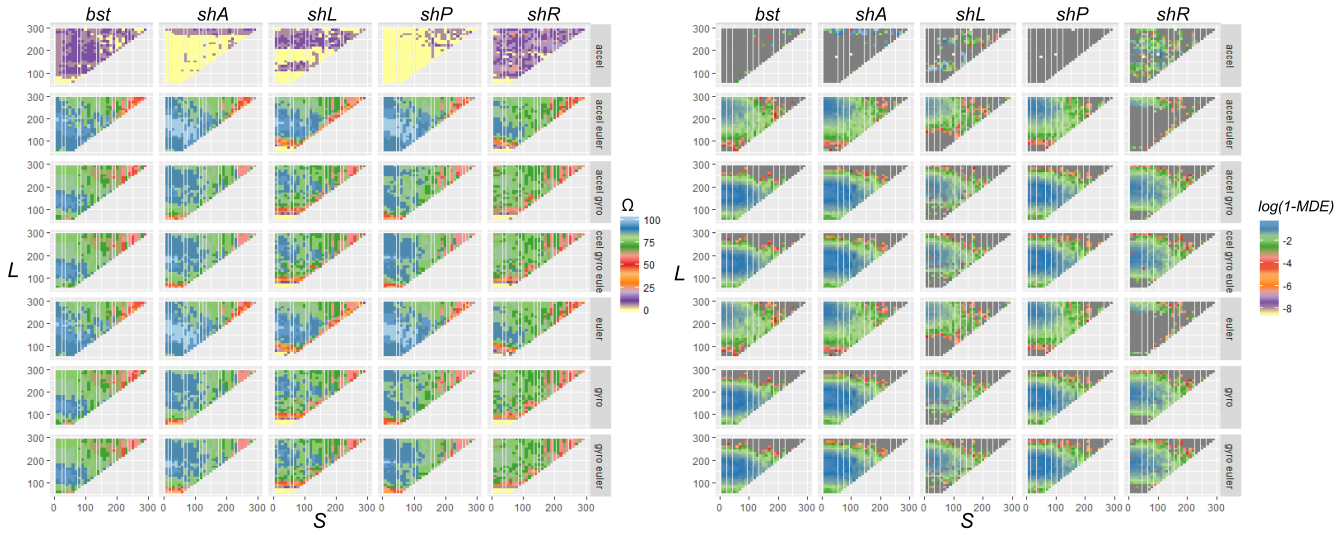

Fig. 2: Results of 2MWT segmentation. Ω_s (left) and MDE_s (right, log-scale) for different values of L_s and S_s and C_s^{input}.

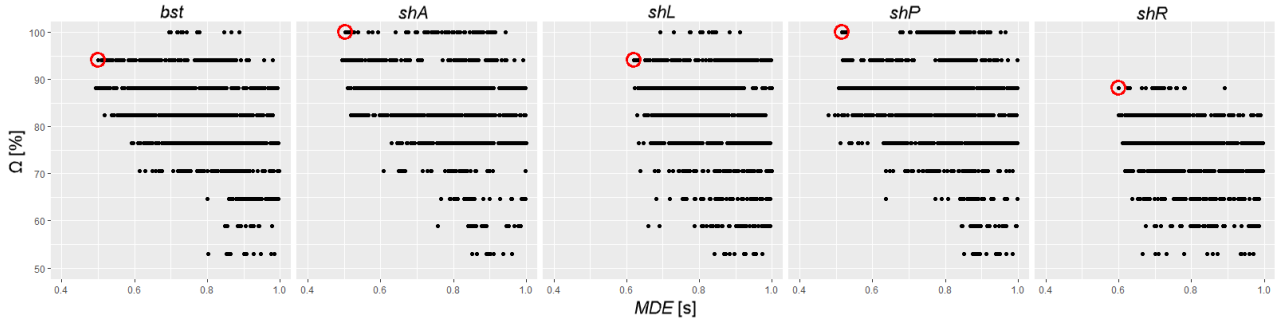

Fig. 3: Parameter search of 2MWT classification via joint maximization of Ω_s and minimization of MDE_s for different values of L_s and S_s and C_s^{input}. The selected optimal points per sensor s are circled in red.

TABLE II: Results of subtask segmentation for 5STS test and all subjects. The best results are highlighted in bold font.

	ACC	Ω	*DE* [s]		
Sensor	**[%]**	**[%]**	**Mean**	**Std**	**Range**
bst	64.01	94.12	0.13	0.16	[0.05, **0.26**]
shA	**98.01**	**100**	**0.13**	0.16	[**0.02**, 0.39]
shP	96.77	82.35	0.13	**0.13**	[0.04, 0.36]

ter search and selection, for the 2MWT, overall ACC_s of GMM was 80.14 to 91.49%, $\Omega_s \in [29.41, 70.59]\%$ and $MDE_s \in [0.53, 1.16]$ s for all sensors. Consistently to the k-means results, the best performing sensor was shA, with $L^{opt} = 150$, $S^{opt} = 60$ and $\mathbf{C}^{input} = \mathbf{C}$. Generally, the performance of GMM was poorer than the one of k-means for Shimmer3 and BioStampRC sensors, especially when considering Ω_s. Yet ACC_{act} and Ω_s of GMM were 5.2% and 66.6% higher, respectively, and MDE_{act} 14.7% lower. Compared to the mean accuracy of k-means, for the 5STS test, the one of GMM was 4.4% and 1.4% higher for bst and shA, respectively, and 4.2% lower for shP. Run-time of GMM clustering was significantly higher.

We tested the classification of TUG data from shA, as shA has shown to have high accuracy in classifying both 2MWT and 5STS data. As a result, we achieved $\Omega_{shA} = 76.47\%$ and balanced overall accuracies ACC_{shA} of 84.17, 66.92 and 72.30% for standing up, walking and turning, respectively.

B. Clinical assessment

Table III shows the summary (mean and std) of clinical metrics errors for gait tests. For long walking (2MWT) act is better than shR, whereas for short tests, v_{shR} comes closer to the clinically evaluated speed. By considering only data from the segmented walking shuttles, the distance errors improved for both 2MWT and TUG.

V. CONCLUSIONS

Incorporating sensors into established clinical motion assessment and hence perform instrumented tests, enables teams to vastly increase the amount of objective motor function performance data collected, with only a small increase in burden to clinicians and patients. We presented an approach to derive quantitative parameters from motor function tests and their subtasks by capturing, segmenting and analyzing IMU data. Compared against expert annotations, our method is device agnostic, automated and unsupervised, and achieved an accuracy of 91% (98%) when segmenting walking vs.

TABLE III: Mean (std) of clinical metrics errors, with shR* based on only segmented walking data.

	SPPB 3m		**TUG**			**2MWT**		
	act	**shR**	**act**	**shR**	**shR***	**act**	**shR**	**shR***
Δv [m/s]	0.39 (**0.10**)	**0.16** (0.11)	0.62 (**0.12**)	**0.19** (0.17)	0.58 (0.26)	**0.06** (**0.05**)	0.13 (0.10)	0.12 (0.11)
Δd [m]	**0.74** (0.34)	1.14 (1.52)	3.02 (0.99)	2.44 (1.36)	**0.83** (**0.63**)	**6.79** (**5.95**)	29.50 (23.97)	16.16 (21.12)

turning (standing vs. sitting) and time errors lower than 0.5 s. Classification accuracy on the TUG test was above 67% for all subtasks. We also extracted clinical metrics from segmented data. Such segmented parameters were in agreement with traditional scores.

In addition to establishing feasibility of these approaches and demonstrating added value for assessing motor function, our results have a number of practical consequences for clinical teams. We recommend to use sensors with gyroscope functionality, worn close to the body center of pressure (waist) when performing chair raise and at the lower extremities or waist for gait tests. In this work, we considered each sensor independently. Future work will investigate on finding an optimal combination of wearable sensors in terms of accuracy of motor function assessment, while still being a practical solution to be deployed into clinics.

Interesting results come from demonstrating the feasibility of deriving tasks from TUG tests by combining segmented subtask models trained on 5STS and 2MWT data. These results could enable removing TUG from batteries of motor tests, reducing site burden, while maintaining captured information. Better combinations of classifiers could improve TUG testing accuracy. We plan on applying this method to assess physical performance to data from patients with gait impairment, such as frail, elderly and post-surgery populations. For these, temporal changes of subtask performance are of clinical interest and lead to a more complete and granular understanding of patient mobility, e.g. assessing balance during the stand-up subtask of 5STS. We believe that experience gained in relatively controlled clinical settings will enable deployment of sensor-based monitoring of motor function in the home environment.

ACKNOWLEDGMENT

We thank our colleagues from the Novartis Institutes for Biomedical Research Jelena Curcic and Holger Hoefling for their help in annotating and processing data; Ronenn Roubenoff and Roland Huber for their help in setting up the study; and Jason Laramie and Scott Kennedy for their support.

REFERENCES

[1] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," *KDD workshop*, vol. 10, no. 16, pp. 359–370, 1994.

[2] M. Bo, G. Fonte, F. Pivaro, M. Bonetto, C. Comi, V. Giorgis, L. Marchese, G. Isaia, G. Maggiani, E. Furno, *et al.*, "Prevalence of and factors associated with prolonged length of stay in older hospitalized medical patients," *Geriatrics & gerontology international*, vol. 16, no. 3, pp. 314–321, 2016.

[3] R. W. Bohannon, "Reference values for the timed up and go test: a descriptive meta-analysis," *Journal of geriatric physical therapy*, vol. 29, no. 2, pp. 64–68, 2006.

[4] P. S. Bradley and U. M. Fayyad, "Refining initial points for k-means clustering," *ICML*, vol. 98, pp. 91–99, 1998.

[5] B. Byrom and D. A. Rowe, "Measuring free-living physical activity in copd patients: deriving methodology standards for clinical trials through a review of research studies," *Contemporary clinical trials*, vol. 47, pp. 172–184, 2016.

[6] CIR systems Inc, "GAITrite," http://www.gaitrite.com.

[7] S. S. Dhillon, C. A. Sima, A. R. Kirkham, N. Syed, and P. G. Camp, "Physical activity measurement accuracy in individuals with chronic lung disease: a systematic review with meta-analysis of method comparison studies," *Archives of physical medicine and rehabilitation*, vol. 96, no. 11, pp. 2079–2088, 2015.

[8] E. P. Doheny, C. Walsh, T. Foran, B. R. Greene, C. W. Fan, C. Cunningham, and R. A. Kenny, "Falls classification using tri-axial accelerometers during the five-times-sit-to-stand test," *Gait & posture*, vol. 38, no. 4, pp. 1021–1025, 2013.

[9] C. Fischer, P. T. Sukumar, and M. Hazas, "Tutorial: Implementing a pedestrian tracker using inertial sensors," *IEEE Pervasive Computing*, vol. 12, no. 2, pp. 17–27, 2013.

[10] A. N. Freire, R. O. Guerra, B. Alvarado, J. M. Guralnik, and M. V. Zunzunegui, "Validity and reliability of the short physical performance battery in two diverse older adult populations in Quebec and Brazil," *Journal of aging and health*, vol. 24, no. 5, pp. 863–878, 2012.

[11] N. Gebruers, C. Vanroy, S. Truijen, S. Engelborghs, and P. P. De Deyn, "Monitoring of physical activity after stroke: a systematic review of accelerometry-based measures," *Archives of physical medicine and rehabilitation*, vol. 91, no. 2, pp. 288–297, 2010.

[12] O. Giggins, I. Clay, and L. Walsh, "Physical activity monitoring in patients with neurological disorders: a review on novel body-worn devices," *Digital Biomarkers*, vol. 1, pp. 14–42, 2017.

[13] C. Godinho, J. Domingos, G. Cunha, A. T. Santos, R. M. Fernandes, D. Abreu, N. Gonçalves, H. Matthews, T. Isaacs, J. Duffen, *et al.*, "A systematic review of the characteristics and validity of monitoring technologies to assess parkinsons disease," *Journal of neuroengineering and rehabilitation*, vol. 13, no. 1, p. 24, 2016.

[14] B. R. Greene, D. McGrath, R. O'Neill, K. J. O'Donovan, A. Burns, and B. Caulfield, "An adaptive gyroscope-based algorithm for temporal gait analysis," *MBEC*, vol. 48, no. 12, pp. 1251–1260, 2010.

[15] 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 Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2918–2926, 2010.

[16] A. Hickey, S. Del Din, L. Rochester, and A. Godfrey, "Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis," *Physiological measurement*, vol. 38, no. 1, 2016.

[17] M. Hirvensalo, T. Rantanen, and E. Heikkinen, "Mobility difficulties and physical activity as predictors of mortality and loss of independence in the community-living older population," *Journal of the American Geriatrics Society*, vol. 48, no. 5, pp. 493–498, 2000.

[18] E. A. Miller and W. G. Weissert, "Predicting elderly peoples risk for nursing home placement, hospitalization, functional impairment, and mortality: a synthesis," *Medical care research and review*, vol. 57, no. 3, pp. 259–297, 2000.

[19] I. Skog, P. Handel, J. O. Nilsson, and J. Rantakokko, "Zero-velocity detection - an algorithm evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2657–2666, 2010.

[20] H. Van Remoortel, S. Giavedoni, Y. Raste, C. Burtin, Z. Louvaris, E. Gimeno-Santos, D. Langer, A. Glendenning, N. S. Hopkinson, I. Vogiatzis, *et al.*, "Validity of activity monitors in health and chronic disease: a systematic review," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 9, no. 1, p. 84, 2012.

[21] Vicon, "Vicon motion capture system," http://www.vicon.com.