Mobile computing technology for monitoring advanced Parkinson’s disease

Mevludin Memedi
Department of Computer Engineering, Dalarna University

mmi@du.se

ABSTRACT
This paper presents the development and evaluation of Information Technology-based methods and systems for supporting assessment of symptoms and enabling remote monitoring of Parkinson’s disease (PD) patients. In this paper, data from a mobile device test battery, collected during a three year clinical study, was used for the development and evaluation of methods. The data was gathered from a series of tests, consisting of self-assessments and motor tests (tapping and spiral drawing). These tests were carried out repeatedly in a telemedicine setting during week-long test periods.

A computer method was developed to process traced spiral drawings and generate a score representing PD-related drawing impairments. The data processing part consisted of using discrete wavelet transform and principal component analysis. When this computer method was evaluated against human clinical ratings, the results showed that it could perform quantitative assessments of drawing impairment in spirals comparatively well.

Furthermore, an information system was designed and evaluated for delivering assessment support information to the treating clinical staff for monitoring PD symptoms in their patients. The system consisted of a patient node for data collection based on the mobile device test battery, a service node for data storage and processing, and a web application for data presentation. The evaluation results for this integrated system indicate that it can be used as a tool for frequent PD symptom assessments in home environments.

Keywords
Information technology, mobile computing technology, Parkinson’s disease, home assessments, patient monitoring, symptom assessments, discrete wavelet transform, principal component analysis, decision support, web application.

1. INTRODUCTION
Recent advances in a variety of disciplines like wireless communications, mobile computing, sensing technology, decision support systems and the web technology enable patients to be monitored remotely while offering reliable and cost-effective home healthcare [2, 6, 15, and 25]. Mobile computing refers to technologies that employ small portable devices and wireless communication networks that allow user mobility by providing access to data “anytime, anywhere” [5]. An instance of mobile computing technology is Personal Digital Assistants (PDA). These are light-weight handheld computers and one of their medical applications is decision support which provides real-time information access, clinical computational programs, and diagnostic data management [15]. Electronic diaries (e.g. PDAs) overcome the issues of poor patient compliance by including functions that remind patients to complete diary entries at the proper time and inflexible data storage and analysis by allowing just one answer per entry, and stamping the time and the date of the entry [7, 16, and 19].

Parkinson’s disease (PD) is a progressive neurological disorder which is caused by degeneration of dopamine producing nerve cells in the region of the brain called substantia nigra. This region of the brain is important for control of body movement and coordination. The four cardinal motor symptoms comprise of bradykinesia (slowness of initiating voluntary movements), rigidity (increased muscle tone), tremor (a 3-5 Hz hand tremor) and impaired postural stability. With the disease progression and long-term therapy, patients start to experience motor fluctuations. Their motor condition fluctuates between the “off” state (as a result of insufficient medication levels) and the “on” state (in which medication levels are enough for the patient to respond as a non-parkinsonian person). In addition to these two motor conditions, patients in the “on” state may develop abrupt involuntary movements, also known as dyskinesias, in response to peak levels of medication.

In clinical settings, the severity of each of the cardinal symptoms can be scored quantitatively using clinical rating scales. These scales are used as instruments by observers to evaluate PD-related disability and impairment in order to provide a comprehensive clinical picture of the patient. The Unified Parkinson’s Disease Rating Scale (UPDRS) is the most commonly used scale [8]. The UPDRS is made up of four parts covering mentation, behaviour and mood (Part I); activities of daily living (Part II); motor performance (Part III); and complications of therapy (Part IV). In general, a major disadvantage of clinical rating scales is their subjectivity; i.e. the scoring is dependent not only on the motor performance of the patient but also on the examiner’s interpretation [10].

The complexity associated with the clinical management of PD is high as a result of random and frequent changes of symptoms during the day. In the presence of symptom fluctuations, detailed and frequent reporting of multiple measurements related to motor and non-motor symptoms is necessary [26]. This is useful in order to reveal the full extent of a patient’s condition and avoid bias in measuring treatment effects [11].

The adoption of Information Technology (IT) tools can be applied to support the management of PD. A computerized system that enables PD patients to be monitored periodically and remotely offers a reliable and cost-effective screening of those patients.

A test battery, consisting of self-assessments and motor tests (tapping and spiral drawing), was implemented on a PDA with touch screen and built-in mobile communication to be used by patients at home as a telemedicine approach [28].

The aim of this paper is two-fold. The first aim is to develop a computer-based method that processes traced spiral drawings and generates a score useful for representing PD-related drawing impairments thus quantifying the fine motor control in such patients. The second aim is to present the development of an information system for delivering assessment support information to the treating clinical staff for monitoring PD symptoms in their patients. The system first compiles the test battery raw data,
concerning self-assessments and motor tests, into summary scores for conceptual symptom dimensions and an “overall test score” reflecting the overall condition of the patient during a week-long test period and then presents these scores graphically in a web application.

The paper is organized in the following format. In section 2, a review of similar methods and systems found in the scientific literature is provided. Section 3 briefly describes the data collection, using a mobile device test battery, and the clinical study. The methodologies for the development of the computer method and the information system are presented in sections 4 and 5 respectively. The general results for the both aims of the paper are summarized in section 6 followed by the discussions in section 7. And finally, the conclusions are presented in section 8.

2. Related work

Fine motor control can be defined as the ability to perform small and precise movements requiring hand-eye coordination. PD affects the fine motor control of an individual by slowing his/her movements and decreasing reaction time leading to the occurrence of involuntary movements. Quantitative assessment of fine motor impairment has been tried using tracking devices [9], sensing technology [21], scanning devices [13], and computer-interfaced musical keyboards [24].

There have been a number of initiatives from different research groups to address the development of information systems that enable PD patient monitoring. Most of the approaches are based on wearable sensors [4, 12, 20, and 22] and mainly target a specific aspect of the disease such as gait, posture, tremor, and cognition. Typically, these approaches do not address the important issue of defining scores that combine different aspects of patient function in order to facilitate monitoring of the patients.

3. Briefly about the test battery

A test battery was implemented in a handheld computer (PDA) with touch screen and mobile communication for enabling monitoring of PD in home environment conditions [28]. The test battery includes a synchronized collection of self-assessments of common symptoms and fine motor tests. Patients answer questions and perform motor tests by touching the touch screen in a certain position with a stylus. The patients entered responses to seven self assessment questions (q1-q7) and performed fine motor tests four times per day during one to ten test periods of seven days’ length.

A detailed description of the test battery can be found elsewhere [28], however an outline of it is given here. Most of the self-assessment questions (q1, “Ability to Walk”, q3, “Off at worst”, q4, “Dyskinetic at worst”, q5, “Painful cramps”, and q6, “Satisfied with functioning”) relate to the last 4 hours/this morning and are of verbal descriptive scale type with answer alternatives ranging from 1 (worst) to 5 (best). There is another question (q2) in which patients are asked to mark the proportion of time they spent in “off”, “on” and “dyskinetic” states during the last 4 hours/this morning. In addition, patients were asked (q7) to mark the right now condition in one of seven categories of motor conditions ranging from “very off” to “on” to “very dyskinetic”. In addition patients performed different tapping tests (with and without visual cueing) and traced a pre-drawn Archimedean spiral three times per test occasion.

This test battery has been used by 65 patients with advanced PD (treated with intraduodenal levodopa/carbidopa gel infusion, Duodopa®, or candidates for this treatment) at nine clinics around Sweden in the clinical study DAPHNE (Duodopa in Advanced Parkinson’s: Health Outcomes & Net Economic Impact, Eudract No. 2005-002654-21). In total there were 379 test periods and 10439 test occasions. For 223 of these test periods, Hoehn and Yahr and UPDRS ratings were performed in afternoons at the start of the week. Baseline characteristics are shown in Table 1.

4. Development of the computer method

4.1 Manual rating of spiral drawing impairment

A separate web application was constructed to display spiral drawings and to allow users (PD specialists) to rate observed drawing impairment. The web application was organized in four “tracks”: “preliminary rating”, “training”, “standardised rating”, and “rater agreement”. In all these tracks, users could display spirals, rate them and change previously given ratings.

Two raters studied the examples in the clinical handbook for assessment of tremor in spiral drawings in ten categories [1]. The same ordinal severity scale (1-3 = mild impairment, 4-6 = moderate, 7-9 = severe and 10 = extremely severe) was applied for rating of drawing impairment in spiral drawings. The general impression of the shape of the spirals determined the level of severity: homogeneous and symmetrical spiral shapes were rated as mild drawing impairment, larger deviations from the pattern of spiral shape were rated as moderate impairment and spirals with large interruptions, skewed or incomplete shapes were rated as severe. Drawings without signs of a spiral shape were rated as extremely severe.

A ‘standardised manual rating’ score (SMR) was defined as the mean of the two raters’ assessments.

4.2 The computer method

A new method, using the Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA), was developed to process the spiral drawings and generate a ‘wavelet spiral test score’ (WSTS) to represent drawing impairments [29]. The steps involved in the development of the method are given as follows:

4.2.1 Calculate spiral radius

The spiral data consists of x (horizontal position) and y (vertical position) coordinates. In order to perform quantitative evaluation of the drawing impairment in spirals, the raw coordinates were transformed into polar coordinates. The radius (i.e. the square root of the sum of squares of x and y coordinates) represents the degree to which the spiral drawing deviates from the pre-drawn spiral. It was subsequently used in analysis and processing. Validation of input data is performed so that only those spiral drawings that contain >50 data points are processed.
4.2.2 Feature extraction using DWT
Three level decomposition using Daubechies (db10) wavelet family was performed on the signal of the spiral radius to obtain approximations and detail coefficients. A feature vector (reconstructed signal), containing 256 features was obtained after the details (Dn, n = 1, 2, 3) at the first, second and third levels (128+64+32 coefficients) and the approximations (A3) at the third level (32 coefficients) were appended to each other in order of descending levels, i.e. A3, D3, D2 and D1 (Figure 1).

4.2.3 Dimension reduction using PCA
Since the analysis of high-dimension data sets is often complex and time-consuming, PCA was performed due to the large number of features extracted by DWT. At first, PCA was applied on a subset of data, preselected on the basis of the 10% worst and 10% best tapping results. Tapping results were assessed based on both speed and accuracy. By doing this, the first principal component of the radius was assumed to provide a desired direction in multidimensional feature space well representing PD symptom severity. Secondly, after having both the feature vectors from all the spirals and the obtained PCA coefficient matrix from the subset of extreme tapping results, the score was obtained as the first element after multiplying the DWT feature vector by the coefficient matrix.

4.2.4 Scaling of the score
The score was calibrated using logarithmic and linear transformations to bring it to a roughly linear interval scale between 0 and 10, comparable to the manual rating scale. Linear regression was performed between the manually rated spirals from a “training” track and scaled log values in order to reduce systematic errors, such as over-prediction of low impairment or under-prediction of high impairment. The resulting spiral score is hence on denoted WSTS.

4.3 Data Analysis
Test-retest reliability and between-rater agreement was assessed by Spearman rank correlations. In order to avoid the problem of multiple test occasions per individual, 200 random samples of single test occasions per patient were drawn and mean correlations in this sample was calculated. Bland-Altman analysis of difference [3] was used to estimate the prediction error of the WSTS versus the SMR. Correlations between the different spiral scores and tapping test results were assessed on test occasion level, first taking mean values of the three spiral drawings, while correlations with UPDRS were assessed after taking mean values over all spirals drawn during each test period.

5. Development of the information system
5.1 Calculation of summary scores
In order to analyze the test battery data and obtain test period summary scores from it, time series of self-assessments and motor tests were summarized and processed using statistical and machine learning methods [18]. Time series of raw data were summarized into scores for conceptual symptom dimensions and an ”overall test score” providing a comprehensive profile of patient’s health during a test period of about one week. This procedure is schematically depicted in Figure 2 and further described in the following sections.

5.1.1 Data Summary
Initially, time series of test battery items were summarized by calculating statistical features, such as the level (MEAN), fluctuation (standard deviation, SD) and the mean squared deviation (MSD) from “the best” answer alternative for test battery questions: question 1, 3, 4 and 6, on a test period level. In total, for each patient and test period, there were obtained 28 features. The rationale for selection of these statistical features was to define scores taking into account: a) the intensity, b) the frequency and c) the importance of occurring symptoms. Mean values are the obvious choice to represent levels and standard deviations are obvious to represent overall variation. The MSD feature was also used aiming at combining both level and variation.

Figure 1. Example of spiral drawings (left column) and corresponding wavelet coefficients (right column) with drawing impairment rated 1 (first row), 5 (middle row) and 10 (bottom row).

5.1.2 Calculation of test battery dimensions
Some of the test battery items are highly correlated indicating that they measure the same concept. Because of this redundancy, it is possible to combine these items and reduce them into a smaller number of principal components without much loss of information. To achieve this, PCA was applied. The information content of a test period with the test battery can be described by six dimensions: walking (based on question 1), satisfaction (question 6), dyskinesia (question 2-dyskinetic and question 4), off (question 2-off and question 3), tapping (all tapping tests) and spiral (all spirals). For each dimension, PCA was performed using
the correlation matrix method applied to the statistical features for the questions or motor tests that the dimension was based on by retaining the respective first principal components.

Linear transformations were used to scale first principal component scores of the dimensions to a scale from 0 (worst) to 1 (best) score. To evaluate whether dimensions measure the same concept of patient’s health status, Cronbach’s Alpha for dimensions was calculated.

5.1.3 Calculation of overall test score

Using the six derived dimensions, the overall score can be calculated by using a regular hexagon where each dimension becomes a corner on it. The geometrical placement of dimensions in the hexagon was decided based on conceptual relationships between them and in collaboration with domain experts (Figure 4). Using the side-angle-side method the areas of each triangle of the hexagon can be calculated separately and then added up to form an unweighted overall score.

The problem with this approach is that all dimensions, concerning self-assessments and motor tests, have the same weight in the assessment of the overall condition of the patient which is in contradiction with the common clinical rating scales, e.g. UPDRS, used in clinical practice where the highest weight is given to the evaluation of motor symptoms.

To overcome this problem, an alternative way to define the overall score was to weight symptom dimensions using the UPDRS. The standard least-squares regression was used for examining the relationship between first components of dimensions and the patient’s UPDRS score. The weighted overall score (UPW) was defined as a linear combination of the first principal components, with numerical weights estimated by regression technique to fit simultaneous clinical ratings on the UPDRS scale.

The UPW was then scaled to a scale from 0 (worst) to 1 (best) score using linear transformations. Correlations between UPW and other clinical rating scores were assessed by Spearman rank correlations.

5.2 System description

Decision support systems comprise components for (a) database management capabilities with access to internal and external data, information knowledge, (b) modelling functions accessed by a model management system, and (c) user interface designs that enable interactive queries, reporting and graphing functions [cf. e.g. 23]. In the developed system, these processes are handled separately where data collection, storage and processing takes places before presentation. The former two processes are handled by the so-called data processing sub system and the later one is handled by the web application.

5.2.1 Data Processing Sub System

For each test occasion, raw test data are sent from the PDA hand unit over the mobile net to the so-called remote device manager (RDM). The RDM is a commercial product responsible for collecting and storing the data from the hang unit. There is a communication protocol between these two systems which handles the transfer of the data. The system architecture is illustrated in Figure 3.

The data processing sub system (DPSS) is a stand-alone application, which incorporates knowledge to analyze and interpret the raw test battery data. It parses, processes and stores the data into relational database tables and at the same time calculates and stores test scores on test period level. A connection with RDM is first established followed by receiving and parsing XML data from files. Once files are received, they can be directly interpreted by the DPSS which runs during a specified time interval, e.g. every single hour. The main reason for processing data centrally instead of locally is the risk of losing raw data in the remote devices if it is not uploaded regularly. Access to raw data centrally also facilitates future research and method developments. The data collection in hand units is designed to minimize upload bandwidth [28]. For computing the spiral score according to the method explained in section 4 a single piece of M-code was written in Matlab® (Mathworks Inc.). This code was then encrypted and wrapped into a C# interface by using the Matlab Builder for .NET to be accessed by the DPSS.

5.2.2 Web application

The web application (WA) is a feedback system comprising a secure web server and a database with web-based access for medical staff. The main role of the WA in the overall system is to present test results to the end-user clients. It is designed in line with users’ requirements and needs for decision support by displaying the test results per patient and test period in graphical and tabular format. Some important user requirements specifications for the WA were: easy to understand and use, user-friendly design, easy navigation, fast response, rapid and convenient screening of patients, a comprehensive overview of a patient’s condition on a single page. They were drafted in collaboration with experts and different prototype versions were developed in several iterations.

![Figure 3. The overall system architecture](image-url)
To enable rapid patient status assessment, the information in WA is displayed and ordered using a top-down approach where the general overview of the patient’s performance per test period is given (Figure 4). Raw data and other detailed information may be accessed in more “advanced” displays. After the user logs in and selects a relevant patient, the main page displays the graphs of the patient scores on a test period level, focusing on the overall test score, dimensions and daily summaries. These graphs can be updated interchangeably.

Figure 4. Patient status report in WA with graphical visualization of time series of test measures [18]

5.2.3 System’s usability evaluations
The potential usefulness of the overall system was evaluated by an advisory board consisting of 14 neurologists after presentation and demonstration of its functionality. They were based in the following countries: USA 3, Germany 3, Italy 2, Spain 1, Netherlands 1, UK 1, Sweden 1, Denmark 1 and Finland 1. They responded to questions by pressing keypads.

The WA was demonstrated to fifteen nurses from the nine clinics in Sweden who had experience using the mobile test battery device in the DAPHNE study, but they had not previously seen the WA. The evaluation was performed as a presentation session for the nurses where they asked the presenter to display certain data from a particular patient in whom they had an interest. The presenter took notes about the reactions from the group. At least one patient per clinic was shown.

One year later, the IBM Computer System Usability Questionnaire (CSUQ) was administered to evaluate the nurses’ satisfaction with the WA. They were asked to perform a series of tasks using the WA, such as (a) login, (b) select a patient, (c) check the patient’s performance by looking at the graphs of summarized scores and compare these results with own clinical observations, (d) select and check the other patients’ results repeating steps (b) and (c), (e) complete the survey and (f) logout. The questionnaire was web-based and more than one person per clinic involved in the study could check and update the answers to the questionnaire giving a consensus response. CSUQ consists of 19 items on a seven-point rating scale ranging from 1 (strongly agree) to 7 (strongly disagree). Four usability scores were derived after responses to the CSUQ items were averaged: Overall Satisfaction, System Usefulness, Information Quality and Interface Quality [14]. Data about gender, age and previous experience with computer applications were also noted.

6. Results
6.1 Computer method
Test–retest reliability coefficients were as follows: 0.77 for WSTS, 0.71 for first rater and 0.70 for second rater. In general, correlations (absolute values) between the WSTS and tapping results were low to moderate (Table 2). Correlations within the tapping results were 0.58 or lower.

The correlations between WSTS and UPDRS scores were as follows: 0.41 to total UPDRS, 0.38 to UPDRS subsection III (motor exam) and 0.51 to UPDRS subsection II (activities of daily living). The correlation between WSTS and SMR for the spiral drawings was (0.89) strong. The 95% confidence interval for the prediction error for the WSTS was ±1.5 and the mean value (bias) was 0.39 scale units. Rater agreements were as follows: 0.87 between raters, 0.91 between WSTS and mean rating.

Table 2. Mean Spearman correlations after repeated random sampling of one occasion per patient [29].

<table>
<thead>
<tr>
<th></th>
<th>WSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMR</td>
<td>0.89</td>
</tr>
<tr>
<td>Total UPDRS</td>
<td>0.41</td>
</tr>
<tr>
<td>T1 speed (Table 3, average speed of test 8 and test 9)</td>
<td>-0.40</td>
</tr>
<tr>
<td>T1 accuracy (Table 3, average accuracy of test 8 and test 9)</td>
<td>-0.45</td>
</tr>
<tr>
<td>T2 speed (Table 3, test 10)</td>
<td>-0.56</td>
</tr>
<tr>
<td>T3 speed (Table 3, test 11)</td>
<td>-0.52</td>
</tr>
<tr>
<td>T3 accuracy (Table 3, test 11)</td>
<td>-0.50</td>
</tr>
</tbody>
</table>

6.2 Information system
The Cronbach’s Alpha test showed that there was a good internal consistency between dimensions with a coefficient of 0.81. The correlations between UPW and total UPDRS (-0.6, p<0.001) were adequate whereas between UPW and Hoehn and Yahr (-0.44, p<0.001) were medium.

Eleven neurologists had a positive impression regarding the overall impression of the integrated system, one had a neutral impression and two had negative impression. The most important benefits they could see were an increased ability to identify patients who are not doing well and facilitated follow-up optimization of an individual’s treatment. The system was seen as most important for complicated patients and for regional patients, that is patients living in regions far away from a clinic. The general conclusion of the board was that the system was recognized as a tool that will assist in management of patients.

At the presentation session, the responses from nurses are summarized in a qualitative manner as follows: (a) WA is very useful, (b) the results during test periods showed agreement with qualitative observations of the patient during that test period, for example, “one patient was in a bad condition in baseline, he improved after starting Duodopa, then he became worse again, 24-hour infusion started and the patient became better again; we can clearly follow this in the web application”, (c) comparisons between patients are possible (one patient is in better/worse condition than another).

Responses to the CSUQ were obtained from seven of the nine clinics and the results were mixed. A majority of the clinics were quite satisfied with the usability although a sizeable minority were
not. All evaluators were female and ages ranged from 38 to 61 (mean value 49). Two out of seven asserted that they had much previous experience with computers, whereas four had some experience and one had little experience.

7. Discussion
The WSTS method could automatically assess the drawing impairment in traced spirals. It was applicable during event-based data acquisition, in the way that spiral data were usable even if spirals were drawn slowly, which is often the case in advanced PD. In another study, event-based data acquisition and signal processing techniques suitable for non-stationary signals were shown to be useful in systems for patient-home monitoring [17]. One limitation of the WSTS method was that it over-predicted small impairments and under-predicted moderate impairments, compared to SMR. Therefore, further optimization and transformation of the data is required. The method could not exactly classify the drawing impairment in spiral drawings into the correct category on the scale 1-10. Using an ordinal scale with fewer categories would improve the classification performance of the method. Another weakness of the method is that it only detects disability but lacks the mechanism to distinguish between specific symptoms causing the particular disability. This problem will be addressed in future research work.

The information system summarized the test battery raw data into summary scores for conceptual symptom dimensions and an overall test score (UPW) per patient and test period. From the regression equation of the UPW, motor test results which consisted of tapping tests and spirals were given the highest weight of 65%, just as motor section (Part III) has the highest weight in UPDRS. The summary scores have been validated in a separate study (in Italy, 30 patients), with the objective of assessing test-retest reliability, correlations to other assessment methods, and the ability to detect differences between patient groups at different stages [27]. When developing and applying web-based systems for telemedicine applications several issues should be taken into consideration such as (a) architectural and technical, (b) security, privacy and confidentiality, and (c) usability and user acceptance [2]. The information system is at a prototype stage and its architecture and technology allows data collection, retrieving and visualization. The WA relies on password-based user authentication but lacks data encryption. However, the current usability evaluation is missing feedback from PD physicians and this limitation will be taken into account in future usability tests for the final version of the system.

8. Conclusions
The WSTS method could assess PD-related drawing impairments well comparable to trained raters. It resulted in a strong correlation with the SMR and a medium correlation with the UPDRS scores. Its approach for classifying the impairment in spirals through multi-resolution representation of the radius and dimension reduction with a selected subset of data is a novel and powerful one.

The information system was designed to provide PD patient status evaluation through convenient access to a patient’s current symptom profile and history. The system processes and summarized various time series of self-assessments and motor tests into different summary scores. The method to take the first principal component of mean, SD and MSD for each feature, is a data-driven general way to automatically obtain weights for combining level and variation. The internal consistency between the six conceptual symptom dimensions was good. Correlations between the UPW and total UPDRS were adequate and significant. The derived UPW may be helpful in (a) facilitating the process of screening patients, (b) avoiding sub-optimization of treatments, and (c) deciding if a treatment change leads to an improvement of a patient’s general condition or not.

REFERENCES


