<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *Biomedical and Health Informatics 2016, Las Vegas, 24-27 February*.

Citation for the original published paper:

Memedi, M., Aghanavesi, S., Westin, J. (2016)

A method for measuring Parkinson's disease related temporal irregularity in spiral drawings

In: *2016 IEEE International Conference on Biomedical and Health Informatics* (pp. 410-413).

<https://doi.org/10.1109/BHI.2016.7455921>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:du-21183>

A method for measuring Parkinson's disease related temporal irregularity in spiral drawings

Mevludin Memedi, Somayeh Aghanavesi, and Jerker Westin

Abstract—The objective of this paper was to develop and evaluate clinimetric properties of a method for measuring Parkinson's disease (PD)-related temporal irregularities using digital spiral analysis. In total, 108 (98 patients in different stages of PD and 10 healthy elderly subjects) performed repeated spiral drawing tasks in their home environments using a touch screen device. A score was developed for representing the amount of temporal irregularity during spiral drawing tasks, using Approximate Entropy (ApEn) technique. In addition, two previously published spiral scoring methods were adapted and their scores were analyzed. The mean temporal irregularity score differed significantly between healthy elderly subjects and advanced PD patients ($P < 0.005$). The ApEn-based method had a better responsiveness and test-retest reliability when compared to the other two methods. In contrast to the other methods, the mean scores of the ApEn-based method improved significantly during a 3 year clinical study, indicating a possible impact of pathological basal ganglia oscillations in temporal control during spiral drawing tasks. In conclusion, the ApEn-based method could be used for differentiating between patients in different stages of PD and healthy subjects. The responsiveness and test-retest reliability were good for the ApEn-based method indicating that this method is useful for measuring upper limb temporal irregularity at a micro-level.

I. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder associated with motor (e.g. bradykinesia, rigidity, tremor) and non-motor (e.g. constipation, depression, eye movement disorders) symptoms. One of the disabilities associated with PD is the impaired ability to accurately time movements [1]. The research suggests that the basal ganglia (BG) play a significant role in temporal processing at milliseconds to seconds range, also known as "interval timing" [2]. The speed of a hypothetical "internal clock" is controlled by basal ganglia and is related to the brain dopaminergic levels [1]. The deterioration of dopaminergic neurons among PD patients is associated with motor symptoms but also with disruption in repetitive movements. Many studies have shown an increase in timing variability among patients when compared to healthy subjects, suggesting the BG has a role in interval timing. For instance, in one study it was shown that treated PD patients had poorer timing control than

untreated patients when modulating gait timing during externally-cued conditions [3]. Therefore, relating objective measures obtained by instrumented tests to pathological BG fluctuations would be beneficial for facilitating the assessment of high frequency motor irregularities that could be difficult to be assessed visually [4] [5].

In the present study we investigated the upper limb temporal irregularity of patients in different stages of PD and healthy elderly subjects during spiral drawing tasks. The amount of temporal irregularity during spiral drawing tasks was quantified using an Approximate Entropy (ApEn)-based method. In addition to this method, two more spiral scoring methods [6] and [7] were adopted and their scores were compared to the ApEn-based method. Specific objectives of the study are to: i) investigate mean temporal irregularities between patients and healthy elderly subjects, ii) assess the responsiveness of the three methods, iii) assess the test-retest reliability of the three methods, and iv) investigate whether long-term trend of the methods can be used as indicators of a possible involvement of BG in temporal control during spiral drawing tasks.

II. METHODS

A. Subjects

The results presented in this paper are based on data from two clinical studies, both of which were approved by the relevant agencies and informed consent was given. In total, 98 PD patients in different stages of PD and 10 healthy elderly (HE) subjects participated. Of the 98 patients, 65 patients with advanced idiopathic PD were recruited in an open longitudinal 36-month study at nine clinics in Sweden [8]. On inclusion, 35 of these patients were treated with levodopa-carbidopa intestinal gel infusion (LCIG) and 30 patients were candidates for switching from conventional oral PD treatment to LCIG. In the latter group, the patients were LCIG treatment-naïve at study start. In the second study, 38 patients with a clinical diagnosis of idiopathic PD in Milan, Italy participated [9]. The Italian patients were divided into two groups: intermediate stage patients experiencing motor fluctuations ($n = 16$) and clinically stable, early PD patients ($n = 17$). Characteristics of the patients and HE subjects are shown in Table I.

B. Experimental setup

Both patients and HE subjects repeatedly used a touch screen telemetry device in their home environments [10]. Assessments with the device were performed four times per day during week-long test periods. The Swedish LCIG-naïve patients used the device at baseline (before LCIG), month 0 (first visit; at least 3 months after permanent intraduodenal LCIG), and thereafter quarterly for the first year and

*Research supported by Swedish Knowledge Foundation and Vinnova.

Authors are with the School of Technology and Business Studies, Computer Engineering, Dalarna University, Falun 791 88, Sweden (e-mail: mmi@du.se; saa@du.se; jwe@du.se).

biannually for the second and third years. The LCIG-naïve patients used the device from the first visit that is month 0. In 23 LCIG-naïve patients, assessments with the device were available during baseline period and at least one test period after having started LCIG. Hence, $n = 23$ in the LCIG-naïve group. The Italian patients used the device for two week-long test periods with a washout week in between. The HE subjects used the device for one week-long test period.

TABLE I. CHARACTERISTICS OF PD PATIENTS AND OF HEALTHY ELDERLY PARTICIPANTS, PRESENTED AS MEDIAN $\pm$ INTERQUARTILE RANGE.

	Subjects			
	Swedish study (advanced patients)	Italian study (Intermediate patients)	Italian study (Early patients)	HE
Patients (n, gender)	65 (43m; 22f)	17 (13m; 2f)	16 (13m; 2f)	10 (5m; 5f)
Age (years)	65 $\pm$ 11	65 $\pm$ 6	65 $\pm$ 6	61 $\pm$ 7
Years with levodopa	13 $\pm$ 7	7 $\pm$ 8.5	5.5 $\pm$ 6	NA
Hoehn and Yahr stage at present	2.5 $\pm$ 1*	2 $\pm$ 0**	2 $\pm$ 0.5	NA

* Assessments performed in the afternoon. ** Assessments performed in the On state. Abbreviation: HE, healthy elderly; NA, not applicable.

On each test occasion, the subjects were asked to trace a pre-drawn Archimedes spiral that was shown on the screen of the device, using the dominant hand. The subjects were instructed to perform the spiral tests, using an ergonomic pen stylus. In addition, they were instructed to trace the pre-drawn spiral from the center and out, as accurately and fast as possible, supporting neither hand nor arm, with the device placed on a table and to be seated in a chair. The spiral test was repeated three times per test occasion and the subjects were instructed to complete it within 10 seconds. The device had a 3.5" touch screen, 240 X 320 pixel resolution. Position and time-stamps (in ms) of the pen tip were recorded and stored for offline processing.

III. FEATURE EXTRACTION

The digitized spiral signals were processed using the following three methods. A new method based on Approximate Entropy (ApEn) was developed to generate a temporal irregularity score (TIS). In addition, two more methods developed by Liu et al. [6] and Westin et al. [7] were adopted and applied on the spiral data and their corresponding scores were compared to TIS.

A. The method for quantification of temporal irregularity during spiral drawing tasks

Initially, drawing speed (DS) was calculated as a rate of spatial change with respect to time, using the following equation:

$$DS = \frac{\sqrt{(x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2}}{t_{i+1}-t_i} \quad (1)$$

where x is the horizontal coordinate of pixels on the screen, y is the vertical coordinate and t is the time in seconds. Next, the ApEn technique was applied on DS signals to generate the score. The ApEn is a statistical method for measuring the repeatability of patterns within a signal [11]. A signal containing a single frequency component is associated with a relatively small ApEn value whereas more complex signals containing multiple frequency components are associated with high ApEn values, indicating a high level of irregularity. ApEn reflects the similarity between a chosen window of a given duration and the next set of windows of the same duration. ApEn requires determination of two parameters: m (length of the window being compared) and r (measure of similarity), which must remain fixed during all calculations. In the current work and after experimentation, m was set to 4 and r to 0.2 (20% of the standard deviation of the signal). The derived ApEn value was then corrected for total drawing completion time. The resulting score is hence on denoted 'temporal irregularity score' (TIS).

B. Other methods for measuring PD motor symptoms based on digital spiral analysis

Two previously published methods by Liu et al. [6] and Westin et al. [7] were adopted and applied on the spiral data and their scores were used in subsequent analysis. The method developed by Liu et al. [6] was designed to quantify the severity of drug-induced dyskinesias in the upper limbs, using digital spiral analysis. Drawing velocity signals in horizontal, vertical, radial and tangential directions were bandwidth filtered in the range of 1-5 Hz to extract frequency components representing dyskinetic movements. The standard deviation of the frequency transformed signals in each direction was then calculated and the mean standard deviation was used to represent the extent of dyskinesia. Since our device had a lower sampling rate than the one used by Liu et al. [6], their method was modified and its score was compared to TIS. Initially, the radial velocity (RV) was calculated using the following equation:

$$RV = \frac{r_{i+1}-r_i}{t_{i+1}-t_i} \quad (2)$$

where r is the radius defined as the square root of the sum of the squares of x and y coordinates. Next, the RV was interpolated and bandwidth filtered using Chebyshev Type I filter in the frequency range of 1-5 Hz. Finally, the standard deviation of the filtered RV signal was calculated to form a dyskinesia score, denoted SD-DV (standard deviation of drawing velocity).

In the work performed by Westin et al. [7], the digitized spiral data were processed to yield a score for representing spiral drawing impairment, using discrete wavelet transform and principal component analysis (PCA). A 3-level decomposition was performed on radius signal to obtain low- and high-frequency components represented by wavelet coefficients. In order to reduce the dimensionality of these coefficients, the PCA was applied and the first principal component was calibrated and linearly transformed. The resulting score is hence on denoted the 'wavelet spiral test score' (WSTS) and used in subsequent analysis.

C. Statistical analysis

Differences in mean scores of the three methods across the four subject groups were assessed using linear-mixed effects (LME) models with subject ID as a random effect and group as fixed effect of interest. For group (HE vs. advanced PD) classification, three machine learning methods: Logistic Regression (LR), C4.5 decision tree, and random Forests (RF). The machine learning methods were tested with stratified 10-fold cross validation. The relative ability of the methods to detect change from baseline (oral treatment) to month 0 (LCIG) was determined by effect size correlation coefficient, representing the magnitude of treatment effect [cf. e.g. 12]. The score that had the highest coefficient value was defined as the most sensitive to treatment response. The long-term progressions of the three scores were assessed using LME models in combined data from the two groups of Swedish advanced patients starting from the first test period (month 0) with LCIG treatment. Correlations between the three scores were assessed using Pearson correlation coefficients. Test-retest reliability of the three scores was assessed after taking mean of the three possible correlations between the three spiral test trials.

IV. RESULTS

When comparing mean scores between the four subject groups, the mean TIS score was significantly different between HE subjects and advanced patients ($P < 0.005$, Figure 1). The mean TIS score did not differ between HE and the other two groups i.e. early ($P = 0.62$) and intermediate ($P = 0.09$) patients. The mean TIS score was significantly different only between HE and advanced groups ($P < 0.05$). In contrast to TIS and WSTS, there were no significant differences in mean SD-DV scores across the groups.

Fig 1. LME fixed effect coefficients of the three scores across the four subjects groups. Y axes: a high score for TIS and WSTS means good function and for SD-DV means bad. P-values (groups: early, intermediate and advanced) with respect to HE subjects for TIS: 0.62, 0.09, < 0.05 ; WSTS: 0.62, 0.22, < 0.05 ; SD-DV: 0.99, 0.33, 0.24. Number of observations: HE ($n = 877$), Early ($n = 2432$), Intermediate ($n = 2458$), Advanced ($n = 30287$).

WSTS and SD-DV scores were correlated to each other with a coefficient of 0.69 ($P < 0.001$). However, TIS was uncorrelated to both WSTS and SD-DV with coefficients of 0.11 and 0.12, respectively, indicating that TIS measures a different aspect of upper limb motor performance in PD patients than WSTS and SD-DV. Table II shows the

performance of the three classifiers when classifying between HE subjects and advanced PD patients, using first WSTS and TIS individually and then both of them as inputs to the classifiers. The rationale for not including SD-DV in the classification analysis was that WSTS and SD-DV were highly correlated to each other and by adding it did not improve the classification accuracy. The three classifiers performed better when using both WSTS and TIS scores combined. The best performing classifier (LR) correctly classified the group with an accuracy of 85% and weighted AUC of 0.89.

TABLE II. CLASSIFICATION RESULTS OF THE THREE CLASSIFIERS WHEN USING WSTS AND TIS SEPARATELY AND COMBINED. THE RESULTS WERE OBTAINED WITH A STRATIFIED 10-FOLD CROSS VALIDATION.

		Accuracy (%)	Weighted Kappa	AUC
Input: WSTS	C4.5	77	0.54	0.75
	LR	74	0.47	0.83
	RF	77	0.54	0.83
Inputs: TIS	C4.5	71	0.42	0.7
	LR	67	0.34	0.4
	RF	70	0.4	0.75
Inputs: WSTS and TIS	C4.5	84	0.68	0.88
	LR	85	0.71	0.89
	RF	82	0.63	0.89

The effect size correlation coefficients were as follows: TIS (0.078), SD-DV (0.011) and WSTS (0.0009), indicating that TIS has greater responsiveness than the other two scores. In Swedish LCIG-naïve patients, the mean TIS improved to the second test period on LCIG treatment and this improvement remained significant throughout the study period (Figure 2). However, the mean scores of WSTS and SD-DV deteriorated over the study period.

Figure 2. LME fixed effect coefficients of the three scores of Swedish patients over the 36 months study period. All scores for follow-up test periods i.e. from 0 to 36 were significantly different from baseline (-3) test period (each $P < 0.001$) and those periods annotated with symbol * ($P < 0.005$). Abbreviation: n.s. (not significant). Number of observations (test periods, months): 0 ($n = 4467$), 3 ($n = 4043$), 6 ($n = 3781$), 9 ($n = 3391$), 12 ($n = 3028$), 18 ($n = 2828$), 24 ($n = 2666$), 30 ($n = 1833$), 36 ($n = 1696$).

Test-retest reliability coefficients of the three scores were as follows: TIS (0.9), WSTS (0.83) and SD-DV (0.55). These results indicate that the TIS score is more stable and consistent over time than WSTS and SD-DV scores.

V. DISCUSSION AND CONCLUSIONS

In conclusion, TIS could reasonably discriminate well between spiral drawings drawn by patients in different stages of PD and HE subjects, as compared to the scores produced by two previous methods i.e. SD-DV [6] and WSTS [7]. The TIS measure quantifies a different aspect of upper limb motor severity as compared to the previous scores since it was weakly correlated to both of them. In addition, the TIS was weakly correlated to simultaneous tapping speed measurements collected by the touch screen device indicating that the TIS does not measure bradykinesia. The WSTS and SD-DV are related to spatial components of the spiral and thus measure overall drawing impairment or shakiness whereas TIS is related to temporal fluctuations during spiral drawing. This difference was noted in the improvement of classification accuracy of the three classifiers when TIS was used as an input along with WSTS. The TIS measure had a better responsiveness and test-reliability than WSTS and SD-DV. These results could be related to the fact that the ApEn technique provides a high resolution of a signal by partitioning it into smaller set of windows (in our case 4 data points) and sliding the windows throughout the signal. This enables the TIS to measure high-frequency irregularities during spiral drawing in the order of milliseconds. Visual assessment of these irregularities is impossible.

In advanced Swedish patients, the mean TIS improved significantly throughout the study period ($P < 0.001$), except for the second test period i.e. month 3 (Figure 2). This result was surprising to us since TIS was worsening with increased disease severity as seen in Figure 1. If as we assume, the TIS relates to pathological BG oscillations, this makes sense and we had expected a similar profile in Figure 2 as for WSTS where drawing impairment seem to increase over time. What we see is a gradual improvement in TIS over the first two years and then a constant level. One possible explanation for this could be that the pathological BG oscillations correlate to the number of remaining dopamine receptors. This number is related to the number of dopaminergic cells but also the number of receptors on each cell is variable. During levodopa tablet treatment there is a high variation in striatal dopamine levels that causes the number of receptors per cell to decrease in response to high peak levels. This variation in dopamine levels is reduced when starting LCIG pump treatment and over time the number of receptors per cell may begin to increase.

A limitation of the study is that there is no gold standard to validate the TIS. This limitation is related to the data collection scheme where subjects repeatedly used the touch screen device in their home environments without clinical supervision. Therefore, there were no test occasion level clinical ratings to assess the validity of TIS. An interesting research question to be investigated in the future would be to assess clinimetric properties of TIS on a new data set, which consists of sensor and clinical measurements on a test occasion level. The plan is to investigate whether mean TIS is different between PD patients and healthy controls and across different levodopa levels. Since the BG oscillations are affected by deep brain stimulation (DBS) treatment

frequencies [4] another interesting research question that could be investigated in the future would be to calculate the TIS on spiral data of patients on DBS treatment. This would allow investigating the relationship between TIS and different DBS frequencies as well as during Off and On stimulation phases and whether TIS could be useful for adjusting DBS treatment.

REFERENCES

- [1] J. T. Coull, R. K. Cheng, and W. H. Meck, "Neuroanatomical and neurochemical substrates of timing," *Neuropsychopharmacology*, vol. 36, pp. 3-25, 2011.
- [2] C. V. Buhusi, and W. H. Meck, "What makes us tick? Functional and neural mechanisms of interval timing," *Nature Reviews Neuroscience*, vol. 6, pp. 755-765, 2005.
- [3] Q. J. Almeida, J. S. Frank, E. A. Roy, A. E. Patla, and M. A. Jog, "Dopaminergic modulation of timing control and variability in the gait of Parkinson's disease," *Movement Disorders*, vol. 22, pp. 1735-1742, 2007.
- [4] A. Eusebio, C. C. Chen, C. S. Lu, S. T. Lee, C. H. Tsai, P. Limousin, M. Hariz, and P. Brown, "Effects of low-frequency stimulation of the subthalamic nucleus on movement in Parkinson's disease," *Experimental Neurology*, vol. 209, pp. 125-130, 2008.
- [5] A. Raz, V. Frechter-Mazar, A. Feingold, M. Abeles, E. Vaadia, and H. Bergman, "Activity of pallidal and striatal tonically active neurons is correlated in MPTP-treated monkeys but not in normal monkeys," *The Journal of Neuroscience*, vol. 21, 2001.
- [6] X. Liu, C. B. Carroll, S. Y. Wang, J. Zajicek, and P. G. Bain, "Quantifying drug-induced dyskinesias in the arms using digitized spiral-drawing tasks," *Journal of Neuroscience Methods*, vol. 144, pp. 47-52, 2005.
- [7] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, and T. Groth, "A new computer method for assessing drawing impairment in Parkinson's disease," *Journal of Neuroscience Methods*, vol. 190, pp. 143-148, 2010.
- [8] S. R. Pålhagen, N. Dizdar, T. Hauge, B. Holmberg, R. Jansson, J. Linder, D. Nyholm, O. Sydow, M. Wainwright, H. Widner, and A. Johansson, "Interim analysis of long-term intraduodenal levodopa infusion in advanced Parkinson's disease," *Acta Neurologica Scandinavica*, vol. 126, pp. e29-e33, 2012.
- [9] J. Westin, M. Schiavella, M. Memedi, D. Nyholm, M. Dougherty, and A. Antonini, "Validation of a home environment test battery for assessments in advanced Parkinson's disease," *Neurological Sciences*, vol. 33, pp. 831-838, 2012.
- [10] J. Westin, M. Dougherty, D. Nyholm, and T. Groth, "A home environment test battery for status assessment in patients with advanced Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 98, pp. 27-35, 2010.
- [11] S. M. Pincus, "Approximate entropy as a measure of system complexity," in *Proc. Natl. Acad. Sci.*, pp. 2297-2301, 1991.
- [12] C. G. Goetz, G. T. Stebbins, K. A. Chung, R. A. Hauser, J. M. Miyasaki, A. P. Nicholas, W. Poewe, K. Seppi, O. Rascol, M. A. Stacy, J. G. Nutt, C. M. Tanner, A. Urkowitz, J. A. Jaglin, and S. Ge, "Which dyskinesia scale best detects treatment response?," *Movement Disorders*, vol. 28, pp. 341-346, 2013.