A method for automatic, objective and continuous scoring of bradykinesia

O. Martinez-Manzanera, E. Roosma, M. Beudel, R. W. K. Borgemeester, T. van Laar and N. M. Maurits

Department of Neurology

University Medical Center Groningen (UMCG)

University of Groningen

Groningen, the Netherlands

Email: o.e.martinez.manzanera@umcg.nl

***Abstract*— The assessment of bradykinesia is a key element in the diagnosis of Parkinson's disease. It is typically performed using the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS). However, despite its importance, the bradykinesia-related items of this scale show very low inter-rater agreement. Therefore, in this study a method for automatic, objective and continuous scoring of three of the bradykinesia-related items of the MDS-UPDRS is proposed. Four clinicians scored these items for 25 patients diagnosed with Parkinson's disease, within a range of 0-4. Orientation sensors were used to record movement during performance of each item. From the recorded data a set of features was derived to represent the movement characteristics that evaluators assess for scoring bradykinesia according to the MDS-UPDRS. These features and the averaged scores of the evaluators were used to create a model for the score on each item using backward linear regression. The estimated generalization errors indicate that the continuous objective scale can obtain an automatic score with an average error of 0.50 compared to the evaluators' averaged scores.**

Keywords—Parkinson's disease; bradykinesia; automatic scoring; linear regression; UPDRS;

I. INTRODUCTION

Bradykinesia is defined as slowness of movement execution. It is present in different movement disorders and its correct assessment is essential for the diagnosis of Parkinson's disease (PD). At present, the gold standard for assessing the severity of bradykinesia is the assessment by a well-trained evaluator using standard clinical rating scales [1]. The most widely used scale is the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [2]. In the third part of this scale, motor ability is evaluated employing different tasks. Each task is assessed according to specific movement characteristics as indicated in the guidelines. Despite its ubiquitous use, the bradykinesia-related items of the MDS-UPDRS have low inter-rater agreement between evaluators [3].

Current assessment of bradykinesia suffers from two important inherent inconveniences: the evaluators' individual bias and the limited number of categories of the scale (0-4 for each item). The individual bias is due to the subjective appreciation of task execution and the subjective interpretation of the guidelines. The limited number of categories induces that if task performance has characteristics of two consecutive categories (e.g. 2 and 3), it might be scored in either of these categories by different evaluators.

In a separate study [4] we focused on the individual bias problem. We proposed an automatic and objective method for assessment of the bradykinesia-related items of the MDS-UPDRS that used a supervised classification algorithm (support vector machine (SVM) based) that reproduced the evaluators' classification results.

In the present study, we propose an approach that overcomes the problem associated with the limited number of categories of the MDS-UPDRS. We employed data recorded using orientation sensors during execution of the bradykinesia-related items of the MDS-UPDRS from the same 25 patients with PD and the scorings of the same four evaluators as in our other, separate study. A set of features defined there [4] [Appendix 1] and the averaged scores of the evaluators were used to define a continuous model for performance on each bradykinesia-related item of the MDS-UPDRS. A continuous scale could be used to quantify small variations in movement not detected by evaluators and/or that cannot be represented in the discrete MDS-UPDRS. In this manner an automatic and objective method to assess bradykinesia on a continuous scale could help in the early detection and monitoring of PD.

II. METHODS

Twenty-five patients diagnosed with PD participated in the study. They performed three of the bradykinesia-related items of the MDS-UPDRS (items 3.4 (finger tapping), 3.6 (diadochokinesis) and 3.7 (toe tapping)) and their movements were videoed and subsequently analyzed by four well-trained evaluators. Each evaluator scored these items according to the MDS-UPDRS. Both left- and right-sided movements were performed and analyzed.

Data were recorded using a nine degrees of freedom inertial measurement unit (IMU; Shimmer2r, Shimmer, Dublin, Ireland [5]). For finger tapping, the IMU was placed on the dorsal side of the proximal phalange of the index finger, for diadochokinesis on the dorsal side of the forearm and for toe

978-1-4673-7201-5/15/$31.00 ©2015 IEEE

tapping on the instep of the foot over the corresponding shoe of the participant. A sensor fusion algorithm [6] was used to combine the signals from accelerometers, gyroscopes and magnetic sensors to obtain a movement orientation representation in form of Euler angles. A set of features representative of the movement characteristics described in the MDS-UPDRS (Table 1) was extracted from the Euler angle signal that best described each movement [Appendix 1].

Intra-class correlation (ICC) coefficients were used to quantify agreement between evaluators. Two-way random effects ICC (case 2) coefficients were calculated [7] and the Landis [8] criteria (Table 2) were used to interpret ICC coefficients. We also determined the number of tasks that were classified differently across all pairs of evaluators with a difference in score of only one point of the scale. A large number of these one point differences can be an indication that (regardless of the individual bias) the task execution has characteristics of two consecutive scoring classes as described in the MDS-UPDRS.

We observed a large number of one point differences between evaluators (see Results). A continuous scale (similar to Giuberti et al. [9]), in which the scores of the four evaluators for each item are averaged, can deal with these differences as it results in intermediate values for movements that possibly had characteristics of two consecutive scoring classes.

TABLE I. FEATURES EXTRACTED FROM EULER ANGLE SIGNALS

Number	*Feature*
1	Slope amplitude RA
2	Mean amplitude RA
3	Standard deviation amplitude RA
4	Slope frequency RA
5	Mean frequency RA
6	Standard deviation frequency SSA
7	Slope amplitude SSA
8	Mean amplitude SSA
9	Standard deviation amplitude SSA
10	Slope frequency SSA
11	Mean frequency SSA
12	Standard deviation frequency SSA
13	Filtered signal fit (SSE)
14	Filtered signal fit (R2)
15	Filtered signal fit (RMSE)
16	Percentage of Hesitations
17	CV of zero crossings
18	Mean maxV during movement initiation
19	CV maxV during movement initiation
20	Mean maxV during movement termination
21	CV maxV during movement termination

List of abbreviations. RA: raw angle, SA: smoothing spline angle, SSE: sum of squared errors, RMSE: root mean squared errors, R2: R-squared goodness of fit, CV: coefficient of variation, maxV: maximum velocity. For features description see Appendix 1.

TABLE II. INTERPRETATION OF INTRA-CLASS CORRELATION COEFFICIENTS (ICC) FOR INTER-RATER AGREEMENT

Coefficient	*Agreement*
ICC > 0.8	Almost perfect
0.8 > ICC > 0.61	Substantial
0.60 > ICC > 0.41	Moderate
0.40 > ICC > 0.21	Fair
0.20 > ICC	Slight

Backward linear regression was applied to obtain a model for performance on each bradykinesia-related item using the features obtained from the Euler angle signals as independent variables and the averaged scores from the four evaluators as dependent variables. Backward linear regression was chosen over other regression approaches because it takes into account suppressor effects, which occur when a predictor has a significant effect but only when another variable is held constant [10]. By averaging the scores of the evaluators we decreased the effect of possible outliers (tasks misclassified) on model estimation.

To estimate the model's performance on new data, leave-one-out cross-validation (LOOCV) was employed. In LOOCV all samples but one are used to build the model and the one remaining sample is used for evaluation. The test error was defined as the difference of the predicted value of the remaining sample from its true value (the averaged scores across evaluators). This process was repeated for each sample and the root mean squared error (RMSE) of all samples was defined as the model performance.

III. RESULTS

Moderate agreement between evaluators was found for finger tapping (ICC = 0.58) and diadochokinesis (ICC = 0.60) while evaluators only had fair agreement for toe tapping (ICC = 0.37). For finger tapping, 52% of item performances were given the same score across all pairs of evaluators and 40% of the item performances were scored with one point difference. The average difference in scores across all pairs of evaluators was 0.57. For diadochokinesis, 55% of the item performances were given the same score across all pairs of evaluators and 38% of the item performances were scored with one point difference. The average difference in scores across all pairs of evaluators was 0.53. For toe tapping, 44% of the item performances were given the same score across all pairs of evaluators and for 44% of the item performances the score differed in one point of the scale. The average difference in scores across all pairs of evaluators was 0.70.

A. Linear Regression

For finger tapping (Table 3), the features selected for the model accounted for 79% of the total variance in the data. The large coefficient for feature 8 ($\beta = -0.70$, $p < 0.001$), which is related to movement amplitude, indicates the importance of amplitude for scoring finger tapping. The cross-validation error was 0.38 (Fig. 1, top), 0.19 less than the average difference in scores across all pairs of evaluators.

978-1-4673-7201-5/15/$31.00 ©2015 IEEE

TABLE III. FEATURES SELECTED FOR THE MODEL FOR EACH ITEM PERFORMANCE, EXPLAINED VARIANCE OF THE MODEL (R^2), COEFFICIENTS, STANDARD ERRORS AND SIGNIFICANCE VALUES

Finger Tapping model (R^2 = 0.79)	*Beta*	*Std. Error*	*Significance*
Constant	.880	.050	p<.001
Mean amplitude RA	-.367	.069	p<.001
Standard deviation amplitude SSA	.145	.081	p=0.79
Mean frequency SSA	-.709	.081	p<.001
Std frequency SSA	.332	.081	p<.001
SSE	-.232	.113	p=.046
R2	-.446	.105	p<.001
RMSE	-.361	.084	p<.001
Diadochokinesis model (R^2 = 0.73)	*Beta*	*Std. Error*	*Significance*
Constant	.700	.064	p<.001
Standard deviation amplitude RA	-.154	.078	p=0.056
Standard deviation amplitude SSA	.252	.118	p=0.038
Slope frequency SSA	-.212	.072	p=0.005
Std frequency SSA	.680	.118	p<.001
R2	.357	.118	p=0.004
RMSE	-.515	.129	p<.001
Percentage of Hesitations	.322	.070	p<.001
CV maxV during movement initiation	-.736	.260	p=0.007
CV maxV during movement termination	.741	.269	p=0.009
Toe Tapping model (R^2 = 0.56)	*Beta*	*Std. Error*	*Significance*
Constant	0.875	0.064	p<.001
Slope amplitude RA	0.151	0.067	p=0.029
Mean amplitude RA	-0.973	0.321	p=0.004
Slope amplitude SSA	-0.199	0.07	p=0.007
Mean amplitude SSA	0.704	0.316	p=0.031
R2	0.217	0.091	p=0.021
Percentage of Hesitations	0.2	0.078	p=0.013

For abbreviations, see Table 1 and for feature descriptions, see Appendix 1.

For diadochokinesis (Table 3), the features selected for the model accounted for 73% of the total variance in the data. In this case the largest coefficients were obtained for features 12, 19 and 21 (β = -0.68, $p < 0.001$, β = -0.74, $p = 0.007$ and β = -0.74, $p = 0.009$, respectively) which are related to the variation in movement velocity. The cross-validation error was 0.56 (Fig. 1, center), 0.03 more than the average difference in scores across all pairs of evaluators.

For toe tapping (Table 3), the predictors selected for the model accounted for only 56% of the total variance in the data. Features 2 and 8, related with movement amplitude, had the largest coefficients (β = -0.97, $p = 0.004$ and β = 0.70, p=0.031 respectively). The cross-validation error was 0.48 (Fig. 1, bottom), 0.22 less than the average difference in scores across all pairs of evaluators.

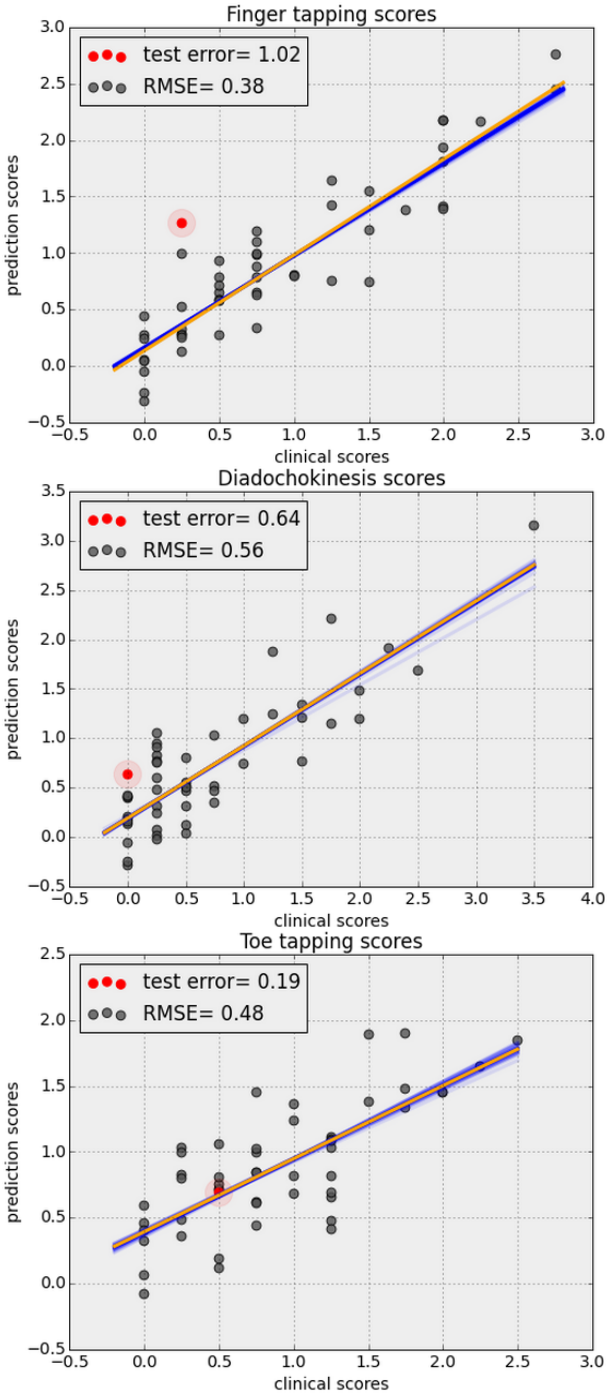

Fig. 1. Illustration of the LOOCV procedure. For each of the bradykinesia-related items finger tapping (top), diadochokinesis (center) and toe tapping (bottom) an example is given of the test error for a specific sample (in red), and the RMSE across all samples. The orange line is the regression model when the specific sample is left out. The pale blue lines are the regression models created on each iteration of the LOOCV sequence. The clinical scores are the averaged scores across all evaluators, the prediction scores are the scores given by the model.

978-1-4673-7201-5/15/$31.00 ©2015 IEEE

IV. Discussion

In this study we employed backward linear regression to create a continuous model for three bradykinesia-related items of the MDS-UPDRS. Using cross-validation we estimated that we can score a new task performance with an average error of 0.38 for finger tapping compared to the averaged scores of four clinicians. The estimated errors for diadochokinesis (0.56) and toe tapping (0.48) are similar. The still limited number of clinical evaluators may induce bias in the obtained model because of their subjective evaluation. A study with a larger number of evaluators could be helpful to derive an unbiased model. Also, the inclusion of a larger number of participants would allow the utilization of a less optimistic cross-validation technique than LOOCV, such as k-fold cross-validation. Reproducing our results in a study with more evaluators and participants is even more important because there are, to our knowledge, no comparable studies in literature.

V. Conclusion

The introduction of objective measures into the clinical diagnosis could abandon current issues with inter-rater agreement. The relatively low errors obtained using cross-validation in this study indicate that the automatic assessment could be useful in the clinic. An automatic objective assessment of bradykinesia using a continuous scale would allow the identification of small variations in performance and can therefore improve the diagnosis and monitoring of PD.

References

[1] N. L. Keijsers, M. W. Horstink, and S. C. Gielen, "Online monitoring of dyskinesia in patients with Parkinson's disease.," IEEE Eng. Med. Biol. Mag., vol. 22, no. 3, pp. 96–103, 2003.

[2] C. G. Goetz et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results.," Mov. Disord., vol. 23, no. 15, pp. 2129–70, Nov. 2008.

[3] D. a Heldman, et al., "The modified bradykinesia rating scale for Parkinson's disease: reliability and comparison with kinematic measures.," Mov. Disord., vol. 26, no. 10, pp. 1859–63, Aug. 2011.

[4] O. Martinez-Manzanera, et al. "A method for automatic and objective scoring of bradykinesia using orientation sensors and classification algorithms," Unpublished.

[5] S. O. H. Madgwick, A. J. L. Harrison, and A. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm.," IEEE Int. Conf. Rehabil. Robot., vol. 2011, p. 5975346, Jan. 2011.

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

[7] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data.," Biometrics, vol. 33, no. 1, pp. 159–74, Mar. 1977.

[8] M. Giuberti et al., "Linking UPDRS Scores and Kinematic Variables in the Leg Agility Task of Parkinsonians," 2014 11th Int. Conf. Wearable Implant. Body Sens. Networks, pp. 115–120, Jun. 2014.

[9] A. Field, Discovering Statistics Using SPSS, Third. Thousand Oaks, California: SAGE Publications Inc., 2009.

[10] K. Shima et al. "Measurement and evaluation of finger tapping movements using magnetic sensors." Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 2008, pp. 5628–31, Jan. 2008.

978-1-4673-7201-5/15/$31.00 ©2015 IEEE

APPENDIX I

The following section describes how the features employed in the regression model are derived from the Euler signal for each task.

All signals were recorded at a sampling rate of 51.2 Hz. For each bradykinesia-related item, most of the movement can be described by a single Euler angle: for finger tapping by the angle that describes the flexion and extension of the index finger, for diadochokinesis by the angle that describes the pronation and supination of the wrist and for toe tapping by the angle that describes the dorsiflexion and plantar flexion of the foot.

Signals were filtered using a band-pass filter (Butterworth, second order, cutoff frequencies 0.3 and 20 Hz). Then, to obtain a smoother version of the signals for feature extraction, spline interpolation was used to fit each signal (Fig. 1a), using a smoothing parameter p = 0.1 (Eq. 1). Two versions of each signal were thus obtained: one with more detail (raw angle (RA) signal) and one with less detail (smoothing spline angle (SSA) signal). Features were subsequently extracted from these two signals. The function that was minimized to obtain the smoothing spline is given in (1):

$$p\sum w_i(y_i+s(x_i))^2 + (1-p)\int(d^2s/dx^2)^2 \qquad (1)$$

Here, p is the smoothing parameter and the first term is the mean squared error (MSE) when the curve s, which is a function of x, is used to predict y. The second term is an added penalty function that limits the curvature of s.

To obtain features related to the characteristics defined in the MDS-UPDRS we identified each movement repetition in the signal by first distinguishing the peaks and valleys in the SSA signal. Then, we defined the amplitude of a single movement as the difference in amplitude from a peak to the next valley and the frequency of each movement as the inverse of the time between consecutive peaks (Fig. 1a). To represent amplitude and speed as mentioned in the MDS-UPDRS, the mean amplitude and mean frequency across all identified movements were calculated. Due to its smoothness, the pattern descripted by the SSA signal resembles more the oscillations pattern expected from the type of movements described in the MDS-UPDRS than the RA signal. However, the low-pass filter effect of the spline interpolation reduced the amplitude of each individual movement repetition in the SSA signal. We therefore decided to calculate features for both the RA and SSA signals.

Other characteristics that are evaluated according to the MDS-UPDRS are decrement of movement amplitude, and slowing of movement. To capture decrement of movement, the slope of the straight line fitted through all movement amplitudes as a function of movement repetition number was taken resulting in the feature slope amplitude. To capture decrement of frequency, a similar procedure was performed for the movement frequencies resulting in the feature slope frequency. These procedures were performed for both the RA and SSA signals.

Another characteristic that is evaluated in the MDS-UPDRS is the regularity of movement (described as rhythm of movement). To represent this characteristic, we estimated amplitude and speed variability by calculating the standard deviations (std) of all movement repetitions amplitudes and frequencies, respectively. This resulted in the features std amplitude and std frequency.

Compared to their corresponding RA signals, SSA signals show a much smoother and regular pattern. The goodness of fit of SSA signals when compared to their corresponding RA signals can give an indication of the regularity of movement. The discrepancy between these two signals is summarized in the following additional features: Sum of Squares due to Error (SSE) which is the sum of the squared deviation of the SSA signal from the RA signal, coefficient of determination (R2) and Root Mean Squared Error (RMSE).

We additionally included features describing regularity of movement. First, to estimate the velocity of each movement, the first derivative of the SSA signal was calculated. According to Shima et al. [10], we determined the maximum velocity during initiation of each movement (extension for finger tapping, dorsiflexion for foot tapping and supination for diadochokinesis) and during termination of each movement (flexion for finger tapping, plantar flexion for foot tapping and pronation for diadochokinesis) and used its mean and coefficient of variation (CV) as well as the mean and CV of the maximum velocity (mean and CV maxV during movement initiation and mean and CV maxV during movement termination).

Finally, hesitations were quantified according to Shima et al, employing zero crossings in the acceleration signal. The acceleration signal was calculated as the second derivative of the SSA signal. An individual movement was considered to contain hesitations if its corresponding acceleration signal contained more than two zero crossings. The percentage of individual movements containing hesitations (percentage of hesitations) and the CV of the number of zero crossings of each individual movement (CV of zero crossings) were determined as features related to hesitations.

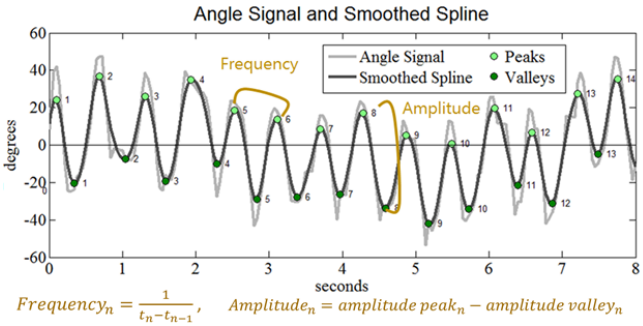

Figure 1a. Example of raw angle (RA: gray) and smoothing spline angle (SSA: black) signal for diadochokinesis. The frequency of each tap was obtained as the inverse of the time (t) between consecutive peaks. In the figure the frequency of tap 6 was defined as the inverse of the time difference between the sixth and the fifth peaks. The amplitude of each tap was obtained as the difference in amplitude from a valley to the next peak. In the figure the amplitude of tap 8 was defined as the amplitude difference between the eighth peak and the eighth valley.

978-1-4673-7201-5/15/$31.00 ©2015 IEEE