Implicit Detection of Motor Impairment in Parkinson's Disease from Everyday Smartphone Interactions

Jing Gao[1, 4]
gaojing162@mails.ucas.ac.cn

Jin Huang[1, 4]
huangjin2014@iscas.ac.cn

Feng Tian[1]
tianfeng@iscas.ac.cn

Hongan Wang[1]
hongan@iscas.ac.cn

Junjun Fan[1, 4]
junjun2015@iscas.ac.cn

Dakuo Wang[3]
Dakuo.Wang@ibm.com

Xiangmin Fan[1]
xiangmin@iscas.ac.cn

Yicheng Zhu[2]
zhuyc@pumch.cn

Shuai Ma[1, 4]
mashuai171@mails.ucas.ac.cn

[1]Institute of Software, Chinese Academy of Sciences, Beijing, China

[2]Peking Union Medical College Hospital, Beijing, China

[3]IBM T.J.Watson Research Center, Yorktown Heights, NY, USA

[4]University of Chinese Academy of Sciences, Beijing, China

Abstract
In this work, we explored the feasibility and accuracy of detecting motor impairment in Parkinson's disease (PD) via implicitly sensing and analyzing users' everyday interactions with their smartphones. Through a 42 subjects study, our approach achieved an overall accuracy of 88.1% (90.0%/86.4% sensitivity/specificity) in discriminating PD subjects from age-matched healthy controls. The performance was comparable to the alternating finger tapping (AFT) test, a well-established PD motor test in clinical settings. We believe that the implicit and transparent nature of our approach can enable and inspire rich design opportunities of ubiquitous, objective, and convenient systems for PD diagnosis as well as post-diagnosis monitoring.

Author Keywords
Parkinson's disease, smartphone, passive monitoring, finger tapping

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Parkinson's disease (PD) is one of the most common neurological diseases (8-18 per 100,000 persons [7]), and the rate is likely to increase as the elderly population

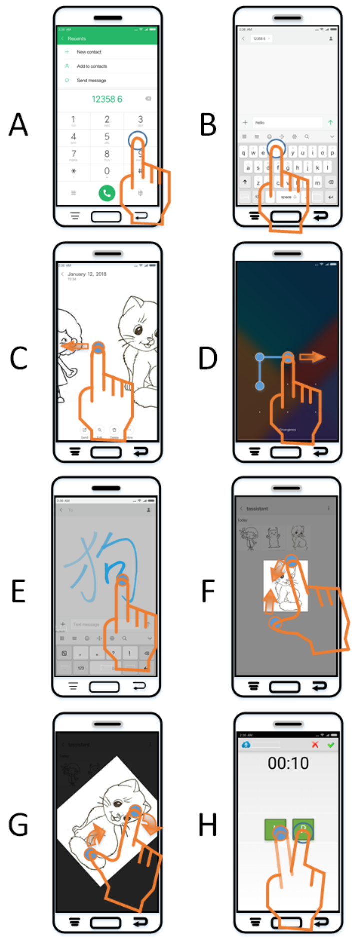

Figure 1. Common smartphone interactions explored in this study. **A**: dial numbers; **B**: type text messages; **C**: swipe screen; **D**: draw graphical password; **E**: handwrite characters; **F**: zoom in/out pictures; **G**: rotate pictures; **H**: alternating finger tapping (clinical reference).

becomes larger in the near future. PD severely affects patients' well-being and quality of life, due to its motor impairment symptoms (e.g., tremor, bradykinesia, postural instability, gait difficulty, and rigidity), and non-motor signs (e.g., cognitive alteration or sleep disturbances [3]). The most commonly used metric in traditional clinical evaluation is the Unified Parkinson's Disease Rating Scale (UPDRS), which consists with a metric of standardized test in each of the aforementioned motor- and non-motor dimensions. However, it suffers from several major constraints. The clinical practitioners need to have significant training to conduct the assessment, yet cannot guarantee the minimum inter-rater variability in interpretation. Meanwhile, when a subject experiences evident symptom and is diagnosed with PD through UPDRS, it is too late for her to receive early interventions at the mild stage of the disease.

Researchers as well as clinical practitioners shared a keen interest in the exploration of technology for objective and low-cost PD assessment. A variety of techniques have been proposed to measure a broad range of PD signs. The most common approach is using accelerometers to develop fine-grained motor tracking tools [6, 13, 12]. However, the use of specialized sensors in an obtrusive way can limit the wide adoption of such approaches. At the same time, researchers have also explored using commodity device such as smartphones for PD evaluation. One of the most well-known studies is mPower [2], which collected longitudinal data from a large number of PD patients and controls when they performed certain tasks including memory tests, voice tests, finger tapping tests, and gait tests. A common limitation of prior work is that they require subjects' *explicit* and *active* participation while it is difficult to sustain users' motivation, especially when the symptoms are not evident. Moreover, the nature of explicit test may impact subjects' physical behaviors which may further introduce the risk of less reliable results.

In this work, we explore detecting PD via sensing and analyzing users' everyday interactions with smartphones (e.g., typing a text message, swiping a photo, etc.). The design rationale behind this approach is that the PD motor symptoms such as bradykinesia, tremor, and rigidity could affect hand and finger performance while interacting with smartphones, which could be detected by the touchscreen and other embedded sensors such as the accelerometer and the pressure sensor. We hypothesize that we can build a machine learning model that can distinguish PD patients from the healthy subjects *implicitly*. We use the term "*implicit*" to differentiate our envisioned scenarios with those that require users to mount sensing equipment, "*explicitly*" launch monitoring apps, and spend an uninterrupted amount of time in data collection. In comparison, our approach can evaluate and monitor PD as a *side effect* during everyday tasks (Figure 1).

We provide empirical evidence to demonstrate the feasibility and accuracy of detecting PD motor impairment from daily smartphone interactions. We also evaluated the accuracies of using various subset of sensor signals in the machine learning models. Through a 42-subject study (20 PD patients and 22 age-and-gender-matched controls), our approach achieved an overall classification accuracy of 88.1% (with 90.0%/86.4% sensitivity/specificity), which was comparable to the performance of common motor tests (e.g., alternating finger tapping test, aka AFT) in clinical settings.

Related Work

Researchers have explored using novel sensor technologies in PD clinical diagnosis, post-diagnosis

	PD	Controls	Significance
Total #	20	22	n/a
Female #	8	9	n/a
Ave Age (yrs)	66.7	65.7	p=0.37
Ave Education Length (yrs)	8.9	8.8	p=0.15

Table 1. Demographic information of the study participants.

Figure 2. Sample pictures of participants in this study.

monitoring, and supporting patients during recent years [5, 14, 13, 8, 9, 1]. One of the common themes is to leverage these technologies to compliment the current clinical diagnosis practice [5, 14, 12]. For example, [5] and [14] are among the earliest works of designing portable sensing equipment to quantify PD patients' response time and abnormal muscle movements in dedicated assessment tasks. Other researchers aimed to utilize widely accessible sensors and commodity devices to detect PD symptoms [1] or support PD patients (e.g., [8, 9]). The most related work with our work is [1], in which the authors proposed an algorithm and a set of useful features to detect PD signs via analyzing the tapping behaviors on smartphones. In comparison, our work provides wider coverages in both sensing channels (e.g., touch pressure, accelerometer) and interaction tasks (e.g., one-finger and two-finger manipulation gestures), which can help the community gain more fundamental and comprehensive understanding of both strengths and limitations of such approach.

In addition to PD diagnosis, researchers have also explored using sensor technologies to empower the PD patients to frequently monitor the natural progression of the disease at home during the gaps between clinical visits [6, 4, 11, 10]. For example, [6] uses 6 sets of accelerometers attached to a patient's body (e.g., arms, chest, and legs) to monitor patient's daily activities at home (e.g., washing hands). [4] uses non-intrusive and inexpensive Kinect sensor to detect and analyze full body gait in PD patients' homes. However, one common limitation in all these technology designs is that the user still needs to allocate specific time and effort to accomplish certain assessment task, though some technology designs (e.g., smartphone, Kinect [4], Fitbit [11], and Google Glass [10]) are less intrusive than the others.

Experiment

We collected smartphone interaction data from 20 subjects diagnosed with PD and 22 age-matched healthy controls. PD was diagnosed via clinical evaluations including the UPDRS-Ⅲ test and magnetic resonance imaging (MRI) test conducted by at least two expert physicians. The experimental procedures were approved by the Institutional Review Board of Peking Union Medical College Hospital. Subjects gave informed consent before the data collection process. Table 1 summarizes the demographic information of the subjects in this study. According to the two-sided Mann-Whitney U tests, PD subjects and controls are statistically similar in age, gender and education length. All PD subjects were tested during the "ON" stage under best medical treatment. Patients with cognitive impairment were excluded for this study.

We used a HUAWEI P9 Plus smartphone running Android 7.0 for data collection. It has a 5.5 inch, 1080*1920 pixels display. It also has a pressure sensor on top of the touchscreen.

Participants first performed the Alternating-Finger-Tapping (AFT) test with the smartphone. They need to alternately tap two rectangles of 12 by 12 millimeters, separated by 6 millimeters, by using their index finger and middle finger on the dominant hand for 10 seconds. They repeated this test for 3 times. The average number of taps was recorded. We included this well-established motor test as an external reference to quantify upper limbs dexterity.

Gesture Type	Interaction Task	Gesture Features	Pressure Features	Accelerometer Features
Static Gestures	Dialing phone numbers	mean and standard deviation of inter-tap dwelling time, time between press and release, screen distance between press and release	mean, standard deviation, and derivative of screen pressure	mean, standard deviation, and derivative of the output from x/y/z axis
	Typing text messages			
One-Finger Manipulation Gestures	Swiping left or right	mean, standard deviation, and derivative of finger movement speed		
	Drawing graphical passwords			
	Handwriting Chinese characters			
Two-Finger Manipulation Gestures	Zooming in/out pictures	mean, standard deviation, and derivative of finger movement speed (for each of the two fingers)		
	Rotating pictures			
Sensor-Enhanced Alternating Finger Tapping (AFT)		mean and standard deviation of inter-tap dwelling time, time between press and release, screen distance between press and release		

Table 2. Tasks and the corresponding features explored in the study.

We collected the interaction activity data when participants performed a variety of touch gestures with a smartphone (Figure 1), including static gestures (e.g., tap), one-finger manipulation gestures (e.g., slide, drag, handwriting gesture), and two-finger manipulation gestures (e.g., pinch, spread, rotate). In order to reflect the actual routine use of smartphone, they were informed to perform these gestures in daily usage scenarios, including dialing phone numbers, typing text messages, swiping the screen left or right, drawing graphical passwords, handwriting Chinese characters, zooming in or out pictures, and rotating pictures. The interaction activity data, including screen pixel positions, screen pressure values, accelerometer outputs, and their corresponding timestamps, were collected. We divided the data collection into five rounds, each round consisted of two repetitions of each scenario in a randomized order. Participants could pause and had a break any time during the study. On average, it took 1.5 hours for each participant to complete the study.

Table 2 shows the tasks, sensors, and features calculated from the outputs of sensors.

Preliminary Results

We applied leave-one (subject)-out-cross-validation (LOOCV) and explored six classification methods (i.e. KNN, SVM, Decision Tree, Random Forest, Naïve Bayes, and AdaBoost) by using the Weka toolkit. For conciseness, we only report the results of AdaBoost (using DecisionStump as base classifier, 10 iterations) in the rest of the paper since it achieved better performance than the other methods.

The overall classification accuracy is 88.1% (with 90.0%/86.4% sensitivity/specificity) using the common smartphone interaction data. The Mann-Whitney U test shows that the accuracy is significantly higher than the clinical references ($p<0.05$), including both the traditional AFT test merely relying on tap counting (accuracy: 71.4% and 66.7%/76.2% sensitivity/specificity, $p<0.05$) and sensor-enhanced-AFT test (accuracy: 83.3% and 85.0%/81.8% sensitivity/specificity, $p<0.05$).

We also conducted a comparison across all the tasks explored in the study. To make the comparison fair, we split the interaction data into 10-second length windows and report the average classification accuracies. Such duration is widely adopted in clinical AFT tests. Figure 3 shows the classification accuracies corresponding to different interaction tasks. Drawing graphical password (**D**) and handwriting Chinese characters (**E**) result in highest average accuracy (85.7%), followed by sensor-enhanced-AFT test (**H’**, 80.9%). Tap counting (**H**) results in the lowest average accuracy (64.3%). Major findings include:

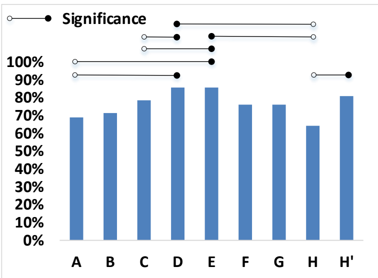

Figure 3. Classification accuracies corresponding to different interaction tasks. **A**: Dialing; **B**: Typing; **C**: Swiping; **D**: Drawing graphical password; **E**: Handwriting characters; **F**: Zooming in/out pictures; **G**: Rotating pictures; **H**: Traditional AFT (i.e. tap counting); **H'**: Sensor-enhanced AFT

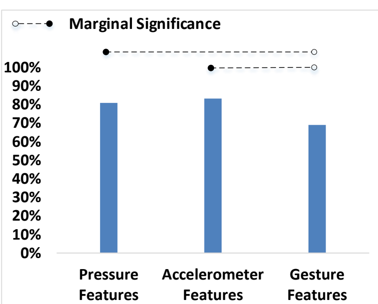

Figure 4. Classification accuracies of using different feature sets.

Finding 1: *Detection accuracy based on sensor-enhanced AFT (**H'**) is significantly higher than tap counting (**H**).*

Finding 2: *Detection accuracies based on drawing graphical password (**D**) and handwriting Chinese characters (**E**) are significant higher than tap counting (**H**).*

Finding 3: *There is no significant difference between detection accuracies based on everyday interaction tasks and the sensor-enhanced AFT (**H'**).*

We attribute the significant improvement achieved by sensor-enhanced AFT (**H'**), drawing graphical password (**D**), and handwriting (**E**), compared with traditional AFT (**H**), to the fact that the richer features can better capture the motor abnormalities that are a direct representation of PD signs. Overall, these findings suggest that analyzing the common smartphone interactions can achieve comparable detection accuracies with widely adopted clinical motor test (i.e. AFT), which shed light on the feasibility of implicit detection of PD impairment from everyday smartphone interactions.

To gain further insights on the relative importance of different types of features, we compared the classification accuracies of using gesture features, pressure features, and accelerometer features. Figure 4 shows the results. We have the following finding from this comparison.

Finding 4: *The detection accuracies of using pressure features (81.0%) and accelerometer features (83.3%) are higher than using gesture features (69.0%) while the significance is marginal*.

We attribute this result to the interference of personal interaction style (e.g., typing skill, familiarity with manipulation gestures, etc.) on the gesture features. It is worth further exploration of how to minimize such effect on the analysis, e.g., normalize the dialing and typing data by subtracting the mean value to every data sample.

Discussion and Future Work

In this work we demonstrate the feasibility of *implicit* detection of PD motor impairment by analyzing users' smartphone interaction activities when they perform a variety of everyday tasks. We believe that the high classification accuracies are achieved because of the hand tremor, finger bradykinesia, rigidity and other PD signs may alter smartphone interaction kinetics in a way detectable through analyzing data from rich sensors such as accelerometers, pressure sensors and touchscreens. Users do not need to wear any dedicated sensor or remember to perform a structured test. The detection happens implicitly without interfering with the normal use of smartphones.

Our methods were tested in a controlled environment. Although participants were asked to perform these tasks as they would normally do to mimic actual routine use of smartphones, further large-scale deployments in real-world settings are necessary to discard a significant impact of the controlled study on participants' interaction behavior. Moreover, we plan to include a larger and balanced cohort that enables a comprehensive review of the influence of the potential confounding variables such as disease stage, medication state and cognitive deficits.

It is worth exploration of developing a regression model based on the current classification model, which could make it possible to conduct fine-grained evaluation on the motor function and the disease progression. Addition information from user's everyday smartphone interaction, such as voice signal during phone calls, accelerometer data during idle periods, could be used to complement our analysis.

References

1. Arroyo-Gallego, T., Ledesma-Carbayo, M. J., Sanchez-Ferro, A., Butterworth, I., Sanchez-Mendoza, C., Matarazzo, M., ... & Giancardo, L. (2017). Detection of Motor Impairment in Parkinson's Disease via Mobile Touchscreen Typing. IEEE Transactions on Biomedical Engineering.
2. Bot, Brian M., et al. "The mPower study, Parkinson disease mobile data collected using ResearchKit." Scientific Data 3(2016):160011.
3. F. L. Campos et al., "Rodent models of Parkinson's disease: Beyond the motor symptomatology," Frontiers Behav. Neurosci., vol. 7, Jan. 2013, Art. no. 175.
4. Gabel, M., Gilad-Bachrach, R., Renshaw, E., & Schuster, A. (2012, August). Full body gait analysis with Kinect. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (pp. 1964-1967). IEEE.
5. Ghika, J., Wiegner, A. W., Fang, J. J., Davies, L., Young, R. R., & Growdon, J. H. (1993). Portable system for quantifying motor abnormalities in Parkinson's disease. IEEE Transactions on Biomedical Engineering, 40(3), 276-283.
6. Keijsers, N. L., Horstink, M. W., & Gielen, S. C. (2006). Ambulatory motor assessment in Parkinson's disease. Movement Disorders, 21(1), 34-44.
7. L. M. L. de Lau and M. M. B. Breteler, "Epidemiology of Parkinson's disease," Lancet. Neurol., vol. 5, no. 6, pp. 525–35, Jun. 2006.
8. Mazilu, S., Blanke, U., Hardegger, M., Tröster, G., Gazit, E., & Hausdorff, J. M. GaitAssist: a daily-life support and training system for parkinson's disease patients with freezing of gait. In Proceedings of CHI2014.
9. McNaney, R., Poliakov, I., Vines, J., Balaam, M., Zhang, P., & Olivier, P. LApp: a speech loudness application for people with Parkinson's on Google glass. In Proceedings of CHI2015.
10. McNaney, R., Vines, J., Roggen, D., Balaam, M., Zhang, P., Poliakov, I., & Olivier, P. Exploring the acceptability of google glass as an everyday assistive device for people with parkinson's. In Proceedings of CHI2014.
11. Mentis, H. M., Komlodi, A., Schrader, K., Phipps, M., Gruber-Baldini, A., Yarbrough, K., & Shulman, L. Crafting a View of Self-Tracking Data in the Clinical Visit. In Proceedings of CHI2017.
12. Patel, S., Lorincz, K., Hughes, R., Huggins, N., Growdon, J., Standaert, D., ... & Bonato, P. (2009). Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. IEEE transactions on information technology in biomedicine, 13(6), 864-873.
13. Patrick, S. K., Denington, A. A., Gauthier, M. J., Gillard, D. M., & Prochazka, A. (2001). Quantification of the UPDRS rigidity scale. IEEE transactions on neural systems and rehabilitation engineering, 9(1), 31-41.
14. Spieker, S., Jentgens, C., Boose, A. & Dichgans, J. (1995). Reliability, specificity and sensitivity of longterm tremor recordings. Electroencephalography & Clinical Neurophysiology, 97, 326–331.