Mälardalen University Press Dissertations
No. 153

FIRST-PRINCIPLE DATA-DRIVEN MODELS FOR ASSESSMENT OF MOTOR DISORDERS IN PARKINSON'S DISEASE

Taha Khan

2014

School of Innovation, Design and Engineering

ISBN 978-91-7485-142-7
ISSN 1651-4238
Printed by Arkitektkopia, Västerås, Sweden

Mälardalen University Press Dissertations
No. 153

FIRST-PRINCIPLE DATA-DRIVEN MODELS FOR ASSESSMENT OF MOTOR DISORDERS IN PARKINSON'S DISEASE

Taha Khan

Akademisk avhandling

som för avläggande av teknologie doktorsexamen i datavetenskap vid Akademin för innovation, design och teknik kommer att offentligen försvaras onsdagen den 16 april 2014, 13.00 i Clas Ohlsonsalen, Campus Borlänge, Högskolan Dalarna.

Fakultetsopponent: professor Gunnar Klein, Örebro universitet

Akademin för innovation, design och teknik

Abstract

Parkinson's disease (PD) is an increasing neurological disorder in an aging society. The motor and non-motor symptoms of PD advance with the disease progression and occur in varying frequency and duration. In order to affirm the full extent of a patient's condition, repeated assessments are necessary to adjust medical prescription. In clinical studies, symptoms are assessed using the unified Parkinson's disease rating scale (UPDRS). On one hand, the subjective rating using UPDRS relies on clinical expertise. On the other hand, it requires the physical presence of patients in clinics which implies high logistical costs. Another limitation of clinical assessment is that the observation in hospital may not accurately represent a patient's situation at home. For such reasons, the practical frequency of tracking PD symptoms may under-represent the true time scale of PD fluctuations and may result in an overall inaccurate assessment. Current technologies for at-home PD treatment are based on data-driven approaches for which the interpretation and reproduction of results are problematic.

The overall objective of this thesis is to develop and evaluate unobtrusive computer methods for enabling remote monitoring of patients with PD. It investigates first-principle data-driven model based novel signal and image processing techniques for extraction of clinically useful information from audio recordings of speech (in texts read aloud) and video recordings of gait and finger-tapping motor examinations. The aim is to map between PD symptoms severities estimated using novel computer methods and the clinical ratings based on UPDRS part-III (motor examination). A web-based test battery system consisting of self-assessment of symptoms and motor function tests was previously constructed for a touch screen mobile device. A comprehensive speech framework has been developed for this device to analyze text-dependent running speech by: (1) extracting novel signal features that are able to represent PD deficits in each individual component of the speech system, (2) mapping between clinical ratings and feature estimates of speech symptom severity, and (3) classifying between UPDRS part-III severity levels using speech features and statistical machine learning tools. A novel speech processing method called cepstral separation difference showed stronger ability to classify between speech symptom severities as compared to existing features of PD speech. In the case of finger tapping, the recorded videos of rapid finger tapping examination were processed using a novel computer-vision (CV) algorithm that extracts symptom information from video-based tapping signals using motion analysis of the index-finger which incorporates a face detection module for signal calibration. This algorithm was able to discriminate between UPDRS part III severity levels of finger tapping with high classification rates. Further analysis was performed on novel CV based gait features constructed using a standard human model to discriminate between a healthy gait and a Parkinsonian gait.

The findings of this study suggest that the symptom severity levels in PD can be discriminated with high accuracies by involving a combination of first-principle (features) and data-driven (classification) approaches. The processing of audio and video recordings on one hand allows remote monitoring of speech, gait and finger-tapping examinations by the clinical staff. On the other hand, the first-principles approach eases the understanding of symptom estimates for clinicians. We have demonstrated that the selected features of speech, gait and finger tapping were able to discriminate between symptom severity levels, as well as, between healthy controls and PD patients with high classification rates. The findings support suitability of these methods to be used as decision support tools in the context of PD assessment.

ISBN 978-91-7485-142-7
ISSN 1651-4238

First-principle data-driven models for assessment of motor disorders in Parkinson's disease

– Doctoral Thesis –

Taha Khan
School of Innovation, Design and Engineering
Mälardalen University, Västerås, Sweden
tkh@du.se

Abstract

Parkinson's disease (PD) is an increasing neurological disorder in an aging society. The motor and non-motor symptoms of PD advance with the disease progression and occur in varying frequency and duration. In order to affirm the full extent of a patient's condition, repeated assessments are necessary to adjust medical prescription. In clinical studies, symptoms are assessed using the unified Parkinson's disease rating scale (UPDRS). On one hand, the subjective rating using UPDRS relies on clinical expertise. On the other hand, it requires the physical presence of patients in clinics which implies high logistical costs. Another limitation of clinical assessment is that the observation in hospital may not accurately represent a patient's situation at home. For such reasons, the practical frequency of tracking PD symptoms may under-represent the true time scale of PD fluctuations and may result in an overall inaccurate assessment. Current technologies for at-home PD treatment are based on data-driven approaches for which the interpretation and reproduction of results are problematic.

The overall objective of this thesis is to develop and evaluate unobtrusive computer methods for enabling remote monitoring of patients with PD. It investigates first-principle data-driven model based novel signal and image processing techniques for extraction of clinically useful information from audio recordings of speech (in texts read aloud) and video recordings of gait and finger-tapping motor examinations. The aim is to map between PD symptoms severities estimated using novel computer methods and the clinical ratings based on UPDRS part-III (motor examination). A web-based test battery system consisting of self-assessment of symptoms and motor function tests was previously constructed for a touch screen mobile device. A comprehensive speech framework has been developed for this device to analyze text-dependent running speech by: (1) extracting novel signal features that are able to represent PD deficits in each individual component of the speech system, (2) mapping between clinical ratings and feature estimates of speech symptom severity, and (3) classifying between UPDRS part-III severity levels using speech features and statistical machine learning tools. A novel speech processing method called cepstral separation difference showed stronger ability to classify between speech symptom severities as compared to existing features of PD speech. In the case of finger tapping, the recorded videos of rapid finger tapping examination were processed using a novel computer-vision (CV) algorithm that extracts symptom information from video-based tapping signals using motion analysis of the index-finger which incorporates a face detection module for signal calibration. This algorithm was able to discriminate between UPDRS part III severity levels of finger tapping with high classification rates. Further analysis was performed on novel CV based gait features constructed using a standard human model to discriminate between a healthy gait and a Parkinsonian gait.

The findings of this study suggest that the symptom severity levels in PD can be discriminated with high accuracies by involving a combination of first-principle (features) and data-driven (classification) approaches. The processing of audio and video recordings on one hand allows remote monitoring of speech, gait and finger-tapping examinations by the clinical staff. On the other hand, the first-principles approach eases the understanding of symptom estimates for clinicians. We have demonstrated that the selected features of speech, gait and finger tapping were able to discriminate between symptom severity levels, as well as, between healthy controls and PD patients with high classification rates. The findings support suitability of these methods to be used as decision support tools in the context of PD assessment.

Acknowledgment

First and foremost, I would thank the Higher Education Commission of Pakistan for enrolling me into their program 'MS leading to PhD, Phase-II Batch-1 for Sweden 2007'. I am grateful to Dr. Ata-ur-Rehman and Dr. Javaid Laghari, former chairmen of the Commission, who faced a lot of hurdles to keep the program running when the country was in the state of financial crisis. Because of their efforts, I was able to make this long journey.

I am deeply grateful to my supervisors, Mark Dougherty, Jerker Westin, Peter Funk and Shahina Begum, for their constant guidance, support, motivation and their unceasing help during the course of my PhD.

My deepest gratitude is to Mark and Jerker for being always ready to listen to my problems and coming up with new and creative ideas based on the immense knowledge they have. I have been extremely fortunate to have you as supervisors who gave me freedom to explore on my own and at the same time guided me to recover when my steps faltered. Your patience and confidence in me helped me overcome many crises and finish this dissertation. I thank you Jerker for arranging all those picnic cum meetings that strengthened the relationship between our research group members and allowed us to discuss many new ideas in a really friendly environment. I thank you again for sharing the office and for always giving attention to my questions. This has helped me in solving many problems quickly and efficiently. I am deeply indebted to you for all the time and effort you gave. I cannot imagine having better supervisors and mentors than you, Mark and Jerker, for my PhD study.

I am thankful to my MS thesis supervisor Hasan Fleyeh for developing my interest in image and signal processing. The skills I gained under his supervision have helped me a lot in my PhD research. I am thankful to Siril Yella for giving fruitful ideas about speech processing. A special thanks to my friend Mevludin Memedi for having many scientific and non-scientific discussions. I acknowledge your support inside and outside the university.

I am thankful to Dag Nyholm for co-authoring my papers and for giving advice, motivation and encouragement regarding the publication of the work, and for sharing the clinical data. I would like to thank the Kinetics Foundation for providing the speech data. Thanks to the Swedish Knowledge Foundation and Abbott for funding the research and for arranging visits to international conferences. I want to thank all the anonymous reviewers of my papers for their valuable feedback.

Finally, I thank my father Dr. Saleem Iftekhar for being a source of encouragement and motivation, and for his everlasting moral and financial support, in bringing me up. It is because of you that I am about to become a PhD. Importantly; I would acknowledge the love and warmth of my mother, my wife, my brothers and sister for all the success of my efforts. Lastly, I am grateful to my wife for the patience in the difficult times we had. My little princess (daughter) Amna, for all the welcomes and hugs I got after returning from work that swept away all my tiredness, and for the time you gave to your 'Baba' to complete his PhD. I love you all!

Taha Khan, Borlange, January 2014

To my family

List of Papers

Publications included in the thesis:

1. **Methods for detection of speech impairment using mobile devices.** Taha Khan, Jerker Westin. *Recent Patents on Signal Processing,* vol. 1, nr 2, p163-171, Bentham Science, 2011, doi: 10.2174/2210686311101020163.

2. **Running-speech MFCC are better markers of Parkinsonian speech deficits than vowel phonation and diadochokinetic**. Taha Khan. Submitted for publication.

3. **Cepstral separation difference: a novel approach for speech impairment quantification in Parkinson's disease.** Taha Khan, Jerker Westin and Mark Dougherty. *Biocybernetics and Biomedical Engineering,* vol. 34, nr 1, p25-34, Elsevier, 2014, doi: 10.1016/j.bbe.2013.06.001.

4. **Classification of speech intelligibility in Parkinson's disease**. Taha Khan, Jerker Westin and Mark Dougherty. *Biocybernetics and Biomedical Engineering,* vol. 34, nr 1, p35-45, Elsevier, 2014, doi: 10.1016/j.bbe.2013.10.003.

5. **A computer vision framework for evaluation of finger tapping in Parkinson's disease**. Taha Khan, Dag Nyholm, Jerker Westin and Mark Dougherty. *Artificial Intelligence in Medicine,* vol. 60, nr 1, p27-40, Elsevier, 2014, doi: 10.1016/j.artmed.2013.11.004.

6. **Computer vision methods for Parkinsonian gait analysis: a review on patents.** Taha Khan, Peter Grenholm and Dag Nyholm. *Recent Patents on Biomedical Engineering*, vol. 6, nr 2, p97-108, Bentham Science, 2013, doi: 10.2174/1874764711306020004.

7. **Motion cue analysis for Parkinsonian gait recognition.** Taha Khan, Jerker Westin and Mark Dougherty. *The Open Biomedical Engineering Journal,* vol. 7, nr 1, p1-8, Bentham Science, 2013, doi: 10.2174/1874120701307010001.

Reprints were made with permission from the respective publishers.

Additional publications not included in this thesis:

Journal

- **Automatic and objective assessment of alternate tapping performance in Parkinson's disease.** Mevludin Memedi, Taha Khan, Dag Nyholm, Peter Grenholm, Jerker Westin. *Sensors*, vol. 13, nr. 12, p16965-16984, MDPI, 2013, doi: 10.3390/s131216965.

Conferences

- **A computer vision framework for finger-tapping evaluation in Parkinson's disease [abstract].** Taha Khan, Dag Nyholm, Jerker Westin, Mark Dougherty. *Movement Disorders 2013*, vol. 28, Supplement 1:302.

- **Quantification of speech impairment in Parkinson's disease [abstract]**. Taha Khan, Jerker Westin, Peter Funk, Mark Dougherty. *Movement Disorders 2012*, vol. 27, Supplement 1:1559.

- **Assessment of PD Speech Anomalies @ Home [abstract]**. Taha Khan, Jerker Westin. *Movement disorders 2011*, vol. 26, Supplement 1:1080.

- **Motion Cues Analysis for Parkinson Gait Recognition [abstract]**. Taha Khan, Jerker Westin. *Movement disorders 2011*, vol. 26, Supplement 1:1074.

Patent

Title of invention:	Cepstral Separation Difference
Publication number:	WO2013187826 A2
Publication date:	19th December, 2013
Inventors:	Taha Khan, Jerker Westin, Mark Dougherty

My contributions to the included papers:

Paper 1 – planning the literature review, conducting the review, writing the first version of the manuscript and revising it.

Paper 2 – data analysis, results interpretation, writing the first version of the manuscript.

Paper 3 – method development, data analysis, results interpretation, writing the first version of the manuscript and revising it.

Paper 4 – method development, data analysis, results interpretation, writing the first version of the manuscript and revising it.

Paper 5 – data collection, method development, data analysis, results interpretation, writing the first version of the manuscript and revising it.

Paper 6 – planning the literature review, conducting the review, writing the first version of the manuscript and revising it.

Paper 7 – method development, data analysis, results interpretation, writing the first version of the manuscript and revising it.

Contents

List of figures

List of tables

Abbreviations

AI	Artificial intelligence
ANOVA	Analysis of variance
AoC	Area under the ROC
COG	Centre of gravity
CSD	Cepstral separation difference
CV	Computer vision
DDK	Diadochokinesis
DFT	Discrete Fourier transform
DSS	Decision support systems
FPDD	First-principle data-driven
HC	Healthy controls
HGA	Holistic-based gait analysis
IDFT	Inverse discrete Fourier transform
ICC	Intra-class correlation coefficient
LLGMN	Log-linearized Gaussian mixture networks
MFCC	Mel-frequency cepstral coefficients
MGA	Model-based gait analysis
PD	Parkinson's disease
RC	Research contributions
ROC	Receiver operating characteristic
RQ	Research questions
RFT	Rapid finger tapping test
SMO	Sequential minimum optimization
SVM	Support vector machines
SVP	Sustained vowel phonation
TPR	True positive rate
TRS	Text-dependent running speech
UPDRS	Unified Parkinson's disease rating scale
UPDRS-FT	UPDRS motor finger-tapping examination
UPDRS-S	UPDRS motor speech examination

1. Introduction

1.1. Computers in medicine

Effective medical treatment begins with a correct diagnosis. Sometimes, the treatment is empirically determined based on a number of disease symptoms, where a successful intervention relies on the ability of experts to accurately assess these symptoms in patients (Wu, 1990). In the current information age, it no longer matters how much information is memorized by the expert because a computer can keep track of more findings, test parameters of the disease and possible remedies to the problem that a person is able to memorize. However, there are cases in which the problem has solutions that 'common sense' can find quicker and with better accuracy than a computer. In such cases a computer can never compete with humans (McCarthy, 1989). Hence if the knowledge of an expert could be partnered with computers, the enhanced accuracy, speed and efficiency of symptom assessment could contribute significantly to the process of medical treatment (Shortliffe, 1976).

The capacity for computers to aid clinicians is growing. Computer-aided systems have become an important part of medical procedures that assist clinicians in symptom interpretation. This interpretation requires computer to process biomedical signals that yield a great deal of hidden clinical information, which allows clinicians to draw accurate and timely conclusions based on the symptom status. Using these biomedical signals, the clinicians can perform manual analysis by observing clinical trends in the signal data. However the level of complexity associated with this type of analysis is huge even for experienced clinicians. The trick is then to use Artificial Intelligence (AI) methods which have shown promise in automating the disease evaluation to support in medical prescription (Reiter, 1987).

In the 1970s, the research in AI was primarily aimed at engineering a 'general problem solver' which if fed with the right input could solve problems in any domain using its problem-solving architecture and all-purpose knowledge base (Gorry, 1973). It was assumed that experts differed not in their method of reasoning but in the content to which they apply their reasoning. This was seen as an effort to encode common sense, which eventually failed because of the intractable volume of information required to emulate human cognition. However, this continued effort gave rise to the emerging concepts of decision support systems (DSS) having manageable domain scopes and the ability to perform complex tasks much faster than human

beings, for example mathematical analysis of biological signals (Weiss et al., 1978). Specifically in the field of health care, the clinical DSS can reduce hospital resources and treatment costs. For instance in circumstances where the number of physicians is limited, such systems can serve as assistants to the physicians, provide a second opinion in diagnosis and give them access to new experience and knowledge (Shortliffe, 1987). Moreover, the DSS can be particularly useful in reducing individual variations and subjectivity regarding the clinical analysis of symptoms. Today, many clinical DSS have been developed to be multipurpose and combine more than one AI method and technique.

Apart from the implication of AI methods in medical diagnosis, telemedicine is an emerging option in general medical care, affording cost effective and reliable screening, and alleviating the burden of frequent visits of patients to the clinic (Bellazzi et al., 2001). It involves remote monitoring of patients who are not at the same location as their healthcare provider. This is accomplished by employing monitoring and computing devices at the patient's home. The results of these devices are transmitted via communication networks to the healthcare provider. The provider can make decisions about the clinical treatment of the patient based on a combination of objective and subjective information. This is similar to what would be revealed during an on-site appointment. Telemedicine devices are capable of recording and providing information about vital signs, which is handy specifically for patients with chronic diseases, provided that the patients have the necessary equipment at their location (Chen et al., 2011). For the patients, it is useful because they can receive feedback regarding their symptoms much more quickly than they otherwise might. An additional advantage is that, since the patients are more involved in their own treatment, they become more knowledgeable about their symptom profiles and gain a better understanding of how and when these symptoms appear, and the ways these can be treated. This eases the communication between clinicians and patients, and helps in enhancing the quality of clinical evaluation, as well as supports in improving the patients' self-care ability.

Some drawbacks of telemedicine include the cost of telecommunication, data acquisition and data management equipment, and possible technical training for medical personnel (Hjelm, 2005). It is also possible that a poor quality of symptom estimate is delivered to the clinician, potentially due to disruption in the transmitting medium or environmental interference in data acquisition such as inclusion of background noise in speech and video recordings (Angaran, 1999). Other obstacles to telemedicine include dubious legal regulation for telemedical practices and difficulty in claiming reimbursement from insurers.

In a telemedicine setting for clinical decision support, it is difficult to formalize human diagnostics into AI models since human reasoning depends on multiple cognitive activities consisting of information collection, pattern

identification, problem solving and decision making with a certain degree of uncertainty (Miller and Geissbuhler, 2007). The early AI systems in medical decision making were mainly developed using truth tables or decision trees. Later on, data-driven approaches such as artificial neural networks, Bayesian statistics etc. were introduced to build clinical DSS that attempt to codify statistical intuitions and the experience of human experts in the medical domain (Little, 2006). These methods utilized clinical information to produce therapeutic predictions in a systematic way that supported clinicians in their decision-making process. Albeit these methods were effective to model relationship between patterns of input attributes (predictors) and medically relevant outcomes (predicted scores), however this was only possible at the cost of great amount of data that was required to generalize this relationship. Another problem associated with the data-driven approach was that the structure and design of the corresponding medical problem being diagnosed was weakly represented. This hindered the understanding of the underlying biological dysfunctions for non-experts, thus creating a barrier in user-friendly interpretation of results that is of paramount importance for any decision support tool (Bemmel and Musen, 1997).

Besides, another AI approach referred to as diagnosis from first principles (Reiter, 1987) use an understanding of physical mechanics to derive mathematical formulae that represent disease effects. The advantage is the depth of insight into the behavior of biological functions that further improves clinical representation of a disease. Such computer tools can be particularly useful to clinicians in tracking fluctuating symptoms in neurological disorders such as Parkinson's disease (PD). The tracking can be further improved using unobtrusive data-acquisition techniques that do not impede the natural movements of patients when performing the test, which improves the accuracy in symptom assessment (Tsanas, 2012). Among different unobtrusive ways of data collection in telemedicine, the processing of speech signals fits ideally the purpose of monitoring. Speech can be self-recorded by the patients and the equipment required to record speech is readily available in the form of mobile phones. Moreover, speech estimates can be easily transmitted on standard cellular mobile networks to a centralized server for clinical evaluation. Other unobtrusive data acquisition methods in telemedicine may include video-recording of motor actions such as gait and hand movements using web-cameras attached to a computer. These videos can be processed using computer vision (CV) methods to extract biometric information, which can provide additional evidence regarding patient's condition and can support clinicians in adjusting the medical treatment.

1.2 Parkinson's disease and treatment

Neurological disorders claim lives at an epidemic rate worldwide, with PD being the second most common disorder after Alzheimer's (de Rijk et al.,

1999). According to sources, PD is more prevalent in men than in women (Haaxma et al., 2007; Baldereschi et al., 2000) and the lifetime risk, considering the current global average life expectancy, is estimated to be 4.4% and 3.7% for men and women respectively (Elbaz et al., 2002). Studies (Rajput et al., 2007) suggest that age is the most important risk factor for PD onset and PD is more ubiquitous in approximately 2% of people over the age of 65. According to Campenhausen et al., (2005), the prevalence and incidence rates of PD in the European population alone are estimated to be 108-257/100,000 and 11-19/100,000 respectively. A further study (Lang and Lozano, 1998) revealed that there are more than one million patients with PD in North America, where an estimated 20% of the patients go undiagnosed. It was speculated that given the growing elderly population, the number of PD patients will double by 2030.

PD is named after James Parkinson (1817) who reported an 'An essay on the shaking palsy'. Parkinson himself referred to the disease as 'paralysis agitans' which was later termed 'Parkinson's disease' by Jean-Martin Charcot in 1876 (Haas, 2001). Numerous surgical and pharmaceutical techniques were developed since then as remedies against PD. However the milestone in PD treatment was set by Arvid Carlsson who introduced Levodopa in 1950s. He discovered that PD results in the loss of dopaminergic neurons in the mid-brain (Carlsson, 1974). These neurons serve as a messenger that allows communication between the mid-brain and other parts of the brain, which is responsible for producing smooth and controlled body movements. A lack of dopamine causes four cardinal motor symptoms comprising of bradykinesia (slowness of movements), rigidity (increased muscle tone), tremor (e.g. 3-5 Hz hand tremor) and impaired postural stability. These motor symptoms are accompanied by non-motor symptoms such as sleep disorders, impairment in cognition, problems in sexual health and fatigue etc. (Wolters et al. 2007). The symptoms advance with the disease progression and demote the quality of life of patients with PD.

No permanent medical cure for PD has been reported till today. Currently available medicine and surgical interventions are capable of alleviating some of the PD symptoms, but only for a short duration of time. The most common therapy is levodopa that acts as a precursor of dopamine. Even 40 years after its discovery (Carlsson, 2002), it remains the most effective PD medication. It has been reported that 70-80% of the PD patients are currently treated with levodopa therapy (Parkinson's disease foundation's webpage: www.pdf.org; last accessed Oct 2013). The standard symptomatic treatment in the initial stage of PD is aimed at restoring depleted stimulation of dopamine receptors, where the induction of levodopa helps effectively improve the patient's motor functions. However in advanced stages, patients continue to experience motor complications within hours or minutes of taking medication.

In the advanced stages of PD, a number of symptoms may occur in varying frequency and duration. It was reported that 50% of the patients may have these problems after 5 years of taking levodopa and nearly 100% of patients after 10 years (Van Laar, 2003). Due to the symptom variation, the PD medication targeting dopamine receptors must be individually tuned (Bayulkem and Lopez, 2010) due to the fact that the under-dosing of medication does not relieve the symptoms and overdosing leads to abrupt involuntary body movements (dyskinesias). The dosage needs to be adjusted daily with respect to time of the day, mood, food intake and daily physical activities. Further, these treatments need to be followed up regularly over time as the interval for the required dosage level narrows as the disease progresses (Mouradian et al., 1988). Albeit in the advanced disease stage, the increased symptom fluctuations result in severe disabilities amongst patients, however the recent experiments reported by Nyholm et al. (2003, 2005) suggest that the continuous delivery of levodopa/carbidopa gel (Duodopa, Abbot laboratories) is capable of controlling motor fluctuations in advanced PD.

1.3 First-principles vs. data-driven models

In general, multiple symptoms can occur simultaneously in neurological disorders, specifically in PD (Wolters et al. 2007). For such cases, the incorporation of AI systems must not only account for the given clinical manifestation of symptom combinations, but it must also satisfy some notion of simplicity and parsimony in symptom interpretation and processing.

The initial phase in modeling of an AI system for clinical decision support is collecting and systematic treatment of available clinical knowledge (Weiss et al., 1978). The a priori knowledge about symptoms comes from the available clinical analysis, comprising of finding all possible connections between the symptoms and physical phenomena of disease. The availability of an a priori in modeling an AI system allows one to develop 1) the final type of model, 2) accuracy validation criteria, 3) the type of specific modeling procedures, 4) the determination of model complexity and 5) the methods and generalization cost. However the availability of a priori is often limited by the complexity of the physical system (Dzitkowski and Dymarek, 2008). Even if the governing physical principles are known, it is sometimes difficult to mathematically formulate the specific relationships to obtain particular parameters that help in choosing appropriate models for developing inference systems. According to the degree to which the a priori is available, the first-principles or the data-driven models, or a hybrid of both models (Czop et al., 2011) can be applied to develop AI methods for disease evaluation.

Theoretically, the biomedical modeling of data can be divided into two categories, data driven and first principle (Little, 2006). The data driven

models, generally termed as statistical machine learning, infer structures in data which can have a meaningful tentative physiological interpretation. However, these models do not reveal direct insights into biological functions. Instead, these models seek for the best features in data to approximate a mathematical relationship between physical principles and measured data. The drawbacks of using this approach for disease evaluation is that, it does not provide a complete visualization of disease symptoms. According to Ottesen et al. (2004), the statistical data analysis may discover correlations between physical principles and selected features but may fail to provide insight into the mechanisms responsible for these correlations. Moreover, the synergy of the mathematical relationship between physical principles and the selected features may likely lead to complex solutions. For these reasons, it can be said that successful symptom evaluation using data-driven methods stems from the codified experience of the human expert being modeled rather than from the deep knowledge of the disease symptoms, which may require volumes of data to generalize the results produced by these methods (Reiter, 1987).

In contrast to data driven models, the first principle models employ physical principles that govern the modeled systems (Reiter, 1987). These models are aimed to discover the underlying mechanisms of the physiological functions of the human body. The symptom evaluation from first principles begins with a description of body function together with an observation of that function's behavior. If this observation conflicts with the way the function is meant to behave, one is confronted with a diagnostic problem i.e., to determine those components that may explain the discrepancy between the observed and correct behavior. This can provide an accurate initial approximation from which inferences can be made to develop clinical DSS (Peng and Reggia, 1986). Importantly, the results of first principle models can be easily interpreted and understood by non-mathematicians such as clinicians due to the fact that the interaction between different body organs can be observed (Reiter, 1987).

Despite of the fact that the first principles models use an understanding of the underlying physics of a biological function to derive its mathematical representation, the development of this representation is expensive since expertise and knowledge at an advanced level is required to derive equations from the physical laws. By contrast, the data driven models use system test data directly to derive its mathematical representation. The advantage in the former is the depth of insight into the system's behavior that supports in correct symptom evaluation, while the advantage of the latter is the speed in which a model can be constructed using the experts' knowledge and experience (Czop et al., 2011). Another advantage of using first principles is that, in terms of simulation in time and space, they provide extrapolation in addition to the interpolation provided by the data-driven models (von Stosch et.

al, 2014; Garud et al., 2013; Wickwar et al., 1999; Adali et al., 1997). This extrapolation can be utilized to optimize clinical DSS through trial and error.

For an accurate and rapid tracking of progressive diseases such as PD, it is important to find a compromise between these two approaches and a combined first-principle data driven (FPDD) model can be more suitable for symptom evaluation. The FPDD model (Czop et al., 2011; von Stosch et. al, 2014; Garud et al., 2013; Wickwar et al., 1999; Adali et al., 1997) utilizes first principles to model parameters that are updated according to the operational data. Besides, the time consumption required for trial and error can be avoided using data driven optimization techniques. On one hand, this can provide a clear interpretation of clinical symptoms. On the other hand, a smaller number of updating first principle parameters can support faster and more accurate convergence of data driven decision making.

1.4 Problem formulation

Chronic neurological disorders are progressive, where symptoms get worse with time. The disease progression cannot be stopped. However, pharmaceutical treatment and surgical intervention can alleviate symptom severities and may help in prolonging the patient's life. Importantly, an accurate prescription requires timely monitoring and quantification of symptom progression. In the case of PD in particular, clinicians have devised a number of methods to quantify symptom severities. In this regard, the most widely used metric is the unified Parkinson's disease rating scale (UPDRS)(Fahn et al., 1987) which reflects the presence and severity of symptoms but does not reveal insight to the underlying causes. At present, PD monitoring has several drawbacks, i.e.

1. It requires frequent visits of patients to the clinic that is expensive in terms of cost and logistics, both for the patients and the clinical staff.
2. It requires the availability of human resources such as expert clinical staff to perform a number of tests to evaluate different symptom dimensions in order to generate an overall UPDRS score.
3. It is expensive for national health systems to accommodate patients and to allocate required clinical resources.
4. The clinical assessment using UPDRS is subjective, i.e. the raters do not agree on symptom scores, resulting in inter-rater variability. (Rajput et al., 1991; Hughes et al., 1993; Ramaker et al., 2002).
5. The assessment using UPDRS is time consuming and normally lasts for more than a couple of hours when assessing PD severity in both the on and off states of medication.

For all these reasons, the UPDRS assessments of patients are performed once in every three to six months. A promising solution is to use remote

monitoring tools that can accurately and efficiently follow disease progression at more frequent intervals with minimal resource utilization. This can mitigate the excessive workload on national health systems, while decreasing the overall treatment cost and increasing the accuracy in clinical evaluation of patient status.

Apart from the problems in monitoring, another important problem in the clinical management of PD is the complexity that arises due to random and abrupt changes in the severity of PD symptoms throughout the day. For this reason, a detailed and timely report of multiple measurements of different PD dimensions is necessary for clinicians to adjust the patient's medication dosage (Weaver et al., 2005). The use of computer tools incorporating AI algorithms can be particularly handy for correctly estimating the PD symptoms, to reveal the full extent of a patient's condition and to avoid bias in measuring the affects of treatment (Isacson et al., 2008). One such example is a test battery system (Westin et al., 2012) developed by a research group at Dalarna University (www.du.se) in collaboration with Malardalen University (www.mdh.se/idt), Abbott Laboratories (http://www.abbott.se/), Animech AB (www.animech.com), Nordforce Technology (www.nordforce.se) and the Swedish Knowledge Foundation (http://www.kks.se/) under the projects 'Evaluation of a motor/non-motor intelligent online system' (E-MOTIONS, 2010-2012) and PAULINA (2013-2015).

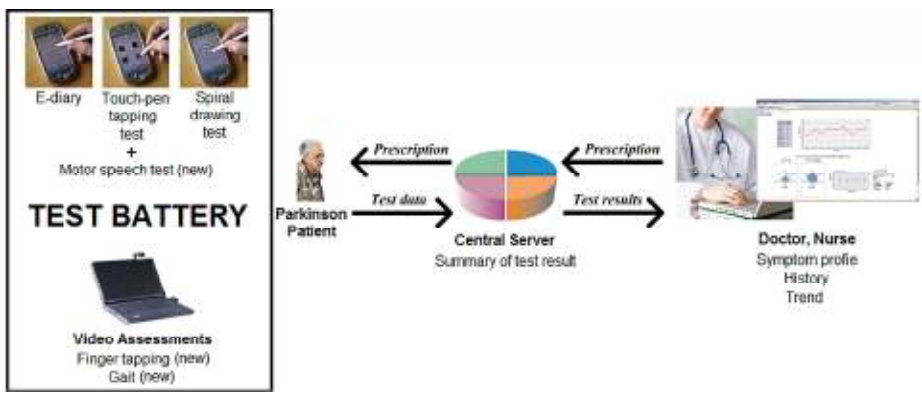


Figure 1: Test battery system. The patient uses the battery to perform motor and non-motor tests on several occasions per day. The test data is transmitted to the server for processing. A web-application provides graphical feedback of the test results to the clinical staff that allows them to monitor and evaluate the patient's profile and suggest an appropriate prescription. The prescription is transmitted back electronically to the patients via the server. The data is saved on the server for future use.

The research presented in this thesis has been done under the projects E-MOTIONS and PAULINA. The goal is to employ mobile technology and AI to develop computerized methods to support clinicians in remote monitoring of PD symptoms, and to support patients in at-home treatment of PD. The system consists of a handheld computer with a touch-screen, for evaluation of self-assessed questionnaires and fine-motor functions in a telemedicine setting (Westin et al., 2012). The device is specifically designed to be used for repeated measurement of symptom fluctuations in the patient's home environment.

Currently the battery employs tapping and spiral drawing test functions that are used to capture upper limb motor abnormalities. Additionally, a web-based system is incorporated in this framework, consisting of a patient node for subjective and objective data collection, a service node for data storage and processing, and a web application for data presentation (figure 1). The system utilizes statistical and machine learning methods to summarize raw symptom data into an overall test score providing a comprehensive profile of the patient's health during a test period of one week. Importantly, the current platform allows processing of recorded speech, as well as processing of recorded videos of gait and finger tapping.

An issue with many of the current telemetry systems used in motor disorder assessment (Haubenberger et al., 2011; Louis et al., 2012; Norman et al., 2013; Ancillao et al., 2013) is that the motor tests are based on data driven approaches; for example tapping and spiral drawing tests used in the test battery system (Westin, 2010). In these tests, the clinical evaluation of data is performed by visual inspection of test results and not by the direct observation of patient's performance. The clinicians involved in inspection required prior training to understand graphical illustrations before scoring the symptoms. The accuracy of subjective scoring is therefore highly dependent on training and experience using these illustrations. Even after that, it is not guaranteed that the computerized estimates would succeed in establishing a relationship between physiological symptoms and clinical scores as these estimates require fine tuning relative to the clinical opinion on physical manifestation of the patient.

The incorporation of speech, gait and finger tapping in the test battery system allows the implementation of first principle models for symptom analysis. For instance, cepstral analysis is a major component of speech processing in this thesis. The cepstrum is based on a source-filter model of speech production (Flanagan et al., 1975) which provides a mathematical framework of physiological interaction between the body organs (from lungs to the lips) to produce sound. Similarly, the gait analysis has been performed by skeletonizing a standard human model (Winter, 1990) using video processing techniques. Likewise, finger tapping has been evaluated using video recordings which allowed extracting physiological patterns of finger movement. Importantly, the first principle inference of speech, gait and fin-

ger tapping allows hands-on assessment of the physical manifestation of patients by the clinicians. The accuracy of patient symptom profiles can thus be greatly improved by incorporating computerized estimates of speech, gait and finger tapping motor exams in the overall test score of the test battery system.

1.5 Aims and objectives

The accurate measurement of symptoms in neurological disorders such as PD requires the estimation of multiple physiological dimensions which is complicated in clinical practice due to the subjectivity of the raters (Olanow et al., 2009). Among different symptom dimensions, for instance, the speech has deficits in multiple sub-dimensions, including problems in respiration, phonation, articulation and prosody (Midi et al., 2008). Besides, an impaired gait manifests gross features such as stooped posture, short shuffling steps and slowness to start walking (Factor, 2008). The fine motor symptoms in finger tapping include tapping arrhythmia, slowness of speed, fatigue and amplitude reduction (Agostino et al., 2003). Most of these symptoms are difficult to be tracked with the 'naked eye' and require clinical experience and expertise for an accurate assessment. The clinical results are largely experienced based and a more experienced clinician likely achieves better results. Albeit the previous clinical research has reported a progressive elevation in the severity of these symptoms over the course of PD (Siddiqui et al., 2002), yet the computerized assessments are necessary for correct estimation of symptom severity and to support in clinical prescription. It also allows transferring of symptom estimates between clinical staff which is important for treatment planning. These objective measurements can also be used as biofeedback for patients, since different patients may have different physical symptoms and methods of treatment can vary between individuals (Olanow et al., 2009).

The main objective of this research is to develop and evaluate unobtrusive computer methods for enabling the remote monitoring of patients with neurological disorders, specifically PD. The study investigates novel signal and image processing methods based on FPDD models for extraction of clinically useful information from audio recordings of motor speech and video recordings of gait and finger-tapping motor examinations. The thesis aims to map between PD symptoms severities estimated using novel computer methods and the standard clinical metric UPDRS part-III motor examination. In the case of speech, the properties of speech signals were inferred to model new acoustic features that can distinguish abnormalities in speech production attributed to the deterioration of neurological control in PD. For finger tapping, the recorded videos of the rapid finger-tapping test (RFT) were processed using a novel CV method to extract symptom information from video based tapping signals. In case of gait analysis, the investigation was

performed on novel CV based gait features to discriminate between a healthy and an impaired gait.

Based on the aims and objective of this work, the main research questions (RQ) that this research tries to explore are:

RQ-1: What computer methods can be used to quantify motor speech, gross motor (gait) and fine motor (finger-tapping) symptoms in PD, and what are their limitations?

RQ-2: How can we develop novel computer methods that address the limitations identified in RQ-1 by feature extraction based on first principle models, to quantify motor speech, gross motor (gait) and fine motor (finger tapping) symptoms in PD, and compare the features to a standardized rating scale?

RQ-3: How can we develop and evaluate computer systems that allow automatic classification of PD symptom severities using first-principle data-driven models and unobtrusive data acquisition techniques?

1.6 Research contributions

The research presented in this thesis contributes to a number of computer science areas for example AI in medicine, clinical DSS, computer speech processing, image signal processing, and in general medical cybernetics. A brief description of contributions is presented in the appended papers. The interconnection between the papers, RQs and contributions is given in table 1. The following are the specific research contributions (RC) presented in this dissertation:

RC1: Literature reviews were conducted to explore and analyze computer applications that are available for evaluating motor speech (section 4.1.1 [Paper 1]), gross motor (gait: section 4.2.3 [Paper 6]) and fine-motor (finger-tapping: section 4.2.1) symptoms in PD (RC to RQ-1).

RC2: A comparative analysis was done using mel-frequency cepstral coefficients (MFCC) computed from recordings of text-dependent running speech (TRS) and other standard vocal examinations including sustained vowel phonation (SVP) and diadochokinesis (DDK) tests (section 4.1.3 [Paper 2]), to show that the use of TRS is the most feasible for assessment of UPDRS motor speech examination (UPDRS-S) (RC to RQ-1 and RQ-3).

RC3: A novel speech processing algorithm called the cepstral separation difference (CSD) is introduced (section 4.1.4 [Paper 3]) to quantify PD speech symptom severities according to the UPDRS-S. The CSD features

rely on the first principles of speech production since they are extracted using the source-filter speech model that characterizes the mechanical properties of vocal-folds and the vocal-tract. Importantly, the CSD utilizes TRS to estimate speech symptoms, which is similar to how speech symptoms are identified using UPDRS-S (RC to RQ-2).

RC4: A system for classification of speech symptom severities according to UPDRS-S is proposed in section 4.1.5 [Paper 4]. This classification scheme involves support vector machines (SVM) and TRS features for classification between the symptom severity levels. The novel CSD and MFCC features, which are based on the first principles of speech production, were utilized to train the SVM for data-driven classification. Besides, the recording of speech using a microphone allows unobtrusive acquisition of patient's data (RC to RQ-3).

RC5: A system for classifying symptom severity according to UPDRS motor examination of finger-tapping (UPDRS-FT) is proposed in paper 5. In this scheme, video recordings of RFT were utilized, which allowed unobtrusive analysis of fine motor symptoms in PD (RC to RQ-3). The patients performed RFT according to the standard clinical procedure and no additional setting was devised for video recordings. The first-principle features computed from the motion analysis of index fingers were able to compute standard UPDRS tapping symptoms including arrhythmia, fatigue, slowed pace and reduction in amplitude (RC to RQ-2).

RC6: A novel tool for Parkinsonian gait and posture analysis is proposed in Paper 7. This CV algorithm processes recorded videos of gait to compute stride variability and posture lean to discriminate between normal and Parkinsonian gait. The proposed features are based on the first principles of human physiology, the utilization of which would allow clinicians to monitor gait symptoms with good accuracy (RC to RQ-2).

1.7 Outline of the thesis

This dissertation is organized as follows. Chapter 1 presents an introduction to the thesis and outlines the goals of this study, the main RQs and the RCs. Chapter 2 discusses human physiology, the effects of PD on motor functions, and a description about how different symptoms are evaluated and treated clinically. Chapter 3 presents methods and approaches utilized to computerize the clinical evaluation of motor functions in PD. Chapter 4 provides a detailed description of computer methods proposed for the quantification of motor functions. Chapter 5 discusses results and analysis. Chapter 6 discusses the significance of this research, along with the conclusions and suggestions for future research.

Table 1: Research questions and corresponding contributions

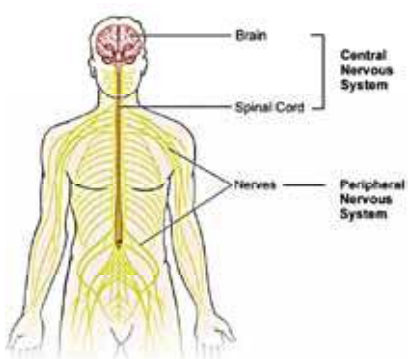
Research questions	Research contributions	Appended paper reference
RQ-1: What computer methods can be used to quantify motor speech, gross motor (gait) and fine motor (finger-tapping) symptoms in PD, and what are their limitations?	Literature reviews and conferences with domain experts were done to understand the application domain, and to determine and analyze suitable computer methods that can be used to quantify gross motor, fine motor and motor speech symptoms in PD.	**Paper 1**: Methods for detection of speech impairment using mobile devices. Taha Khan, Jerker Westin, *Recent Patents on Signal Processing,* 1 (2), 163-171, 2011. **Paper 6:** Computer vision methods for Parkinsonian gait analysis: a review on patents. Taha Khan, Peter Grenholm and Dag Nyholm. *Recent Patents on Biomedical Engineering* 6 (2), 97-108, 2013.
RQ-2: How can we develop novel computer methods that address the limitations identified in RQ-1 by feature extraction based on first principle models, to quantify motor speech, gross motor (gait) and fine motor (finger tapping) symptoms in PD, and compare the features to a standardized rating scale?	New methods employing feature extraction based on first principle models were developed to quantify motor speech, fine motor (finger-tapping) and gross motor (gait) symptoms in PD. The features were able to quantify PD symptoms according to the clinical ratings based on the UPDRS.	**Paper 3:** Cepstral separation difference: a novel approach for speech impairment quantification in Parkinson's disease. Taha Khan, Jerker Westin, Mark Dougherty. *Biocybernetics and Biomedical Engineering,* Elsevier 34(1), 25-34, 2014. **Paper 5:** A computer vision framework for evaluation of finger tapping in Parkinson's disease. Taha Khan, Dag Nyholm, Jerker Westin, Mark Dougherty. *Artificial Intelligence in Medicine,* Elsevier 60(1), 27-40, 2014. **Paper 7:** Motion cue analysis for Parkinsonian gait recognition. Taha Khan, Jerker Westin, Mark Dougherty. *The Open Biomedical Engineering Journal,* 7(1), 1-8, 2013.
RQ3: How can we develop and evaluate computer systems that allow automatic classification of PD symptom severities using first-principle data-driven models and unobtrusive data acquisition techniques?	Computer systems based on FPDD models were developed that were able to classify symptom severity according to the UPDRS. Audio recordings of speech and video recordings of finger-tapping and gait allow unobtrusive data acquisition.	**Paper 2:** Running-speech MFCC are better markers of Parkinsonian speech deficits than vowel phonation and diadochokinetic. Taha Khan. **Paper 4:** Classification of speech intelligibility in Parkinson's disease. Taha Khan, Jerker Westin, Mark Dougherty. *Biocybernetics and Biomedical Engineering,* Elsevier 34(1), 35-45, 2014. **Paper 5:** A computer vision framework for evaluation of finger tapping in Parkinson's disease. Taha Khan, Dag Nyholm, Jerker Westin, Mark Dougherty. *Artificial Intelligence in Medicine,* Elsevier 60(1), 27-40, 2014.

2. Human physiology and Parkinson's disease

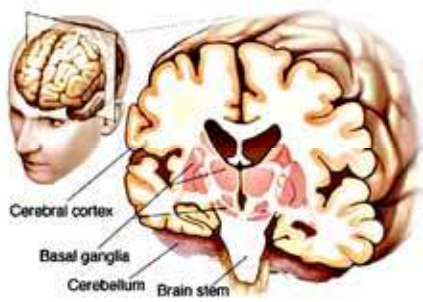
This chapter presents a brief description of the human physiological system with respect to neurological deficits caused by PD. The functionality of nervous system has been discussed in relation to the ability to produce speech and to perform gross and fine motor tasks. Further, the standard clinical procedures used to evaluate PD-related motor dysfunctions have been addressed.

2.1. The nervous system

The nervous system consists of a sophisticated web of dedicated cells (neurons) whose main function is to coordinate voluntary and involuntary body movements by transmitting signals between different parts of the body (Nebylitsyn, 1972). The system is responsible for processing sensory inputs, coordinating actions to achieve a desired movement and processing all other cognitive functions. Theoretically, the nervous system can be divided into two parts, the central nervous system and the peripheral nervous system (figure 2a). The central nervous system consists of the brain and the spinal cord. The rest of the neuronal circuitry belongs to the peripheral nervous system consisting of nerves that connect the central nervous system to every other part of the body.



a. The nervous system.



b. Position of the basal ganglia in the human brain.

Figure 2: The human nervous system (Mayo clinic, 2013)

At the cellular level, the nervous system is comprised of two specialized cells, the neurons and the glia. The neurons have special structures that allow them to transmit signals rapidly and precisely to the other cells in the body. These signals are transmitted in the form of electrochemical waves travelling along fibers called axons. The axons lead to the nerve terminals and the dendrites which receive signals from other neurons using junctions called synapses. A cell that receives a synaptic signal from other neurons may become excited, inhibited or modulated. Besides, the function of glia is to protect, assist and support the electrochemical communication between the neurons.

The nervous system is susceptible to impairment in many ways. This can be as a result of physical damage due to trauma, infection, genetic defects or simply aging (Organisation mondiale de la santé, 2006). In the peripheral nervous system the most common disorder is the failure of nerve conduction that causes diabetic neuropathy, multiple sclerosis, lateral sclerosis etc. Besides, in the central nervous system, problems can be caused by malfunctioning of basal ganglia that may further lead to a wide range of disorders including PD, Huntington's disease and schizophrenia.

The basal ganglia consist of highly interconnected anatomical structures located near the mid-brain (Stocco et al., 2010) including striatum, globus pallidus, subthalamic nucleus and substantia nigra. The ganglia receive input from the cerebral cortex and project their output back to the cortex and the brainstem (figure 2b). Although small, the brainstem is an extremely important part of the brain as it connects the motor and sensory nerves between the central nervous system and the rest of the body. These ganglion interconnections are either inhibitory or excitatory and are controlled by the neurotransmitter dopamine, a substance produced by the dopaminergic cells that are affected in PD.

2.2. Parkinson's disease symptoms and their assessment

The underlying cause of PD is currently unknown (Lang and Lozano, 1998). However, research supports that PD symptoms are caused by a substantial reduction in the amount of dopaminergic neurons in the basal ganglia. The dopaminergic neurons support the transmission of information between neurons to control physiological functions. Their declination causes malfunctions in the central nervous system that deteriorates coordination and motor abilities. With the disease evolution, the progressive dopaminergic loss results in an increase in the severity of PD symptoms.

The main symptoms are muscular rigidity, tremor and problems with body movement. Apart from these cardinal symptoms, research suggests that impairment in speech is one of the early onset indicators of PD (Hanson et al., 1984; Logemann, 1978; Hartelius and Svensson, 1994). According to Ho

et al. (1999), approximately 90% of patients with PD develop speech symptoms in the early stages of disease. Moreover, 29% of the patients consider vocal impairment as the worst symptom associated with the disease (Hartelius and Svensson, 1994).

The clinical management of PD involves administration of a number of physical tests to assess the motor and cognitive abilities of patients. These tests are devised in a way that enables clinicians to quantify and track symptom severity in relation to the disease progression. In these tests, the severity of each symptom can be scored quantitatively using standard clinical rating scales. Importantly, the use of rating scales provides a comprehensive picture of the disease that supports accurate administration of treatment. A common rating scale used to evaluate symptom progression is the Hoehn and Yahr scale (Hoehn and Yahr, 1998), which is a five-point scale that describes impairment in PD. However there are weaknesses in this scale, including the mixing of different symptoms, having a non-linear severity estimate, a lack of information delivery on non-motor problems and a superfluous emphasis on postural instability over other motor symptoms (Goetz et al., 2004). For these reasons, the Hoehn and Yahr scale has been largely supplanted by the UPDRS. The UPDRS is a multi-modular scale widely used in scoring PD motor impairment in clinical trials (Mitchell et al., 2000; Goetz et al., 2003). Presently, the UPDRS is increasingly used as a gold standard reference scale (Ramaker et al., 2002) as approved by the movement disorder society (MDS). The UPDRS has lately been revised as the MDS-UPDRS scale (Goetz et al., 2008). The revision contemplates some deficiencies of the current version which were previously publicized by Goetz et al. (2003). Nevertheless, in the study performed by Ramaker et al. (2002), it was found that the most complete picture of the status of PD symptoms can be achieved using the UPDRS. The UPDRS provides overall better reliability and validity as compared to the other clinical scales.

The UPDRS metric is made of 44 sections divided in three major parts. Each section addresses different symptoms in different body parts and spans the range 0-4, with 0 representing no symptoms and 4 representing severe impairment. The three major parts in the UPDRS cover (I) mentation, behavior and mood, (II) activities of daily living, and (III) motor examination. Parts I and II are assessed by interviewing the patient whereas part III is assessed by physical examination. Summing up the scores in 44 sections generates the total UPDRS-score, representing the global severity of impairment. The total UPDRS-score spans the range 0-176, with 0 representing a perfectly healthy individual and 176 total disability.

In this study we have dealt with UPDRS (part III) motor examination that is comprised of sections 18-31 and ranges from 0-108, with 0 representing ‘normal’ and ‘108’ representing severe motor impairment. The motor examination encompasses tasks such as speech, facial expression, finger tapping, gait and posture etc. This part of the UPDRS contributes most of the

points in the UPDRS scale. For this reason, many studies focus exclusively on the motor examination because motor symptoms are the most prominent aspect of PD.

A major disadvantage of using UPDRS and other rating scales is the subjectivity of clinical opinion, i.e. the scores are dependent not only on the subject's performance on the motor test but also on the clinician's interpretation (Hagell, 2000). A study by Shulman et al. (2006) depicted that the subjective ratings on a patient's performance may not correlate well with the objective measurements of a patient's performance. However, a good basis for PD assessment can be formed by combining computerized estimates and subjective assessment using rating scales such as the UPDRS (Hagell, 2000).

2.3. Speech production system

This subsection presents a brief overview of concepts and definitions associated with the mechanism of speech production. Before proceeding with the discussion of the organs involved in producing speech, it is important to clarify some terminology used to refer to the speech model. Two terms are generally used: speech and voice. Sometimes they are used interchangeably, and as Titze (2000) affirmed: "*in the broader sense voice is synonymous with speech*". However, he added that there is a slight difference between voice and speech, i.e. voice refers to the sound produced by the vocal organs with less cognition than when formulating speech. Since the running-speech is the basic ingredient for speech processing in this work, the term speech is used to describe anatomical parts involved in speech production. Additionally, some physiological aspects related to pathology and their influence on the speech process has been discussed. The following sub-sections present a brief overview of the anatomical components involved in speech production (figure 3).

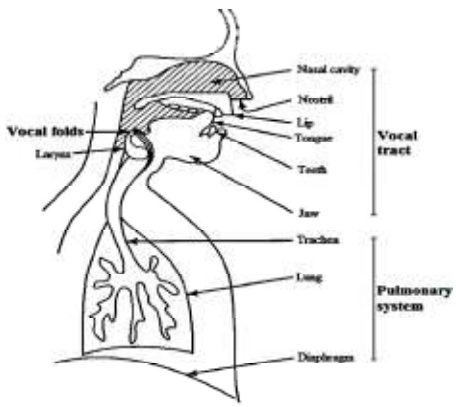


Figure 3: Speech production system. A general scheme of the speech production system and location of its subsystems in the human body is shown.

2.3.1. Pulmonary system

The pulmonary system serves as a source of energy for speech production (Titze, 2000). It is comprised of the lungs and respiratory pathways (trachea) which allow the intake and expulsion of air during respiration. Speech production begins with inhalation of air, where the volume of the lungs expands. The air flows into the lungs and fills this space. During exhalation, the lungs collapse and allow the air to expel through the trachea to the larynx. To produce a voiced sound, the flow and air pressure on exhalation must be sufficient to separate the vocal folds causing the vocal fold membrane to vibrate (Godino-Llorente et al., 2003). According to Hixon et al. (2008), this muscular coordination between the pulmonary organs serves to control speech intensity (loudness), linguistic stress (emphasis), frequency (pitch) and segmentation of speech into units (syllables, words, phrases). With direct relevance to this study, the respiratory muscle control is known to be affected in PD (Apps et al., 1985), which explains why the patients fail to comply with the demanding linguistic stress levels when reading long and difficult paragraphs (section 4.1.5.1 and section 5.3.1).

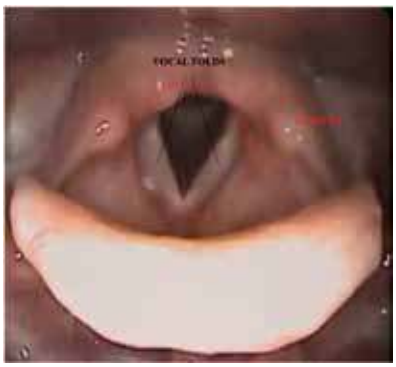
2.3.2. Vocal folds

The vocal folds are two elastic bands of muscle tissue located just above the trachea in the larynx. They serve as a self-sustained oscillator that modulates the airflow from the lungs as it travels through the glottis (the space between the vocal folds) during the opening and closing phases of phonation (Titze, 2000). Because of their elasticity they are capable of generating a wide range of frequencies by using the air that passes through the glottis. In the closing phases of phonation, they move to the midline to decrease the glottal space to close the laryngeal airway (Hixon et al., 2008). While during the opening phases, the vocal folds move away from the midline, allowing a free passage to the air. These movements are symmetric and simultaneous on both vocal folds.

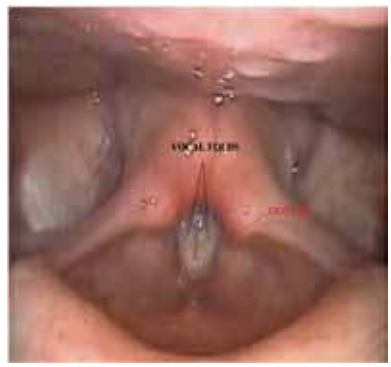
The number of times the vocal folds vibrate during a second is called frequency (in Hertz) of oscillation (Titze, 2000). Similarly, the time taken for the vocal folds to complete one oscillation is called the pitch period (T0). The fundamental frequency (F0) of voice can be calculated as $F0=1/T0$. In this way, the time varying movement of vocal folds can be described in the frequency domain, where frequencies that belong to pathological speech can be identified.

According to Guyton and Hall (2006), there is a direct connection between vocal fold muscles and the brainstem connecting to the basal ganglia through a nerve called vagus. Research has shown that symptoms in PD results in unintended glottal constrictions on phonation (figure 4) which hinders the airflow and creates air-turbulence at the glottis (Ho et al., 1999). As a consequence, a normal voice turns into a harsh-strangled voice which

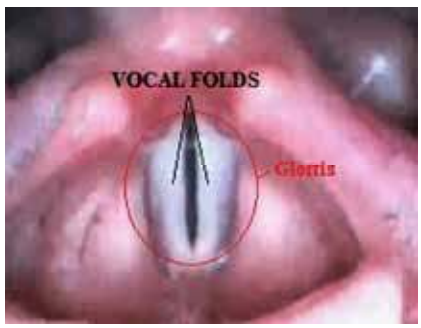
has a perceptual impression of breathiness due to an audible escape of air. The noise in phonation generated as a result of this turbulent air is termed as the aspiration noise (Murdoch, 1998). The turbulent air remains throughout the entire oscillation cycle rather than just after the moment of vocal fold closure, thus causing an increased amount of aspiration.



a. Normal glottal opening



b. Normal glottal closure



c. Glottal constriction

Figure 4: Laryngoscopic pictures of normal opening and closing phases of the glottis are shown in figure 4a and 4b respectively. A picture of glottal constriction in PD speech is shown in figure 4c.

The study of the vibration of vocal fold patterns includes the evaluation of biomechanical parameters that are susceptible to malfunction in the presence of pathology. It should also be acknowledged that the glottal constriction and aspiration are not necessarily caused by neurological deficits. They could be the result of vocal fold tissue problems such as nodules; or can occur as symptoms of more serious problems such as vocal fold cancer (Dubuisson, 2011). This suggests that the computer algorithms that can estimate glottal constrictions in PD speech can also be used to diagnose chronic diseases such as cancer.

2.3.3. Vocal Tract

The vocal tract consists of nose, mouth (oral cavity), tongue and lips, located at the top of the larynx. These components are involved in determining the phonetic qualities of speech sounds (Clark et al., 2007). If the vocal folds can be termed as an oscillator, then the vocal tract can be called a resonator that amplifies certain acoustic frequencies depending on its shape while attenuating the others.

The source-filter speech model provides a mathematical framework to formulate the coupling between vocal-fold and vocal-tract characteristics (Titze, 2008); where the vocal folds are considered as the sound source and the vocal tract is considered as the filter. The research on speech (Titze, 2008; Fant, 1960) assumes that the relationship between source and filter is linear. However, recent research challenges the linear assumption of source-filter theory claiming that the collision between the vocal-folds in the closing phases of phonation contains nonlinear effects produced by the amount of retro-reflected waves generated in the vocal tract and propagated back to the vocal-folds. These waves deform the delicate vocal-fold membrane while contacting in the midline. Besides, no explicit formulation of such mechanism has been made yet and it is an open area of research. Some researchers have used pseudo equations to model the behavior of the collision of the vocal folds, representing nonlinear elastic components related to the tissue deformations (Avanzini, 2008; LaMar et al., 2003).

Albeit previous research on speech has reported a correlation between anomalies in vocal tract organs and PD (Hanson et al., 1984; Logemann, 1978), this evidence is rather limited compared to the investigation of PD effects on the vocal folds. Besides, current literature (Tsanas, 2012; Rusz et al., 2011) agrees that the speech symptoms involving short rushes of speech and articulation blurring arise as a consequence of disturbed coordination between vocal tract organs that is caused by PD. These symptoms can be identified using MFCC (section 3.1.1.2) that are aimed at detecting subtle changes in the motion of vocal tract articulators that interfere with speech intelligibility. In paper 4 (section 4.1.5) we have shown that important clinically relevant information can be extracted from recorded TRS signals by analyzing vocal tract related signal characteristics.

2.3.4 Motor examination of speech

The intelligibility of speech can be disturbed by a number of PD symptoms (Harel et al., 2004; Holmes et al., 2000). Pinto et al. (2004) identified the relationship between anatomical substrates of speech components and PD. According to them, vocal impairment in PD is associated with pathological changes to mainly three components of speech including respiration, phonation and articulation, attributed to the dysfunction of muscles at the subglottis (pulmonary system), glottis (larynx and vocal folds) and supraglottis

(vocal tract organs) levels respectively. The collective dysfunction in these speech components give rise to the dysfunction in the fourth speech component called prosody. Previous research has reported progressive deterioration of speech symptoms over the course of the disease (Harel et al., 2004; Holmes et al., 2000). Some investigations have supported that speech degradation and general PD symptom severity are strongly interlinked (Skodda et al. 2009).

The symptoms in PD speech can be manifested in the speech quality of patients in the form of distorted vowels and consonants, harsh and hoarse voice quality, hypernasility, reduced speaking rate, monoloudness and monopitch (Pinto et al., 2004). Previous methods of estimating vocal impairment typically utilized SVP tests (Tsanas, 2010-12; Gelzinis et al., 2008), where the speaker is asked to sustain phonation /a/ for as long as possible with steady frequency and amplitude. Admittedly, the SVP can provide symptom estimates related to the vocal fold vibration, however a clear picture of impairment relevant to the vocal tract components cannot be captured (Krom, 1995). The laryngeal DDK is another test used to capture vocal tract anomalies. In this test, the patient is requested to utter /pa/-/ta/-/ka/ syllables repeatedly as long as possible. Although the SVP and DDK can provide symptom estimates for tracking voice pathology, they are artificial in the context of natural communication. In both these methods, the dynamic aspects of continuous speech such as onset-offset effects, co-articulation etc. are not present.

Besides from the perspective of speech quality, the TRS contains much more impairment related information than SVP and DDK as it reflects dynamic aspects of continuous speech and is a better representative of communication (Klingholtz, 1990). In the recent computer methods to evaluate PD speech impairment (De Looze et al., 2012; Rusz et al., 2011), the TRS is analyzed to demonstrate deficits in motor speech, suggesting that PD can affect all different subsystems of speech including respiration, phonation, articulation and prosody.

The speech item utility is section 18 in part III (motor examination) of the UPDRS (Fahn et al., 1987). It was previously examined by Zraick et al. (2003). According to them, a standard speech protocol to identify speech symptom severity should include reading of an unfamiliar paragraph. This paragraph should contain different linguistic structures and a description to assess the reading ability of subject. When this protocol was utilized as part of the speech examination, a strong inter-rater reliability coefficient was produced between symptom severity ratings performed separately by the medical (neurologists) and the non-medical (speech pathologists) experts. It was concluded that TRS with standard formulation has the capability to exploit capacious symptoms in PD speech, providing a broader perspective of evaluation.

In the standard clinical method to assess speech using the UPDRS motor examination (Fahn et al., 1987), the patient is given a standard paragraph to read. The clinician listens to the recitation and rates the reading performance between the symptom severity scale of '0' to '4', where

'0' = normal speech.
'1' = slight loss of expression, diction and/or volume.
'2' = monotone, slurred but understandable.
'3' = marked impairment, difficult to understand.
'4' = un-intelligible.

It is important to observe that many of the characteristics of impaired speech mentioned in the UPDRS speech utility cannot be captured using both SVP and DDK tests. For instance 'loss of expression' can only be captured by monitoring continuous speech. Likewise, 'understandability' refers to the ability of the clinician to perceive information from the formulation of sentences and words, which is only possible if the patient produces continuous speech. This infers that, for an accurate computation of speech symptoms, the utilization and processing of TRS is important to quantify underlying deficits in speech that as a whole disturb speech intelligibility.

Readers may find details about previous computer methods used to identify pathological speech in the review study presented in paper 1. A comparative analysis on the performance of TRS, SVP and DDK MFCC in identifying PD vocal symptoms is presented in paper 2. A new computer method to quantify PD-related vocal fold anomalies is introduced in paper 3. A computer algorithm for classification of speech intelligibility between UPDRS-S severity levels is presented in paper 4.

2.4 Quantitative assessment of fine motor impairment

The reduced muscular movement in neurological disorders affects fine motor tasks such as buttoning, using utensils, playing a piano etc. All these examples employ movements that have to be executed as integral actions requiring strong coordination between different motor organs. If any complexity is added to a simple movement, for instance, by repeating the movement or by combining the movement with another task, the condition of the patient may become worse (Berardelli et al., 2001). Rapid sequential movements involving hand or finger tapping often become more hypokinetic (slowed) with repeated movements. Some patients describe the condition of hypokinesia as feeling as if, the "brain is sending message to the body, but the body won't listen" (Koop, 2011).

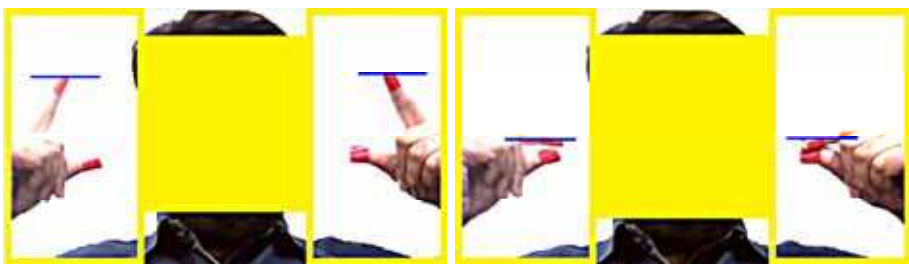
The utilization of tapping tests, to evaluate accessory muscular control and coordination, dates back to the 19th century. Hollingworth (1914) used tapping tests that incorporated an electric counter to characterize the influ-

ence of menstruation in female subjects. Since then, tapping tests have been utilized to assess acute strokes (Heller et al., 1987), the testing of subjects with Korsakoff's alcoholic syndrome (Welch et al., 1997), the characterization of upper limb motor functions (Giovannoni et al., 1999), and in particular to quantify neurological disorders such as Alzheimer's disease (Ott et al., 1995), PD (Volkow et al., 1998) and ataxia (Notermans et al., 1994).

In the case of PD, Volkow et al. (1998) found a strong correlation between the lack of dopamine receptors and motor task ability characterized by finger tapping. However, the quantification of tapping symptoms in clinical practice is difficult (Rao et al,. 2003) due to the fact that the movement of a tapping finger is evaluated visually which results in a coarse resolution to determine the PD status (Muir et al., 1995; Jobbagy et al., 1997). On the other hand, research supports that stable tapping is unaffected by cognitive and perceptual demands and clinical impairments are readily apparent in tapping signal patterns (Collyer et al., 1994). For assessment of PD in particular, finger tapping is widely applied in clinical settings as the rhythm of the dominant hand finger movements acts as an efficient index for evaluating brain motor function (Shimoyama et al., 1990).

2.4.1 Motor examination of finger-tapping

The motor examination of finger-tapping is section 23 in part III of the UPDRS (Fahn et al., 1987), abbreviated as UPDRS-FT. In a clinical setting to examine UPDRS-FT, the patients are seated on a chair with a plain wall directly behind them (figure 5). Their hands are positioned above the shoulders beside the face. While keeping this position, they are requested to tap their index-finger against the thumb repeatedly and as fast as possible. The index-finger should be fully stretched vertically against the thumb while tapping. In medical terms, this type of tapping examination is known as the rapid finger tapping test.



a. Opening phase of tapping b. Closing phase of tapping

Figure 5: Clinical examination of finger tapping.

During the RFT, the clinician considers multiple aspects of tapping motion including speed, amplitude, fatigue and rhythm. They combine these aspects into a single score using the UPDRS-FT severity scale ranging between 0 and 4, where:

'0' =	normal
'1' =	mild slowing (in pace) and /or reduction in amplitude
'2' =	moderately impaired, definite and early fatiguing, may have occasional arrests in movement (i.e. mild arrhythmia)
'3' =	severely impaired, frequent hesitation in initiating movements or arrests in ongoing movement (i.e. severe arrhythmia)
'4' =	unable to perform tapping

Previous research suggests that clinicians rate the UPDRS-FT severity levels differently, thus creating a considerable degree of inter-rater variability (Goetz et al., 2007). Moreover, the symptom evaluation is subjective and may take several hours to conduct the examination, which may also reduce the ability of clinicians to detect subtle pathological changes. Furthermore, the accuracy of symptom evaluation may worsen if the patients have dyskinesias or chorea (jerky movements) which may divert the attention of clinicians from focusing precisely on the finger movements. For such reasons, the objective measurements of tapping are important to support clinicians in symptom evaluation. A CV based algorithm that is able to solve these clinical problems as well as assist clinicians in tracking fine motor tapping symptoms in PD is presented in paper 5.

2.5 Quantitative assessment of gross motor impairment

Gross motor impairment in neurological disorders is typically assessed using gait analysis (Factor, 2008). In the case of PD, gait symptoms develop over time as a result of progressive PD features including bradykinesia, postural abnormality and muscular rigidity (Nieuwboer, 2006). For sake of demonstration, two actors performing normal and Parkinsonian gait are shown in figure 6.

A typical PD patient stands with a bent posture with flexed knees and hip (figure 6). As he tries to walk, his arms fail to swing and trunk rotation becomes restricted. Sometimes, he stops and stands still with a rigid face. He walks with short angular movements at all the joints and his feet shuffle due to his reduced stride length and step height. His trunk bends forward as the body weight is carried out onto his toes. As a result, the centre of gravity (COG) of the body is pushed ahead of his feet which creates a risk of falling (Bloem et al., 2001). He also finds trouble in gait initiation and due to psychological stress he becomes 'frozen' (Nieuwboer, 2006).

In an observational gait analysis by a clinician, approximations are made by comparing abnormal and normal gait patterns. Besides, computerized algorithms can help a clinician produce more precise and quantifiable measurements of the gait variables. Among different spatial and temporal gait variables, the easiest variable to compute is the walking speed (meter/sec), which is useful due to the fact that the alignment and motion of body joints are strongly affected with the increased levels of walking velocity (Pasquina and Cooper, 2009). Other parameters that can be used in gait characterization are cadence, stride length and step length. Step length is referred to as the distance between heels of the person during the double support phase of gait (figure 7). Stride length is the distance covered by the heel from the initial contact point of one foot to the subsequent contact point of the same foot. One stride length is equal to two step lengths. Cadence is the number of steps within a given duration of time that can be multiplied by step length to determine walking velocity. A single sequence of functions of one limb that helps the human body to move forward is called a gait cycle. The gait cycle can be further divided between phases of swing and stance. During the swing phase, the limb is moved forward, while in the stance phase, the limb supports the body weight.

a. Normal gait b. Parkinsonian gait

Figure 6: A comparison between normal and Parkinsonian gait (performed by two actors) is shown. A normal person walks with swinging hands, erect posture and wide stride length. By contrast, a patient with PD walks with bent posture, rigid hands, masked face and reduced stride length.

From the point of view of using a telemonitoring system for gait analysis, CV algorithms can be particularly useful since they can provide quantitative data on postural and locomotive abnormalities using image processing techniques on gait video images (Green et al., 2000). These videos can be recorded ‘at home’ and can be transmitted to a centralized server via the inter-

net, where clinicians and experts can visualize a subject's gait status and rate the severity of the subject's symptoms. Importantly, the visualization of real life gait datasets, accompanied by the computer estimates of symptom severity, can help clinicians to improve the symptom assessment.

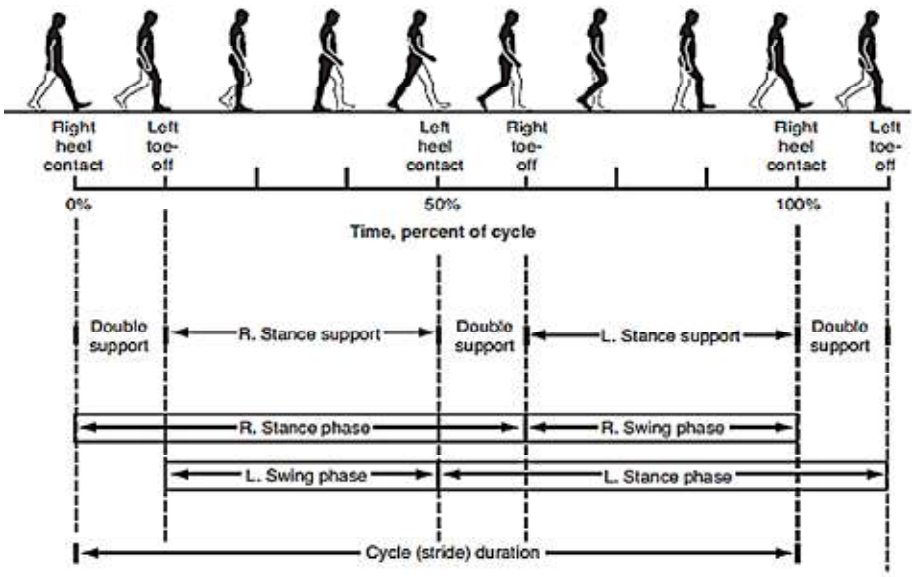


Figure 7: Time dimensions of a gait cycle (Nieuwboer, 2006).

2.5.1 Motor examination of gait and posture

The motor examination of posture and gait are item 28 and 29 respectively in part III of the UPDRS (Fahn et al., 1987). In the motor gait examination, the clinician asks the patient to walk over the gait platform. Based upon the gait performance of the patient, the clinician rates the symptom severity between the scale '0' and '4', where:

'0' = normal gait
'1' = walks slowly, may shuffle with short steps but no festination (hastening steps) or propulsion
'2' = walks with difficulty, may have festination, short steps or propulsion
'3' = severe disturbance of gait
'4' = cannot walk at all

At the same time, the clinician examines the motor posture control of the patient and rates the symptom severity between the scale '0' and '4', where:

‘0’ = normal erect
‘1’ = not quiet erect, slightly stooped posture
‘2’ = moderately stooped posture, definitely abnormal
‘3’ = severely stooped posture
‘4’ = marked flexion with extreme abnormality of posture

Readers may find a detailed systematic literature review on gait analysis from the perspective of ‘at-home’ assessment and treatment of PD in paper 6. Currently available computer methods for pathological gait assessment were investigated and their limitations in terms of their feasibility to be utilized in a home environment were discussed. In order to cope with these limitations, a CV prototype for the assessment of gait impairment is proposed in paper 7.

3. Methods and approaches

For an accurate and rapid assessment of progressive diseases such as PD, computer methods involving first-principle features and data-driven classification approaches can be particularly useful. On one hand, the first principle features can provide a clear interpretation of symptoms to the clinicians. On the other hand, machine learning classifiers can be trained using these features for faster convergence of results and to automate clinical decision making.

In this work, different first principle models were investigated and utilized to process clinical speech, finger-tapping and gait. The features were computed using the first principle models and were used to train an SVM for classification of symptom severity. In order to optimize the SVM convergence, a sequential minimum optimization (SMO) algorithm was used. Additionally, the Guttmann correlation model was emphasized for feature selection and analysis. Further, the consistency and reproducibility of clinical features were assessed using the intra-class correlation coefficient and analysis of variance.

3.1 First principle models for clinical assessment

3.1.1 First principle model for speech assessment

According to Titze (2000), the first principle in speech production is to establish that speech sounds are flow induced, i.e. the airstream generated by the lungs (section 2.3.1) flows through the glottis to the vocal tract, where it is modulated by the vibrating vocal folds (section 2.3.2) and filtered by the vocal tract resonators (section 2.3.3). The resulting pressure at the lips or nostrils, as a function of time, is in the form of a speech sound wave. This mechanical system (figure 8) can be modeled using the source-filter theory of speech production (first developed by Gunnar Fant, 1960) that provides a computational framework of physiological interaction between the vocal organs to produce sound.

According to the source filter model, the source of acoustic energy is at the larynx, and the modulation of this energy by the vocal folds generates a source excitation signal that is filtered by a spectral envelope of vocal tract resonances to form a speech signal. Both the source and filter signals have independent spectral properties in a speech signal. For instance, the source

signal depicts a speech component representing vocal fold vibration where the amplitude of the source signal corresponds to the pressure of airflow at the glottis. On the other hand, a filter signal represents the characteristics of vocal tract resonances that are regulated by the movement of articulators such as tongue, jaw, lips etc.

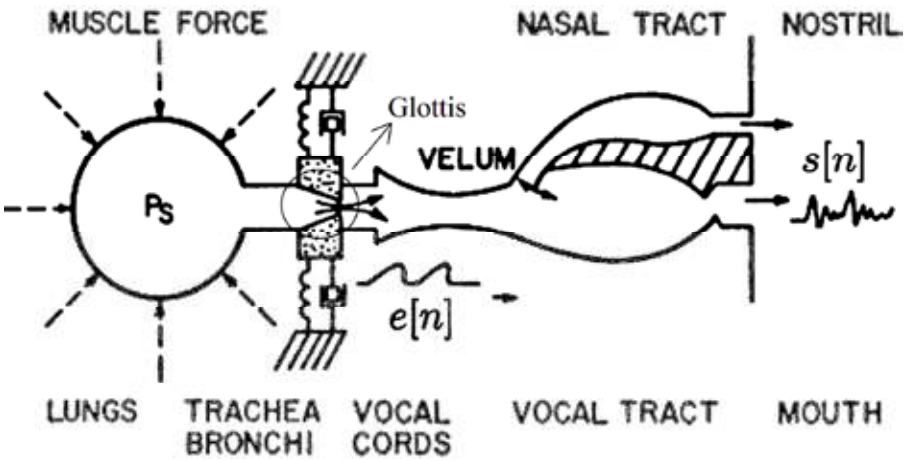


Figure 8: Mechanical speech production (Fant, 1960).

3.1.1.1 Cepstral homomorphic filtering

It is important to split the source and filter components from speech signal to be able to assess symptoms in the vocal-fold and the vocal-tract functions in speech. Mathematically, the source-filter model assumes that speech signal s [i] is a convolution between the source signal e [i] and the filter signal h [i] (Flanagan et al., 1975). This convolution can be resolved using cepstral homomorphic filtering on speech signal s [i] to split it into two independent spectrums of e[i] and h[i].

In the first step, the application of the discrete Fourier transform (DFT) over speech signal s[*i*] which is sampled in the time domain transforms the convolution between e[*i*] and h[*i*] into a multiplication in the frequency domain. Taking the logarithm of speech frequency further transforms this multiplication into a simple linear addition in the log-power domain. Let ω be the log-power coefficient of speech log-spectrum S [ω]. The speech log-spectrum S [ω] can now be separated by applying the inverse discrete Fourier transform (IDFT) on linearly combined source and filter log-spectrums represented as E [ω] and H [ω] respectively (equation 1).

$$IDFT\left(\log\left|S[\omega]\right|\right) = IDFT\left(\log\left|E[\omega]\right|\right) + IDFT\left(\log\left|H[\omega]\right|\right) \quad (1)$$

The application of IDFT on the log-spectrum S [ω] transforms the speech waveform into the quefrency domain where the IDFT of S [ω] log-spectrum is termed cepstrum c[n] of cepstral coefficients n. In the quefrency domain, the lower end of c[n] corresponds to the filter components whereas the higher end of c[n] corresponds to the source-excitation components (Bogart et al., 1963). The filter cepstrum $c_h[n]$ (equation 3) can be estimated by multiplying c[n] with the low-quefrency lifter $L_h[n]$ (equation 2).

$$L_h[n] = \begin{cases} 1, & 0 < n < L_c \\ 0, & L_c < n < N \end{cases} \tag{2}$$

$$c_h[n] = L_h[n] \times c[n] \tag{3}$$

Where, L_c is the cutoff length of lifter $L_h[n]$ and N is the total cepstrum length. The source cepstrum $c_e[n]$ (equation 5) can be estimated by multiplying c[n] with the high-quefrency lifter $L_e[n]$ (equation 4).

$$L_e[n] = \begin{cases} 1, & L_c < n < N \\ 0, & else \end{cases} \tag{4}$$

$$c_e[n] = L_e[n] \times c[n] \tag{5}$$

The log-magnitude frequency response (in decibels) of source and filter waveforms can be recovered by re-applying a real DFT separately on $c_e[n]$ and $c_h[n]$ respectively. The procedure results in the separation of the log-magnitude spectrum of speech frequency between source and filter log-spectrums as shown in equations 6 and 7 respectively.

$$\log |E[\omega]| = real\ (DFT\ \{c_e[n]\}) \tag{6}$$

$$\log |H[\omega]| = real\ (DFT\ \{c_h[n]\}) \tag{7}$$

Since, the log-spectrums log $|E[\omega]|$ (equation 6) and log $|H[\omega]|$ (equation 7) are computed from $c_e[n]$ and $c_h[n]$ respectively, and $c_e[n]$ and $c_h[n]$ are multiplications of c[n] by a step function (equations 3 and 5), it can be deduced that the log-spectrums log $|E[\omega]|$ and log $|H[\omega]|$ are simple convolutions of log-spectrum log $|S[\omega]|$ (equation 1). An example of source and filter log-spectrums and the resulting speech wave (spread by the lips) log spectrum is shown in figure 9.

In paper 3, equations 1 to 7 were utilized to develop the 'cepstral separation difference' algorithm that has shown a strong capability to discriminate between speech symptom severity according to the UPDRS-S.

In paper 4, the CSD features were used as measures of phonatory symptoms for the classification of speech intelligibility in PD.

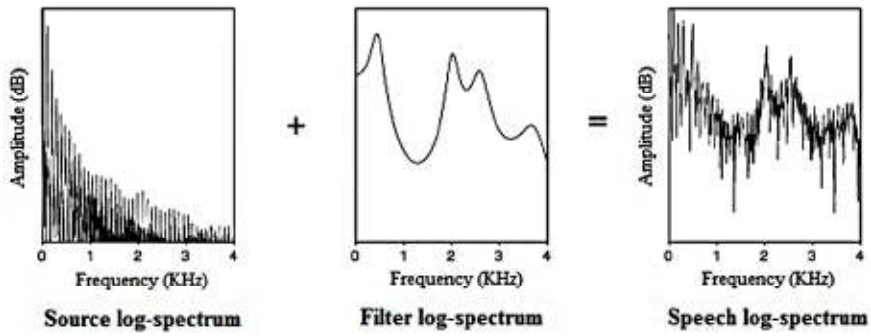


Figure 9: The source-filter model (Nooteboom and Coden, 1983).

3.1.1.2 Mel-frequency cepstral coefficients

The MFCC are other features that rely on the source-filter model of speech production (Tsanas, 2012), with an addition that they partition the speech frequencies into overlapping mel-frequency banks followed by the application of cepstral and cosine transformations on each bank (Davis and Mermelstein, 1978). For their ability to partition between speech frequencies, MFCC are suited to effectively quantify the potential problems in speech articulation (Tsanas et al., 2012; Londono et al., 2011). For instance, the speech signal is characterized not only by the vibration of the vocal folds but also by the resonating frequencies, formed by the controlled movement of articulators (tongue, jaw, lips etc). Besides, it is known that PD affects movement of the articulatory muscles in addition to the movement of the vocal folds (Ho et al., 1999). The subtle dislocation in the movement of articulators deteriorates speech intelligibility and results in varying energy in specific frequency bands of the speech signal. The MFCC compute energy differences between the bands of speech frequencies which can be used to discriminate varying energy levels of disturbed resonances.

Another advantage of partitioning speech frequencies into mel-frequency filter banks is that, the mel-filters provide a simulation of the human auditory system (Stevens and Volkman, 1940). In the case of pathological speech assessment this is advantageous, because it allows estimating the auditory perception of patient's speech by the clinician in the form of MFCC. Importantly, these MFCC can be correlated with the clinical ratings in order to automate the classification of speech symptom severity according to the UPDRS-S.

The mel-frequency filter banks are triangular in shape and compute the energy spectrum around the center frequency in each individual band of

speech frequency. The boundary frequencies of filter banks are spaced using the mel-scale given in equation 8.

$$m = 1127 \ln\left(1 + \frac{f}{700}\right), \ 0 \leq f \leq F_s \tag{8}$$

Where F_s is the sampling rate of frequency f in hertz.

The mel-frequency cepstrum is computed by taking a discrete cosine transform on the logarithm of the mel-scaled DFT of speech signal S_i. In order to compute MFCC, the log-energy at the output of each mel-filter is computed. The MFCC is the discrete cosine transform of filter energy outputs, given in equation 9.

$$MFCC_n = \sum_{k=1}^{K} E_k \cos[n(k - 0.5)\,\pi/K], \qquad n = 0 \ldots L \tag{9}$$

Where L is the number of MFCC, K is the number of filter banks and E_k is the log energy of the kth filter. Typically, a value of L between 10 and 16 is used (Davis and Mermelstein, 1978). The 0th MFC coefficient represents the original signal energy and is generally ignored. The value of K is chosen between 20 and 40. The block diagram of MFCC computation is shown in figure 10.

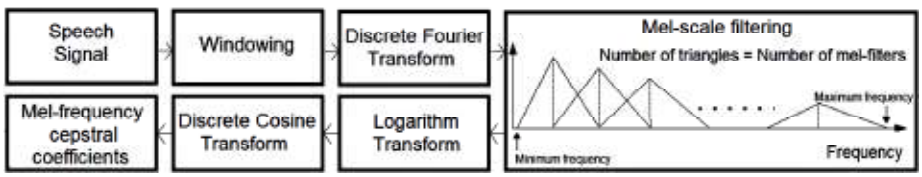


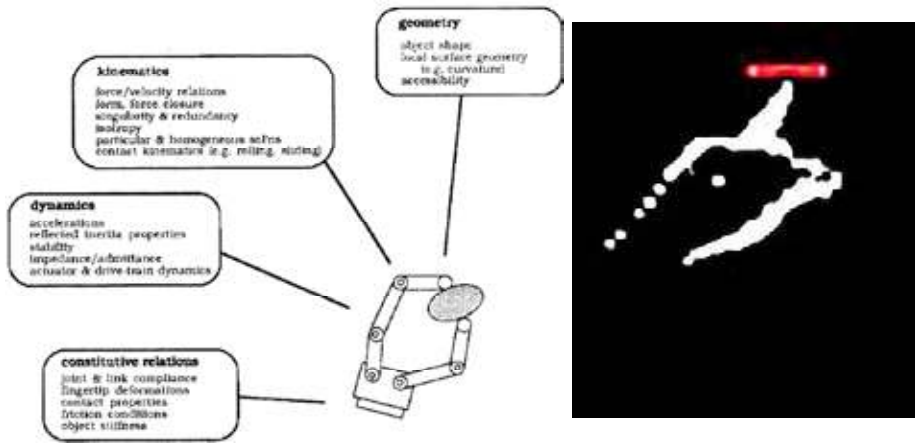
Figure 10: Block diagram of MFCC extraction.

In paper 4, the MFCC were used as measures to estimate articulatory symptoms in PD speech. In paper 2, the MFCC were utilized to assess the classification ability of different types of speech tests, including TRS, SVP and DDK tests, to identify speech symptom severity according to the UPDRS-S.

3.1.2 First principle model for finger-tapping assessment

In the field of robotics, the study of the first principles of hand mechanics has long been an area of interest for designing structures of prosthetic hand devices, robotic hands for surgeries and for quantifying the extent of disability in individuals with injuries or idiopathic disorders such as PD. In particular, there is substantial literature available on the 'precise grasping' capabilities of the human hand in which the coordinative interaction between the prosthetic index-finger and thumb is analyzed using the first principles of kinematics, contact forces and other physical laws of motion (Cutkosky &

Howe, 1990). A particular robotic hand model and the first principle features associated with this model is illustrated in figure 11a.



a. First principle mechanics of a robotic hand (Cutkosky & Howe, 1990).

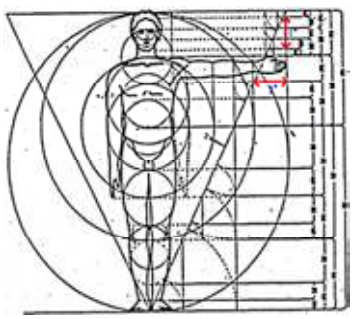
b. Motion estimation of the right hand index-finger.

Figure 11: First principle mechanics of finger-tapping.

The video based motion analysis of the index finger allows conceptualizing of the first principle mechanics used in developing the robotic hand. For instance, having gathered position-time data from a number of frames in the digitized video clip, the motion of the moving index-finger can be studied (figure 11b). Importantly, the first principle motion parameters such as time, displacement, velocity, acceleration, energy etc. can be measured. This is especially useful when conducting a motor examination using the UPDRS-FT (section 2.4.1), where the finger movement is so rapid that sometimes the underlying symptoms go unnoticed by an unaided eye. The high-speed cameras and the first principle motion parameters allow to capture these symptoms with extraordinary sights.

For our work, it was enough to track the moving index-finger using motion gradients (Bradski & Davis, 2002) since the motion of the index-finger provides all the details required to compute the symptoms, including the symptoms of speed, amplitude, fatigue and rhythm, according to the UPDRS-FT. However, one difficulty we had in locating the exact position of a moving index finger was to eliminate the rest of moving pixels, including the pixels of head and other body movements, in the video-frames. It was also difficult to simultaneously separate motion information from both hands in one image (video-frame). Keeping in view that during tapping examination, subjects position their hands above the shoulders besides the face and face directly towards the camera (figure 12), a face detector (Viola & Jones,

2004) was applied to produce a face-rectangle in the video-frames (figure 12b) which was used as a reference point to locate regions of interest for the detection of index-finger motion of left and right hands respectively (figure 12c). Specifically in the case of PD patients, one may notice the involuntarily movement of the head that occludes the motion of index-fingers. By locating the face, the motion pixels related to the head movements can be deleted.



i. The Vitruvian man (Livio, 2008)

ii. Hand vs. face

a. The model is based upon the golden ratios called the proportions of man. ii. Face height (marked with blue arrows) can be used to approximate a hand's length.

b. Face-rectangle

c. Virtual framework for tapping analysis

Figure 12: The golden ratio and the role of face detection in finger-tapping quantification.

A major problem with the previous video based methods for tapping analysis was that the tapping signal amplitude was affected by the distance of the camera from the subject's hand. This problem could not be corrected (Criss & McNames, 2011; Kupryjanow et al., 2010). This amplitude calibration problem made an estimation of tapping features for symptom characterization problematic. We decided to utilize Da Vinci's golden ratios (Livio, 2008) which indicated that the length of an adult person's hand is approximately equal to the height of that person's face (figure 12a). The resulting face-rectangle height enabled us to normalize the tapping amplitude in order to cope with the varying positions of the camera during recording. For instance, the nearer the camera is to the subject, the larger the finger-tapping amplitude and the face height. The opposite is true if the camera is placed at a farther distance. The tapping signal normalization calibrates the tapping amplitude resulting in a correct estimation of tapping features. A further advantage of having a face rectangle is that, the face regions can be blurred to comply with ethical restrictions when publishing the videos (figure 12b). Moreover, the subjects may be assessed for dystonia symptoms by analyzing their facial expressions (a possible extension to this work). In short, the de-

tection of face enables to use a simple camera to develop a virtual framework for the assessment of finger-tapping symptoms in PD [Paper 5] (figure 12c).

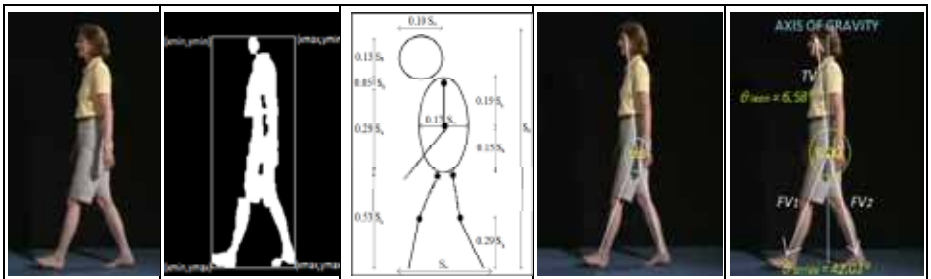
3.1.3 First principle model for gait assessment

The first-principle physiological models of the human body are commonly used for interpretation of limb functions in CV based gait analysis (Brubaker, 2012). An important issue in the design of these models is the explosion of dimensionality if one is to learn the temporal, spatial and photometric variability during the gait analysis. The dynamics, geometry and appearance can change drastically as the person moves from one position to another. One solution is to strongly constrain these phenomena by assuming that the gait platform is a linear subspace of infinite dimension of appearance and configuration (Doretto & Soatto, 2006). By this way a simple representation of appearance, for instance the silhouette of a binary image, evolves on a linear space in some feature representation. Then, a prior model of human physics can greatly help in separating the body parts of the binarized silhouette in order to estimate the severity of symptoms in different gait dimensions.

When combined with a suitable control mechanism, the first-principle physical models offer several advantages. First, they ensure that the estimated motions are physically plausible. Secondly, they allow estimating Newtonian and biomechanical principles of human locomotion, which can provide a second opinion on gait and posture condition during a clinical gait examination (section 2.5.1). Specifically when evaluating stooped posture in PD, 'posture-first principles' (Woollacott & Shumway-Cook, 2002) can be computed by mapping between physical models and binarized silhouettes in video-frames to estimate the risk of falling.

In this work, the binary silhouette in gait recordings was obtained by segmenting between the foreground and background pixels of video-frames using brightness thresholding [Paper 7]. The obtained silhouette which represents the foreground in white was isolated using a bounding rectangle (figure 13b). This rectangle was then mapped to a gait model (previously described by Tafazzoli & Safabakhsh, 2010) using values of the silhouette's height S_h and width S_w. The head, torso and leg segments in the binarized silhouette were estimated by dividing the bounding rectangle into anatomical proportions as shown in figure 13c. According to these anatomical proportions, the head segment comprises of the upper 13% portion of silhouette's height S_h. The width of head segment is 10% of silhouette's width S_w. The torso segment is comprised of 28.8% of S_h. The width of the torso segment is 17% of S_w. The legs are comprised of the lower 53% of the silhouette's height S_h.

Once the proportions of the body segments were separated, the COG of the binarized silhouette was computed and the body was skeletonized by joining tangents of the silhouette's medial points to the COG as shown in figure 13d. The tangent which joins the head segment to COG is called the torso vector. The angle between the axis of gravity and the torso vector denotes the lean angle of silhouette's posture. The two tangents that join the left and the right feet to the COG respectively are called feet vectors. The angle between two feet vectors represents the stride angle. Notice that the larger the stride angle of subject during the gait, the larger the stride length will be in the gait cycle.



a. Original Image | b. Bounding Rectangle | c. Human Model Proportions | d. Skeleton encircled with COG | e. Normal Gait parameters

Figure 13: First principle modeling of gait parameters.

As discussed in section 2.5, the easiest first principle parameter to be measured quantitatively among different temporal and spatial gait parameters is the walking speed (meters/sec) because the motion and alignment of body joints are strongly influenced by the velocity throughout the gait cycle (Pasquina & Cooper, 2009). Studies revealed that cadence and speed are gait features which are robust to dopa-medication (Ghassemi et al., 2006). The speed during the gait is related to two variables; the stride cycles and the stride length. Further, the patients with PD have a tendency to lean forward during the gait (Riva et al., 1998). For these reasons, two gait features (1) posture lean and (2) stride cycles were emphasized in the gait analysis presented in paper 7.

3.2 Data-driven classification of symptom severities

3.2.1 Support vector machines

The most recent advance in data-driven modeling to garner widespread application by fields outside of AI itself is support vector machine, first developed by Vapnik (1992). Presently, the SVM has gained immense popularity in biomedical decision systems for their ability to produce optimal training

results by implementing flexible decision boundaries in high dimensional feature space (Guan, 2011). Previous classifiers separated classes using hyper-planes. The SVM widened this idea of separating hyper-planes to data that cannot be separated linearly by mapping predictors (support vectors) onto a new higher-dimensional feature space in which the data can be separated linearly. This non-linear classification and mapping of data into high-dimensional feature space is performed by SVM using a trick called kernel. Computationally, finding the best location of hyper-planes to facilitate a kernel function, to create linear boundaries through non-linear transformation, may lead to a convex quadratic programming optimization problem. This optimization problem can be solved using the SMO algorithm (Scholkopf et al., 2001). A typical example of SVM linear separation of a two-dimensional two class problem is illustrated in figure 14.

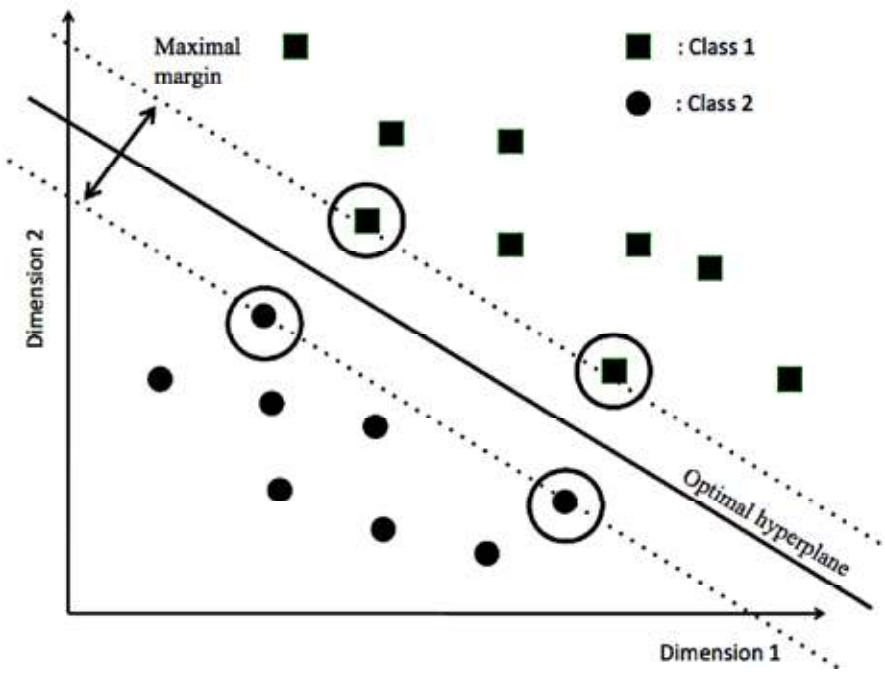


Figure 14: SVM linear separation of a two-dimensional two class problem. The solid line represents the optimal hyper-plane, the black circles represent support vectors and the dotted lines represent the maximal margin.

The principle benefits of SVM over other machine learning algorithms (e.g. artificial neural network) are as follows: (1) The SVM's solution for finding the optimal location of hyper-planes has a single minimum. For this reason the system cannot fall into a sub-optimal solution provided by a 'local

minimum', which is possible when using neural networks. (2) The SVM does not over-train if the 'right data' is given (Cristianini & Shawe-Taylor, 2000), i.e. the systems over-train if a large feature set is provided to the system for training having a low sample size; or if the data is too noisy to relate the training compounds to the property of interest. In this regard, the utilization of first principle models is handy since they provide a limited but clinically rich feature set to train the SVM. (3) The SVM does not suffer from the susceptibility of too few neurons that truncate the numerical range of coverage to cluster predictions, as is in the case of neural nets.

In short, the SVM optimized by SMO, provides a fast training algorithm that guarantees the optimality of training results (Scholkopf et al., 2001). It requires only a little a priori knowledge, i.e. only a labeled dataset. The implicit regularization of the classifier's complexity avoids over-fitting and leads to good generalizations. A brief introduction of SVM is given. Details can be found elsewhere (Decoste & Schölkopf, 2002).

Considering an n-class classification problem and a set of training vectors $\{V_i\}_{i=1\ldots M}$ with the corresponding label S_i, the SVM classifier assigns a new label $\hat{S}$ to a test vector T by evaluating

$$\hat{S} = \sum_i \alpha_i S_i K(T, V_i) + b \tag{10}$$

Where, the weights α_i and bias b are SVM parameters which are maximized during SVM training. K (;) is the SVM kernel function. The maximization of weights α_i in training SVM leads to a quadratic programming optimization problem which can be expressed in the dual form as:

$$\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j S_i S_j K(V_i, V_j) \tag{11}$$

Subject to

$$0 \le \alpha_i \le C \text{ and } \sum_i \alpha_i S_i = 0 \quad \text{for } i=1,2,\ldots n \tag{12}$$

Where, C is a positive constant called the SVM hyper-parameter that weights the influence of training errors.

SMO is an iterative algorithm that breaks the optimization problem expressed in equations 11 and 12 into a series of the smallest possible subproblems which are then solved analytically in each iteration. The SMO treats SVM weights α_i as Lagrange multipliers. The idea is that, for the smallest possible optimization problem involving two Lagrange multipliers α_1 and α_2, the linear equality constraints between the two multipliers should be reduced to

$$0 \le \alpha_1, \alpha_2 \le C \tag{13}$$

And

$$S_1\alpha_1 + S_2\alpha_2 = k \tag{14}$$

Where, k is the equality constraint variable. SMO finds the globally optimal solution by following these steps: 1) the algorithm finds a Lagrange multiplier α_1 that violates Karush-Kuhn-Tucker (Kjeldsen, 2000) conditions, 2) picks a second multiplier α_2 and optimize the pair (α_1, α_2), and 3) repeats steps 1 and 2 until convergence. The quadratic programming optimization problem is solved when all the Lagrange multipliers satisfy the Karush-Kuhn-Tucker conditions.

Albeit the SMO guarantees convergence of the SVM function, the choice of kernel function is important for transforming non-linear feature space into a straight linear classification solution. The choice of function is based on the nature of feature space that can be linear, polynomial or radial basis. In our case the underlying specificity regarding the qualitative nature of data could not be determined due to a high variability of pathological signals computed in different clinical examinations. To circumvent this limitation, a universal kernel function based on Pearson VII function (Ustun et al., 2006) was utilized. The Pearson VII function is generally used to solve curve fitting problems and has a general form that is given in equation 15.

$$f(x) = \frac{H}{\left[1 + \left(\frac{2(x - x_0)\sqrt{2^{1/\omega} - 1}}{\sigma}\right)^2\right]^{\omega}} \tag{15}$$

Here H is the peak height at centre x_0 of the peak and x is an independent variable. The variables σ and ω control the tailing factor and half-width of the peak respectively. Importantly, a curve with ω equals to 3 and σ equals to 1, is comparable to a sigmoid function used in the neural network modeling (Ustun et al., 2006).

Ustun et al. (2006) modified equation 15 to formulate a kernel function for SVM in equation 16. For a given set of training vectors $\{V_i\}_{i=1\ldots M}$, the single variable x in equation 15 is replaced by two training vectors V_i and V_j. A Euclidean distance between these vectors is introduced so that two identical training vectors would have a zero distance. The peak height H is replaced by 1 and the peak-offset x_0 is removed.

$$K(V_i, V_j) = \frac{1}{\left[1 + \left(\frac{2\sqrt{|V_i - V_j|^2}\sqrt{2^{1/\omega} - 1}}{\sigma}\right)^2\right]^{\omega}} \tag{16}$$

The SVM configured with the Pearson VII kernel function and optimized by the SMO algorithm was trained using the first principle features to classify symptom severity according to UPDRS.

3.3 Statistical tools for data analysis

3.3.1 Guttman's coefficient of monotonicity

The UPDRS based assessments are subjective and the raters may differ in opinion when identifying the symptom level. These assessments become more complicated when a rater has to follow a range of symptoms to discriminate between levels of severity. For instance, in the case of speech, a rater has to assess symptoms involving different speech components including anomalies in respiration, phonation, articulation and prosody (Midi et al., 2008), each of which is computed using a different acoustic feature.

Previous research suggests that impairment in phonation is the most common manifestation in PD speech (Midi et al., 2008), albeit PD causes disruption in the speech production system as a whole. For instance, hoarseness and harshness are common symptoms of phonation in PD speech. Besides, PD symptoms of articulation involve short rushes of speech, articulation blurring, improper consonant articulation etc (Ho et al., 1999). Sometimes it is possible that two speech samples are given a similar symptom severity rating by a clinician, yet these samples have anomalies in different speech components. For example, strong symptom of vocal harshness in one sample may be weak in another similarly rated sample which has strong articulation blurring. Another issue is that some symptoms may likely remain unnoticed by the clinician as these symptoms may not interfere so much with speech intelligibility. For example, hoarseness ('soft speech') may remain a consistent attribute in PD speech, but harshness ('rough speech') may cause more of a disruption in the clinician's perception of the intelligibility of the patient's speech (Stewart & Selesnick, 2010; Khan et al., 2013).

All these symptoms are estimated using different acoustic features. In a given situation, it is possible that feature quantities do not follow a monotonic trend relative to their corresponding symptom severity levels of 0' (normal) to '4' (unintelligible) on the UPDRS scale. A similar situation is associated with the finger-tapping assessment. A clinician rating the symptom severity using the UPDRS-FT scale, rates rhythm and fatigue in the 2nd and 3rd levels of severity, besides he rates speed and amplitude in the 1st and 2nd levels only. This causes a disruption in the monotonicity of feature quantities with respect to the increasing order of symptom severity which complicates the choice of the correlation model.

In the given situation, the one-to-one mapping between a computerized feature (representing an individual symptom) and the corresponding clinical

rating (based on multiple symptoms) is not possible through the use of rank-order or the Likert scale. One may choose Spearman's rho which utilizes a rank-order scale to correlate between two variables in an ordered dataset (Myers et al., 2010). A restriction with the rank-order scale is that, an agreement between two variables on one class level is strictly based upon the agreement between these variables on the former class level. If an agreement between a rater and a computerized feature has to be made in a succeeding class level (say level '3') based on a class property (e.g. a speech symptom) which does not exist in the former class level (say level '2') but exists in levels preceding the former level (say levels '0' and '1'). In that case the Spearman's rho would penalize the correlation value since the agreement between rater and feature value in the succeeding class level (level '3') could not be reached due to an absence of this agreement in the former class level (level '2').

This problem of monotonicity in feature quantities can be solved using the 'Guttman scale' (Guttman, 1944), developed by Louis Guttman while he was serving as an expert consultant to the US Army's research branch during World War II. During his service in the army, he developed scale analysis and techniques that were primarily used to study the psychology of American soldiers involved in the war (Guttman et al., 1950). Later on, these methods were utilized in statistical correlation analysis of qualitative rank-order variables, specifically to study human behavior in different aspects of life. In relation to this thesis, Guttman's approach is well-suited for the assessment of qualitative ranked nature of a clinical dataset where a human rater examines the proportions of different symptoms to choose between the severity levels.

According to Guttman's definition of scale:

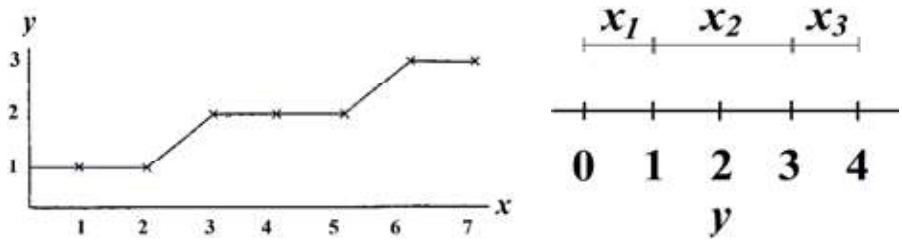
"*For a given population of objects, the multivariable frequency distribution of a universe of attributes (attitudes) will be called a 'scale', if it is possible to derive from the distribution a quantitative variable with which to characterize the objects such that each attribute is a simple function of that quantitative variable. Such a quantitative variable is called a scale variable*" (Guttman, 1944).

This definition gave rise to the concept of a 'perfect scale' according to which a correlation analysis in the real-world can be performed by one-to-one correspondence between the scale variables and the scale ranks (Guttman, 1944). For instance, let us consider the situation where the human rater examines the proportions of different symptoms to choose between the severity levels and each symptom is estimated using a different computer feature. On the Guttman scale, a variable y (say a human rater) with h distinct rank ordered values (say UPDRS severity levels) can be said to be a simple

function of the scale variable x (say a computerized feature) with i distinct ordered values, if for each value of x there is only one value of y. The converse needs not to hold and for the same value of y, there may be one or multiple values of x.

Guttman called this relationship between x and y Guttman's monotonic regression, which according to him is a relationship in which values of x increase in a particular direction as the value of y, without assuming that the increase is exactly according to a straight line (Guttman, 1944). Importantly, this trend is always in one direction, either upward or downward with the possibility that variable y can occasionally stand still. This phenomenon is closely compatible with subjective assessment using UPDRS where some symptoms may not be evaluated at certain severity levels and the level of severity must stand still without penalizing the feature computing the symptom.

The regression in figure 15a depicts the Guttman's monotonic trend between the scale variable x and the scale ranks y (Guttman, 1944). This trend can also be presented graphically as a perfect scale by plotting y (UPDRS -S or –FT levels) on a straight line and cutting this line into numerical intervals (scale scores) according to which the quantities of x (a computerized feature) can be ordered. Figure 15b is a perfect scale because it has the characteristic of perfect interval consistency, i.e. the intervals along the straight line are one-dimensional regions of scale (Guttman, 1944). Now, since there is a one-to-one correspondence between the severity of symptoms over the observed two-item categories, x (feature values) and y (UPDRS levels). This implies that the predictability of severity from the scale ranks y is exactly the same as from the scale variable x.



a. The monotone trend: A simple function of variable y on variable x.

b. Graphical presentation of perfect scale.

Figure 15: Guttman's monotonic regression (Guttman, 1944).

The Guttman's monotonic regression between x and y can be expressed in the form of an equation as:

$$\mu_2 = \frac{\sum_{h=1}^{n}\sum_{i=1}^{n}(x_h - x_i)(y_h - y_i)}{\sum_{h=1}^{n}\sum_{i=1}^{n}|x_h - x_i||y_h - y_i|} \quad (17)$$

Where, μ_2 is known as Guttman's monotonicity coefficient. The value of μ_2 varies between -1 and +1, where -1 implies a perfect monotonic trend in a negative direction and +1 implies a perfect monotonic trend in a positive direction.

In this dissertation, the primary use of μ_2 was to perform a correlation analysis between the clinical ratings and the computer features developed using the first-principle models. Additionally, the μ_2 was utilized for selecting the features that were used in training the SVM for classification between UPDRS severity levels.

3.3.2 Other statistical tools used in this thesis

3.3.2.1 The chi-squared test

In addition to using μ_2 for feature selection and analysis, in order to boost the performance of algorithm in classifying symptom severity, a chi-squared test of fitness was utilized to prune the selection of features used for classification (Geng et al., 2007). The chi-squared statistic calculates the goodness of fit of how well an input is correlated with the target class. A discrete value can be yielded for each feature fitting the test, and 0 for features failing the test. Additionally, the method returns a ranking of each feature in decreasing order by the value of the chi-squared statistic in relation to the class label. This way the uncorrelated data can be discarded prior to classification, resulting in improved classification accuracy. Readers may find a detailed review on the chi-squared test by Plackett (1983).

3.3.2.2 Analysis of variance

Speech is highly non-stationary and due to its underlying variability it is possible that the results of a test are non-repeatable. One should expect to get the same (or a similar) value of speech features with repeated measurements each time using a different speech test. An analysis of variance (ANOVA) (Bolton, 1997) was performed to test repeatability of speech features in characterizing symptom severity levels in different speech tests.

ANOVA relies on assumptions that the data distribution in each population group should indicate a normal (Gaussian) shape and the variances of population between these groups must be equal. One technique to stabilize the variance and to make the data more normal-distribution like is to apply a power transform over the dataset that normalizes the data densities. One

such power function is the Box-Cox transform (Box and Cox, 1964) which is defined as a continuously varying piecewise function with respect to the power parameter λ that makes the function continuous at the point of singularity (λ=0).

The variances of feature values between symptom severity levels were stabilized using the Box-Cox power transform prior to using ANOVA. The equivalence of mean feature values for each symptom severity level in different speech tests was compared to assess the accuracy and repeatability of features in estimating speech symptoms.

3.3.2.3 Intra-class correlation coefficient

To be a reliable discriminator of speech symptoms, the feature values must remain consistent along the symptom classes in different test scenarios. The intra-class correlation coefficient (ICC) (McGraw and Wong, 1996) is generally used to assess consistency and reproducibility by quantifying the degree to which feature values resemble each other at similar levels in different test occasions. Usually the ICC is calculated for single and average measurements. In the single measurement, a class value in one test is mapped to a corresponding class value in another test to estimate consistency. In average measurements, a class value in one test is mapped with all the class values in the other tests to estimate the consistency.

In paper 2, the single and average ICC values were computed to assess the consistency of MFCC in categorizing symptom severity levels using different vocal tests, including TRS, SVP and DDK. In paper 3, the ICC values of the novel CSD features in different TRS tests were compared with the ICC values of non-CSD features H1-H2, HNR and means of 3rd and 4th MFCC.

3.3.2.4 Receiver operating characteristic curves

The area under the receiver operating characteristic (ROC) curves (AoC) is generally used to assess feasibility of a classification model, independent of cost context and class distribution (Metz, 1978). An ROC curve can be plotted by taking the true positive rate (TPR) of a symptom severity class on y-axis against the false positive rate of that class on x-axis. Each predicted instance in the severity class represents one point in the ROC space. The best classification model produces a point in the upper left coordinate (0, 1) of ROC space, which means that there is no false negative or false positive in classification. A point along a diagonal line between the coordinates (0, 0) and (1, 1) represents a 'complete random guess' by the model. This diagonal line divides the ROC space into two halves. An ROC curve above this line represents an AoC greater than 50% of ROC space area. An AoC of 100% represents a 'perfect classification model', whereas an AoC of less than 50% represents a 'worthless model'. An AoC between 80% and 100% represents an 'excellent classification model'.

In this thesis, the ROC analysis was used in papers 2 and 4 to evaluate the performance of speech algorithms in classifying UPDRS-S severity levels. In paper 5, the ROC analysis was used to evaluate the performance of the finger-tapping algorithm in classifying UPDRS-FT severity levels.

4. Computer methods for symptom assessment of Parkinson's disease

Different methods based on FPDD models were developed for the quantification of speech, finger-tapping and gait symptoms in PD. For the sake of convenience, in this thesis, these methods are categorized into two groups: (1) clinical speech processing algorithms and (2) clinical image processing algorithms. The clinical speech processing algorithms include methods that process audio recordings of motor speech examinations in order to categorize the severity of speech symptoms according to the UPDRS. The clinical image processing algorithms include CV methods that process video recordings of motor finger-tapping and gait examinations for the classification of the severity of tapping and gait symptoms according to the UPDRS.

4.1 Clinical speech processing algorithms

4.1.1 Methods for detection of speech impairment using mobile devices: A review (Paper 1)

A systematic literature review on methods for assessment of speech impairment using mobile devices was done. A two-tier review methodology was utilized. The first tier focused on real-time problems in speech detection. In the second tier, acoustic features that respond to medication changes in levodopa-responsive patients were investigated. The choice of publications for review was based on validation of methodology, credibility of experimental techniques for speech analysis and portability of the algorithm to modern mobile devices. Additionally, patents related to speech disorder assessment methods and biofeedback devices were reviewed and synthesized. The synthesis was performed chronologically, i.e. starting from an investigation of essential acoustic parameters for speech impairment recognition, to the systems implemented for classification between impaired and normal speech.

The investigation of patents revealed that speech disorder assessment can be made by a comparative analysis between the patterns of normal and pathological speech acoustics. The vowel and consonant frequencies are the most relevant acoustic parameters to reflect PD medication changes. However, a major challenge is to separate between frequencies of vowels and consonants in the speech signal. Another challenge is to discriminate be-

tween environmental noise and impairment-related noise in the speech signal so that the environmental noise can be removed without affecting important information regarding the symptoms that is found in the speech signal.

The literature suggests that an automatic speech processor must be able to identify silences and stop gaps, as well as vocalic and fricative sounds that provide gains in speech intelligibility. Especially in processing impaired speech, voice segmentation is difficult because syllable units are spread roughly via intensity changes. Another difficulty is that acoustic features like the sound pressure level and formant analysis are gender dependent. Moreover, since speech signals are highly non-stationary, the quantification of speech signals using time frequency methods is difficult. It was demanded that speech signals should be processed in domains (e.g. cepstral domain) where anomalies can be traced in speech subsystems.

Due to such difficulties, previous research utilized SVP recordings for voice analysis since the periodic patterns of vowel signals are computationally easier to analyze. For example, a method for classification between normal and pathological voice was introduced (Little et al., 2009). This method utilized fundamental frequency and pitch period parameters (e.g. pitch period entropy), along with some other acoustic features including jitter, shimmer and noise-to-harmonic ratio that were computed from SVP recordings. These parameters were used to train an SVM to distinguish between normal and pathological voice. Results indicated that the parameters were robust to perplexed acoustic changes and gender effects. Other studies (Salhi et al, 2008; Tsanas et al., 2012) suggested that an artificial neural network and a SVM are useful tools for voice characterization.

The study of previous work advocates that future research is aimed to develop mobile devices capable of identifying PD speech symptoms. The advanced sound processing capability of these devices would allow delivering auditory and visualized feedback to the patients. Importantly, the medicine adjustments would be made based on the computerized rating of symptom severity, supporting timely treatment of patients in their home environment.

4.1.2 Data for speech analysis

The speech recordings in this thesis were obtained from a feasibility study of an at-home treatment device, through a procedure described by Goetz et al. (2009). The data acquisition was conducted at the University of California, San Francisco in collaboration with the Parkinson's Institute. Over the course of one year (i.e. from June 2009 till June 2010) speech data was acquired from a total of 80 subjects, 48 males and 32 females with an average age of 63.8 years, using a computer-based test battery called QMAT. 60 subjects (40 males and 20 females) had PD duration of 75.4 weeks. 20 other subjects were normal controls.

The recorded speech samples consisted of four types of SVP, two types of DDK and three types of TRS tests (table 2). In the SVP tests, the vocal breathiness of patients in keeping the pitch (e.g. 'aaaah...') constant for 12 seconds was examined. Four types of SVP tests, the first at the comfortable constant loudness, the second with twice the initial loudness, the third with thrice the initial loudness, and the fourth with quadrice the initial loudness, were recorded. In DDK tests, the ability of patients to produce rapid alternating speech (e.g. 'puh-tuh-kuh…puh-tuh-kuh…') were assessed. Two types of DDK tests, one at the comfortable level of loudness and the other with twice the initial loudness were recorded.

In the TRS tests, the subjects were asked to recite standard phonetic paragraphs (International Phonetic Association, 1999) that were displayed on the QMAT screen. The paragraphs included 'the north wind and the sun', 'the rainbow passage' and 'the grandfather passage' in tests 1, 2 and 3 respectively (table 2). These paragraphs were structured in a way such that the textual difficulty in reading the paragraphs increased from test 1 to 3. Specifically, the text involving fricatives are difficult to pronounce, as the fricatives require strong coordination of vocal muscles to produce sound (Silbert and de Jong, 2008). The paragraph 'the north wind and the sun' used in TRS test-1 consisted of 5 sentences involving 61 fricatives and was recorded for 40 seconds from each subject. The paragraph 'the rainbow passage' used in TRS test-2 was more difficult to read than TRS test-1 paragraph, and consisted of 6 sentences involving 65 fricatives, recorded for 50 seconds. The paragraph 'the grandfather passage' used in TRS test-3 was the most difficult to read. The third paragraph consisted of 8 sentences involving 87 fricatives. This last paragraph was recorded for 60 seconds.

A clinician examined the speech samples and rated the symptom severity using the UPDRS-S (section 2.3.4). Out of 80 subjects, 24 subjects were rated '0', 25 subjects were rated '1', 28 subjects were rated '2' and 3 subjects were rated '3'. As there were few samples in the severity group '3', they were merged into severity group '2'. This resulted in three severity groups for classification, where group '2' represents 'moderate-severe' impairment.

The sampling rate of speech samples was 48 kHz with 16 bit resolution. In total 720 speech samples, including 320 SVP, 240 TRS and 160 DDK samples, were used in the feature analysis and classification.

Table 2: Motor speech examination structure.

Test Type	Number of sentences	Number of fricatives	Total vowels	Length of recordings	Linguistic stress level
[1]TRS test-1	5	61	136	40 seconds	Low
[2]TRS test-2	6	65	145	50 seconds	Medium
[3]TRS test-3	8	87	185	60 seconds	High
SVP test-1	-	-	(Sustained) 1	12 seconds	Comfortable constant loudness
SVP test-2	-	-	(Sustained) 1	12 seconds	Twice the comfortable loudness
SVP test-3	-	-	(Sustained) 1	12 seconds	Thrice the comfortable loudness
SVP test-4	-	-	(Sustained) 1	12 seconds	Quadrice the comfortable loudness
DDK test-1	-	(Repetitive) 3	(Repetitive) 3	12 seconds	Comfortable constant loudness
DDK test-2	-	(Repetitive) 3	(Repetitive) 3	12 seconds	Twice the comfortable loudness

[1]TRS test-1: "The north wind and the sun were disputing which one is the stronger when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler's take his cloak off should be considered the stronger. Then the north wind blew as hard as it could but the more he blew the more closely the traveler pulled his cloak around him and at last the north wind gave up the attempt. Then the sun shined out and immediately the traveler took off his cloak. And so the north wind was agreed that the sun was the stronger of the two."

[2]TRS test-2: "When the sunlight strikes rain drops in the air, they act like a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch with its path high above and its two ends apparently beyond the horizon. There is according to a legend a boiling part of gold at one end. People look but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the part of gold at the end of the rainbow."

[3]TRS test-3: "Do you wish to know all about my grandfather; well he is nearly 93 years old. He dresses himself in an ancient black frock coat usually minus several buttons. Yeah he still thinks he is swiftly as ever. A long flowing beard clings to his chin giving those who observe him a pronounced feeling of an outmost respect. When he speaks, his voice is just a bit cracked and covers the trifle .Twice each day, he plays skillfully and with a zest upon a small organ except in the winter when the ooze or snow or ice prevents he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less but he always answers 'banana oil'. Grandfather likes to be modern in his language."

4.1.3 Running-speech MFCC are better markers of Parkinsonian speech deficits than vowel phonation and diadochokinetic (Paper 2)

As discussed in section 3.1.1.2, the MFCC are relied on for their capability to estimate anomalies in pathological speech. Previous research on PD speech (Jafari, 2013; Tsanas et al., 2012) used MFCC from SVP signals to separate speech symptom severity. However; these methods did not consider computing the MFCC from the TRS in order to evaluate speech symptoms. This was possibly due to complexities in processing the TRS signals (section 4.1.1). On the other hand, the research on vocal disorders, other than PD, used TRS-MFCC to model speech aberrations. For example, Llorente et al. (2009) parameterized the MFCC from 140 recorded TRS samples that classified between 117 dysarthric and 23 normal samples with an accuracy of 96%. Similarly, Paja et al. (2012) used the MFCC computed from TRS samples to evaluate spastic dysarthria. Their experiments indicated a strong correlation between MFCC and 2-level ('Low-Mid' and 'Mid-High') subjective ratings of speech intelligibility. In another experiment to estimate speech depression (Cummins et al., 2013), the TRS-MFCC displayed the strongest discriminatory characteristics compared to other acoustic features, when classifying the presence and absence of depression.

Previous studies using vowel phonations (Jafari, 2013; Tsanas et al., 2012) support the feasibility of using the MFCC for scoring PD voice symptoms. However, these studies ignored the fact that the vowel /a/ has fundamental frequencies that are concentrated only on a part of frequency scale; roughly between the range 50 Hz to 300 Hz (Whalen & Levitt, 1995; Titze, 1994; Ye et al., 2003), which can provide a relatively smaller amount of symptom information in speech. Besides, the mel-frequency filters can be applied over the complete speech spectrum to compute disturbances in consonantal frequencies that lie above 1 kHz on the speech spectrum (Fry, 1979). Importantly, it should not be neglected that the sounds of consonants are formed by the movement of vocal tract articulators, and symptoms in articulation could be better estimated using TRS that consist of consonants and fricatives. Moreover, it is known that vowels carry the power of voice, but consonants provide intelligibility in speech (Burkle et al., 2004). A clinician, using the UPDRS, scores symptoms based on the perceived intelligibility of synthesized speech, and not only by hearing the power of voice.

Supported by the fact that the mel-filters provide a simulation of the human auditory system (section 3.1.1.2) that enables quantifying the (clinician's) perception of (the patient's) speech, a larger amount of impairment related information can be captured using the MFCC that are obtained from TRS signals, rather than by using the MFCC from the SVP and DDK. Paper 2 aims to utilize MFCC computed from TRS recordings for classification of speech symptom severity according to the UPDRS-S. A secondary objective

is to perform a comparative analysis between the classification performance of the MFCC computed from recorded TRS tests, and the MFCC computed from recorded SVP and DDK tests in discriminating between speech symptom severity levels.

The MFCC were computed from audio recordings of speech tests using equation 9. In order to compute MFCC, the audio signals were divided into frames of 50 ms each. A filter bank of $K=24$ was applied to extract up to 16^{th} order ($L=16$) MFCC from each frame. The mean of each MFCC between the frames was computed and used for analysis.

4.1.3.1 Feature Analysis

The μ_2 (equation 17) was utilized to map between the MFCC computed from TRS, SVP and DDK samples, and the clinical ratings based on UPDRS-S. Jackknife cross validation (Berger, 2007) was used to stratify correlation estimates. The results are given in section 5.1.1.

Apart from the correlation analysis, the test-retest reliability of MFCC was assessed using the ICC coefficients. The MFCC in the order 1 to 16 were computed separately from recordings of TRS test 1, 2 and 3 respectively and were merged to form 16 different TRS-MFCC groups, such that each group consisted of an order of MFCC computed from three different test occasions (test 1, 2 and 3) of TRS. Likewise, the MFCC in the order 1 to 16 were computed separately from recordings of SVP test 1, 2, 3 and 4, and were merged to form 16 different SVP-MFCC groups. Similarly, the MFCC in the order 1 to 16 were computed separately from recordings of DDK test 1 and 2, and were merged to form 16 different DDK-MFCC groups. A comparison of single and average ICC coefficients for each MFCC group is presented in section 5.1.2.

4.1.3.2 Classification

The SVM described in section 3.2.1 was trained using the MFCC to classify between speech symptom severity levels. Two different classification experiments were performed using the MFCC computed separately from TRS, SVP and DDK recordings. In the first experiment, the data were stratified using 10 fold cross validation to allow computing unbiased generalization estimates of MFCC in different speech tests. In the second experiment, the data were separated between training and testing sets to validate the generalization performance of MFCC in different speech tests. The classification ability of MFCC in both these experiments was assessed using the AoC. The results (section 5.1.3) support the use of TRS for subjective as well as for objective assessment of speech pathology in PD.

4.1.4 Cepstral separation difference: A novel approach for speech impairment quantification in Parkinson's disease (Paper 3)

A laryngeal-videoscopic examination in a previous study (Midi et al., 2008) revealed that the uncontrolled glottal closure pattern is the most frequently manifested symptom in PD speech. In paper 3, a novel algorithm called the 'cepstral separation difference' is introduced that is capable of estimating pressure wave disturbances caused by the uncontrolled glottal closures in speech. Importantly, it computes the source and filter log spectrums by performing cepstral homomorphic filtering on speech signal (section 3.1.1.1) that enables first principle estimation of speech dysfunction.

In the first step, the real cepstrum c[n], of cepstral coefficients n, was computed by applying an inverse DFT on the real log of the DFT of speech signal s [i] (equation 1). In the cepstral domain, the source cepstrum $c_e[n]$ and the filter cepstrum $c_h[n]$ were separated by liftering the cepstrum c[n] by applying a low-time and a high-time lifter on c[n] respectively (equations 3 and 5). A cutoff value of 20 cepstral coefficients was used for liftering, which is generally used for identifying the speech source in speech recognition systems (Kim et al., 2004). In the third step, the source and filter log spectrums were computed by applying the real DFT on the source cepstrum $c_e[n]$ and the filter cepstrum $c_h[n]$ respectively (equations 6 and 7).

In the case of impaired speech, an unintended glottal closure hinders the airflow from the lungs and creates turbulence at the constriction point, resulting in meager propagation of pressure waves to the vocal tract. The turbulence generates aspiration noise at the glottis (section 2.3.2) that increases energy in the source component of speech. Additionally, the meager propagation of pressure waves in the vocal tract reduces energy in the filter component of speech. The increased energy in the source and the decreased energy in the filter imply that the mathematical difference between the magnitudes of source and filter log-spectrums should be larger in impaired speech compared to normal speech. Accordingly, the residual log-spectrum log |r[ω]| was computed by taking the difference between the magnitudes of source and filter log-spectrums within the frequency range of 0Hz-1000Hz (the fundamental frequency range of the human voice; Catford, 1964) as given in equation 18.

$$R[\omega] = \log|r[\omega]| = \log|E[\omega]| - \log|H[\omega]| \qquad (18)$$

Where, $R[\omega]$ is called the 'cepstral separation difference'. The log $|E[\omega]|$ and log $|H[\omega]|$ in equation 18 represent the source and filter log-spectrums respectively.

It is also well known that the speech power spectrum log |S [ω]| drops off approximately at a rate of 8 dB/octave above 500 Hz (Beranek, 1986). In this case, the irregularities in the speech power spectrum, such as dispersion, cannot be estimated directly using statistical measures. The baseline of speech spectra needs to be corrected first; otherwise a steep declination of the baseline would result in the wrong calculation of dispersion. An essence of CSD is that, the subtraction between log E [ω] and log H [ω] log-spectrums, both dropping at the same rate above 500 Hz, cancels out the dropping effect and calibrates the baseline of log-spectrum $R[\omega]$ so that it lies parallel to the horizontal axis. This enables to estimate the modulation changes, as well as the increased magnitude, in the residual log-spectrum R [ω] using the standard statistical measures.

Two separate measures were developed using log-spectrum R [ω] to identify 'harshness' and 'hoarseness' in PD speech respectively. Harshness ('rough speech') causes irregularity in source modulation that is evident in the residual log-spectrum. A measure to estimate irregularity in the modulation of the residual log-spectrum is the mean absolute deviation (represented as δ_{CSD}) between the magnitudes of R [ω] (for ω = 1…1000Hz) given in equation 19, where, $\overline{R}$ is the overall mean of R [ω]. Our experiments have shown a marked increase in δ_{CSD} relative to the increasing symptom severity in speech.

$$\delta_{CSD} = \frac{1}{1000}\sum_{\omega=1}^{1000}\left|R[\omega]-\overline{R}\right| \qquad (19)$$

On the other hand, hoarseness ('soft speech') results in depressing the speech frequency and can be measured by estimating the raised level of log-magnitude in R. A peak-detector was applied on R to locate peaks representing the level of residual log-magnitude in each frequency. The average peaks' magnitude (AP_{CSD}) was found elevated in pathological speech samples and was rising with increasing symptom severity. The δ_{CSD} along with AP_{CSD} were used for further analysis.

4.1.1.1 Feature validation and analysis

The CSD-based features were thoroughly analyzed and compared with speech features, such as the harmonic-to-noise ratio (HNR), H1-H2, and the means of 3rd and 4th MFCC, that are known to have the capability of classifying PD speech with a high degree of accuracy. Further analysis was done to validate the performance of CSD features in identifying symptom severity levels in different speech tests.

- Test of validity

The μ_2 was utilized to correlate the CSD features with the UPDRS-S ratings, as well as with the other recognized speech features. The features were extracted from the recited paragraphs in TRS test-1 (difficult to read), test-2

(more difficult to read) and test-3 (the most difficult to read) separately so that the monotonicity in correlation with respect to the increasing textual difficulty, which demands a greater laryngeal stress in pronunciation, can be observed. Jackknifing was used with the correlation analysis as a cross-validation technique to estimate the precision of μ_2. The results are discussed in section 5.2.1.

- Test of reliability

The test-retest reliability of CSD features in different TRS tests was assessed using ICC. Single and average ICC measures were computed for severity classes of each feature between TRS test 1, 2 and 3 respectively. The ICC measures for the CSD features were compared with the measures of non-CSD features H1-H2, HNR and means of 3rd and 4th MFCC. The results are presented in section 5.2.2.

- Test of repeatability

ANOVA was performed to test repeatability of δ_{CSD} and AP_{CSD} means in each UPDRS-S severity group among the three different TRS tests. Before performing the analysis, the ANOVA assumptions were fulfilled by stabilizing the variances of δ_{CSD} and AP_{CSD} using the Box-Cox power transform. The results are presented in section 5.2.3.

4.1.5 Classification of speech intelligibility in Parkinson's disease (Paper 4)

A difficulty in the clinical assessment of running speech is to track underlying deficits in individual speech components, including disturbances in respiration, phonation, articulation and prosody, which as a whole disturb the speech intelligibility (section 2.3). The aim of paper 4 is to extract signal features from TRS samples computing deficits in individual speech components, and to utilize these features to train an SVM for classification between speech symptom severity levels in accordance with the UPDRS-S. Importantly, this paper incorporates CSD features to categorize the level of speech impairment.

Several acoustic features were extracted from TRS signals to estimate symptoms in each speech sub-system. For the sake of description, these features were organized into groups as: 1) measures relating to the phonatory symptoms, 2) measures relating to the articulatory symptoms, and the 3) measures relating to the prosodic symptoms. The respiratory symptoms (e.g. reduced loudness) were estimated along with the symptoms of prosody.

The phonatory measures represent symptoms which emerge due to the incoordination between phonation and respiration and cause harshness and hoarseness in speech. Harshness was estimated using the CSD feature δ_{CSD}, whereas hoarseness was estimated using the other CSD feature AP_{CSD}.

The articulatory measures represent symptoms that emerge during subtle changes in the motion of articulators and cause imprecise articulation and short rushes of speech. An MFCC vector of 10 coefficients was used to estimate the problems in articulation.

The prosodic measures represent symptoms categorized by reduced vocal stress, monopitch intonation, monoloudness and abnormality in speech rate (Jones, 2009). Loudness and speech rate were estimated using short-term dynamics of the speech signal including the number of pauses (N_p), pause intervals (P_I), zero-crossing rate energy (ε_{ZCR}) and mean spectral centroid (SC_{AVG}). Pitch symptoms were estimated using standard deviation in fundamental frequency ($F0_{std}$), interval entropy between pitch periods (I_{ent}) and the jitter perturbation quotient in pitch periods (J_{PPQ}).

A total of 19 acoustic features capable of estimating symptoms in each subsystem of speech were computed. Some of these features were perfectly correlated between each other and were removed, leaving behind a set of 13 features for classification.

4.1.5.1 Classification between UPDRS-S levels

The SVM described in section 3.2.1 was trained using the selected acoustic features for classification between UPDRS-S severity levels. Two important investigations were made: First, to analyze the influence of reading stress on symptom severity classification with a hypothesis that the classification rate should improve relative to the increasing textual difficulty which would demand a greater stress in reading. Since the level of textual difficulty increases from TRS test 1 to test 3, three different classification tests were performed each using the samples of a different TRS. The second investigation was to estimate the classification performance of SVM on classifying the complete speech dataset. Additionally, the ROC analysis was used to evaluate the classification performance.

Investigation-1:

In the first investigation, three different classification matrices of dimensions 13 (features) x 80 (samples) for TRS test 1, 2 and 3 respectively were prepared for classification between the 3-levels of the UPDRS-S. Each classification matrix was used separately to train the SVM classifier using 10-fold cross validation. A comparison between the classification rates produced by each matrix is presented in section 5.3.1.

Investigation-2:

In order to analyze the performance of computed features and the SVM model in classifying the total speech data set, a new matrix with the dimensions of 13 (features) x 240 (total speech samples; i.e. 3 TRS x 80 samples=240 samples) was formed for separation between the 3 levels of the UPDRS-S. This matrix was then used as an input vector to train the SVM

using 10-fold cross validation. The classification results are presented in section 5.3.2.1.

Another unbiased approach to validate the generalization performance of the selected features is to introduce novel unseen data to the classification model, with a statistical assumption that the new data will have a similar distribution to the data used in training the classifier. The selected 13 features were computed from 80 samples of TRS tests 1 and 2 respectively, and were used to form a training set matrix of 13 (features) x 160 (samples; i.e. 80 TRS-1 samples + 80 TRS-2 samples = 160 samples). This matrix was then used to train the SVM classifier against the UPDRS-S targets '0', '1' and '2'. Another set of same features computed from 80 samples in TRS test 3 was used to form a test set matrix of 13 (features) x 80 (samples). This test set matrix was used for testing the trained classifier. The results are given in section 5.3.2.2

4.2 Clinical image processing algorithms

4.2.1 Methods for finger-tapping quantification: a review

Previous research generally used contact sensors for the quantification of finger-tapping (Muir et al., 1995). In recent years, lightweight sensors and accelerometers were utilized for finger movement analysis (Okuno et al., 2006, 2007; Kandori et al., 2004; Shima et al., 2008). However, the feature extraction methods in these systems did not provide measures that could be used to evaluate finger-tapping in accordance with the UPDRS. These systems were able to compute tapping velocity and acceleration. Besides, some other important features including tapping arrhythmia and fatigue were not incorporated, which are essential for the characterization of symptom severity according to the UPDRS-FT (section 2.4.1). It was further discovered that accelerometer readings are affected by gravitational artifacts (Elble, 2005).

As an alternative to sensor-based systems, CV based methods were introduced for the assessment of the finger-tapping test. A recent example is a virtual touchpad interface (Kupryjanow et al., 2010), which utilizes a webcam for video recording and a test-pad for tapping. Tapping assessments using this system were made using image contour algorithms to determine if the index-finger and thumb were touching or apart. However, this system had drawbacks similar to the sensor-based systems, i.e. it could not take into account tapping rhythm and amplitude that were important for symptom level discrimination. Another problem with this system was that, it was unable to measure tapping symptoms for dyskinetic patients who are unable to keep their hands consistently at one location, for instance over the test-pad.

An advantage of using CV algorithms is that it is feasible to incorporate them into telemonitoring systems such as the 'objective PD measurement system' (Goetz et al., 2009) comprising of dexterity and mobility measure-

ment devices which communicate to a central server through the internet. The PD related motor symptoms can be recorded on the server and can be accessed by clinicians and researchers using a web interface. The symptoms can be processed using computer algorithms installed on the server to produce objective estimates. A web-based video analysis of finger-tapping can be particularly useful when incorporated in such systems where tapping videos can be stored on a centralized server for computerized processing.

A further advantage of using CV algorithms is that they are low-cost and do not require any specialized hardware. Moreover, they allow the subject to perform tapping in a more natural setup; for example, they can perform tapping while sitting in front of their webcam equipped computers. Or, a modern mobile phone can be used as a device to collect data since they have a built in camera. The fact that a phone is portable makes it easy for a relative or healthcare worker to film a patient in their home. So even if the patient doesn't have a computer or finds it hard to sit in front of a webcam, the data can be easily collected and can be transferred to the server.

4.2.2 A computer vision framework for finger-tapping evaluation in Parkinson's disease (Paper 5)

4.2.2.1 Data for finger-tapping analysis

The data were acquired from a multi-center clinical study called 'DIREQT' (Duodopa Infusion: Randomized Efficacy and Quality of life Trial) (Nyholm et al., 2005) that was conducted during 2002-2003 at 5 different medical facilities in Sweden. The study consisted of RFT video-recordings from 13 patients (5 females and 8 males) who were suffering from advanced PD. The mean total score of these patients was 50.45 (range 14-92; SD±18.12) on the UPDRS-III scale 0 (normal state) to 108 (total motor impairment). The patients were aged between 50 and 75. The recordings took place at different timings of the day i.e. from 9 am till 5 pm with a rest of 30 minutes in between. A total of 17 recordings took place each day for each patient. Six patients were recorded for 2 days while two others were videotaped for 3 days with a gap of a week in between the recording days. Five patients could show up for a single day of recording. The tapping videos were rated by two clinicians based upon the tapping performance of each patient using the UPDRS-FT.

During the video recordings, the patients were seated on a chair with a plain wall behind them. The videos were recorded with each patient facing towards a pivoted camera. During RFT, the patients positioned their hands above their shoulders beside their face. They were asked to stretch the index finger vertically against the thumb as much as possible. The tapping had to be done as fast as possible. The visual features of interest for the two clinicians were the tapping rate, amplitude between index-finger and thumb, he-

sitations, halts, and decrement in the amplitude. The raters classified the tapping symptoms according to the four symptom severity levels in the UPDRS-FT (i.e. '0', '1', '2' and '3'). Each patient was videotaped for 10 seconds at a frame rate of 25 fps and a frame resolution of 352 x 288 pixels.

In addition to the patient group, RFT from 6 healthy controls (HC), aged between 40-60 years, were recorded with the same video configuration. A total of 84 recordings were made over a span of two weeks such that each individual was recorded once every day for a week. The recorded videos from the two groups (patient and HC) were fed to the CV-based system to produce and analyze the tapping signals. A total of 471 videos, 387 patient (i.e. 23 days x 17 videos/day=391 videos; 4 bad quality videos were removed) and 84 HC (i.e. 7 days x 2 video/day x 6 subjects = 84 videos) were utilized for tapping classification.

4.2.2.2 System description

The first step in the processing of RFT videos was to detect the subject's face in the video-frame. Next, this video-frame was split into two images from the center coordinates of the detected face-rectangle. Two regions of interest were located in each of the images representing areas of index-finger movement of the left and right hands respectively. In the next step, the motion of the moving index fingers of both hands was estimated in each region. The movement of index fingertip coordinates produced a tapping time-series which was normalized using the face-rectangle height. The baseline of this time-series was calibrated to eliminate the effects of varying hand position. Tapping features representing the UPDRS-FT symptoms of speed, amplitude, fatigue and arrhythmia were extracted from the right (dominant) hand tapping time series and were utilized for symptom analysis and classification.

4.2.2.3 Feature analysis

The μ_2 was utilized to map between the tapping features and the clinical ratings from the two raters. Jackknifing was used as a cross-validation method to estimate the precision of μ_2. The jackknife estimates of μ_2 are presented in section 5.4.1. According to this analysis, some of the tapping features showed a perfect correlation between each other. These features were removed from the final list of features used in the classification of symptom severity.

4.2.2.4 Classification between UPDRS-FT levels

The SVM described in section 3.2.1 was utilized for tapping classification. Three different classification experiments were performed on the tapping features selected using the Guttman correlation analysis and chi-squared test. In the first experiment, the selected features were used to classify between UPDRS-FT severity levels. These classification tests were performed sepa-

rately for rater-1 and rater-2 using their ratings as classification targets. In the second experiment, the classification test was performed on the tapping features to separate between PD and HC samples. Further tests were made in the third experiment to compare between the proposed CV-based SVM classification scheme and a sensor-based log-linearized Gaussian mixture networks (LLGMN) classification scheme reported by Shima et al. (2009). Additionally, the AoC was utilized to analyze the feasibility of the classification models. The results are given in section 5.4.2.

4.2.3 Computer vision methods for Parkinsonian gait analysis: a review of patents (Paper 6)

An important application of neurological diagnosis and management is CV based gait analysis that is able to provide first-principle data on postural and locomotive abnormalities (Green et al., 2000). A systematic literature review on methods for gait disorder analysis was performed. The feasibility of marker-less CV based systems was examined for their use in the at-home evaluation of gait. This analysis takes into account the physical restrictions of patients that arise due to PD. Other objectives of the review were to evaluate cost and resource efficiency in gait analysis, portability of the equipment and comfort in carrying out the experiments.

The CV based systems for gait analysis can be characterized into two types; holistic-based gait analysis (HGA) and model-based gait analysis (MGA). In the HGA, diagnosis is performed with a marker-free system. In this approach, posture and motion features are extracted from the segmented silhouette of a test subject using image processing techniques. Importantly, biometric features can be computed (Huang et al., 1999) to compare between normal and pathological gait patterns. By contrast, the MGA uses markers placed on the human body to calculate temporal characteristics of gait. Stride frequency, cadence and velocity can be measured using MGA (Melnick et al., 2002). These features are extracted from stick figures and skeletons produced by the placement of markers on the body (Cunado et al., 1999).

Wearable accelerometers and sensors are other methods used in gait analysis. A problem in using the sensors and accelerometers, as well as in using the MGA systems is that, they require a subject to wear obtrusive equipment that impedes body movement. Also, the analysis using this equipment requires significant setup time and additional cost (Cho et al., 2009). These problems can be tackled using the HGA systems as these systems do not require physical contact. Especially in biomedical applications (like rehabilitation monitoring after orthopedic surgery and neurological disorder assessment) the hardware setup can be effectively reduced through marker-free gait analysis, allowing the patient maximum possible amount of comfort during the examination.

The study of literature revealed that the main challenge in developing an HGA system is the process of image segmentation. Previous research suggests that color or grey level segmentation methods for separating a human silhouette from the background in the video frames are flawed because of artifacts such as bad illumination or object occlusion. Inaccuracy in segmentation leads to coarse gait features that further leads to incorrect gait classification. Microsoft Kinect is the latest technology which provides an alternative solution to the segmentation problem. However, the sensor in its current form cannot assess sideways body movements, which is essential to locate leg bendings, and specifically the flexions in the knees and hip.

An important aspect in review methodology was to explore gait features that can possibly be extracted using the HGA systems. Among different gait features, biorhythmic features were successful in classifying neuromuscular symptoms. Other relevant features were stride details, cadence and angular movements of joints during locomotion. In one study (Wang et al., 2009) walking speed, step-length and step-time were extracted from a HGA tool and a GAITRite (sensor-based) mat simultaneously. Comparisons showed that HGA systems can perform as equally well as the sensor-based systems in analyzing gait in the home environment.

In the reported methods on the marker-based systems, although a comparative analysis between pathological and normal gait features was done, however, none of these methods could classify gait disorder on a clinical rating scale. This was possibly due to the fact that not enough clinical details can be obtained from stick figures. Notably, important information like the body's COG or lean in posture was required to detect anomalies in gait. Some MGA systems used illuminated markers along with foot pressure sensors to estimate gravitational artifacts. Some other methods used pressure sensors attached to the conveyer belt of a treadmill. All these requirements make MGA expensive, non-portable and time consuming to operate.

In contrast to the obtrusive gait analysis tools, marker-less systems can provide a comfortable, cost-efficient and user-friendly setup which can be employed at home since these systems require just a single web-cam or a Kinect sensor for gait analysis. In paper 7, a prototype for a gait recognition system is presented. This system is able to estimate posture lean, stride fluctuations as well as the COG of a human silhouette. A preliminary test on sample videos showed that this algorithm can discriminate between pathological and normal gait patterns with high accuracy.

4.2.4 Motion cue analysis for Parkinsonian gait recognition (Paper 7)

The first step in the image processing of recorded gait videos was the segmentation of the foreground subject from background in video frames based on pixel brightness. Image enhancements were done to extract a 2D sil-

houette of the subject in each frame. In the next step, the silhouette was isolated using a bounding rectangle to estimate its height and width (figure 13b). Once the height and width of this silhouette was computed, it was compared with a human model (Tafazzoli & Safabakhsh, 2010) to separate the head, torso and leg segments (figure 13c). A skeleton was formed by computing the COG of the human silhouette and the medial points of the silhouette segments (figure 13d). The movement of this skeleton between the video-frames was used to estimate clinical gait features. Two clinical features that were analyzed were the cyclic motion of legs and the posture lean of the subject between the video-frames (figure 13e).

The main idea behind the formulation of this gait model was to compute cosine similarity between the subject's gait pattern and an imaginary perfect gait pattern. Accordingly, the values of stride angle θ_{stride} and posture lean angle θ_{lean} for a perfect gait, represented as θ_{PS} and θ_{PL} respectively, were derived using the gait dimensions reported by Murray et al. (1966). Based on the assumption that an imaginary perfect gait exhibits a constant stride frequency with a stride angle $\theta_{PS} = 45^{o}$, the subject's strides can be matched with the perfect strides by computing a cosine distance d_i between θ_{stride} and θ_{PS} for gait cycle i as given in equation 20.

$$d_i = \left|\cos\theta_{stride_i} - \cos\theta_{ps}\right| \qquad \textit{for, } i=1,\ldots,n \qquad (20)$$

Where n is the total number of gait cycles. This distance equation suggests that the larger the value of d_i, the larger will be the cosine difference between the patterns of the subject's and the perfect gait. Stride variability in the subject's gait pattern can be represented by computing the mean of the overall distance $d_1 \ldots d_n$ as given in equation 21.

$$d_{\text{avg}} = \frac{\sum_{i=1}^{n} d_i}{n} \qquad \textit{for, } i=1,\ldots,n \qquad (21)$$

A periodic gait pattern should yield a constant interval between each gait cycle. Time variation between gait cycles can be computed by calculating the residual time between the intervals. The time residuals r_i are computed by subtracting between the time interval t_i in a gait cycle i and the average of total intervals t_{avg} for gait cycles n (equation 22). The mean r_{avg} of residual values $r_1 \ldots r_n$ for gait cycles n (equation 23) is yet another feature for assessment of gait anomalies as it detects shuffling and festination in gait.

$$r_i = \left| t_i - t_{avg} \right| \qquad \textit{for, } i=1,\ldots,n \qquad (22)$$

$$r_{\text{avg}} = \frac{\sum_{i=1}^{n} r_i}{n} \qquad \textit{for, } i=1,\ldots,n \qquad (23)$$

Apart from computing the stride features, the cosine of posture lean angle θ_{lean} was used to produce posture patterns for the normal and pathological gait respectively. It was assumed that a perfect gait exhibits an erect posture throughout the gait with angle $\theta_{lean} = 0^{o}$, denoted as θ_{PL}. Matching between the patterns of a subject's leaned posture θ_{lean} and a perfect erect posture θ_{PL} was performed by computing the cosine distance l_j between them as given in equation 24.

$$l_j = \cos\theta_{PL} - \cos\theta_{lean_j} \qquad \textit{for, } j=1,\ldots,m \qquad (24)$$

Where, m is the total number of gait video-frames. The mean l_{avg} of cosine distance $l_1 \ldots l_m$ was computed to yield a single value to represent posture lean (equation 25).

$$l_{\text{avg}} = \frac{\sum_{j=1}^{m} l_j}{m} \qquad \textit{for, } j=1,\ldots,m \qquad (25)$$

A total of 49 video recordings (7 samples per subject) from 4 normal controls and 3 PD patients, were utilized to compute gait features including stride variability d_{avg}, residual mean r_{avg} and posture lean l_{avg} respectively. The gait recordings of patients were acquired from the training video library provided by the Movement Disorder Society. The feature values were normalized within a range 0-to-1. An SVM classifier based on recursive feature elimination (Guyon et al., 2002) was used to weight gait features using 10-fold cross validation. Based on the given dataset, the classifier ranked the features l_{avg}, d_{avg} and r_{avg} as 3, 2 and 1 with the weights as 2.5, 2.5 and 1 respectively. A formula to compute gait error E was devised using the weighted average of gait features as given in equation 26. The difference of E from a perfect gait G yields a gait match percentage (equation 27). A comparison between the gait parameters of normal controls and PD patients is given in section 5.5.

$$E = \frac{2.5l_{avg} + 2.5d_{avg} + r_{avg}}{6} \qquad (26)$$

$$G\% = (1 - E) \times 100 \qquad (27)$$

5. Results and analysis

This section presents the results and analysis obtained on applying the methodology proposed in chapter 4 to the clinical dataset, including the audio recordings of speech (section 4.1.2), and the video recordings of finger-tapping (section 4.2.2.1) and gait (section 4.2.4) motor exams.

5.1 The MFCC analysis of pathological speech

5.1.1 Feature validation

A higher number of TRS-MFCC have shown a stronger ($\mu_2 > 0.5$) statistically significant ($p<0.05$) correlation with the UPDRS-S ratings in all the TRS tests as compared to the MFCC computed from recordings of the SVP and DDK tests. Specifically, the 4th TRS-MFCC produced the highest correlation estimates in TRS test-1 ($\mu_2 = 0.60$, $p = 0.004$), test-2 ($\mu_2 = 0.67$, $p = 0.001$) and test-3 ($\mu_2 = 0.70$, $p < 0.0001$) respectively. Importantly, the value of μ_2 between the 4th MFCC and UPDRS-S increased and the p-value decreased from TRS test-1 to test-3, suggesting that the higher demand of linguistic stress in a speech test results in a higher degree of precision when estimating speech symptoms.

5.1.2 Test retest reliability

The MFCC groups computed from TRS test occasions showed the highest ICC, both in the single and average measurements, as compared to the MFCC groups that were computed from the SVP and DDK test occasions. Specifically, the group representing the 8th order TRS-MFCC showed the highest values, in both single (ICC = 0.94) and average (ICC = 0.98) ICC measurements, on different test occasions of TRS. The mean between the single ICC measurements of the MFCC groups were 0.64, 0.80 and 0.87 for test occasions of SVP, DDK and TRS respectively. Similarly, the mean between the average ICC measurements of the MFCC groups were 0.87, 0.89 and 0.95 for test occasions of SVP, DDK and TRS respectively. Noticeably, the MFCC groups from SVP recordings were the most inconsistent.

5.1.3 Classification

First-experiment: Classification using 10-fold cross validation:

In the first classification experiment, the MFCC computed from samples of TRS tests 1, 2 and 3 were merged together to form a classification matrix with the dimensions of 16 MFCC x 240 samples (80 subjects x 3 TRS tests). Data stratification with 10-fold cross validation on this input vector in SVM produced an overall classification accuracy of 78% with TPR of 80%, 60% and 90% for classes '0', '1' and '2' respectively. In a similar test on the MFCC computed from the DDK samples, the 10-fold cross validation on the classification matrix with the dimensions of 16 MFCC x 160 samples (80 subjects x 2 DDK tests) produced a classification rate of 66% with TPR of 54%, 75% and 68% for classes '0', '1' and '2' respectively. Likewise, in a further classification test on the SVP-MFCC, the 10-fold cross validation on the input vector with the dimensions of 16 MFCC x 320 samples (80 subjects x 4 SVP tests) produced a classification rate of 83% with a TPR of 77%, 80% and 90% for classes '0', '1' and '2' respectively.

Second-experiment: Classification using training/testing sets:
In case of the TRS, the SVM trained using the matrix of 16 MFCC x 160 (TRS tests 1 and 2) samples and tested using the matrix of 16 MFCC x 80 (TRS test-3) samples classified between the UPDRS-S severity levels with an accuracy of 74%, with a TPR of 64%, 88% and 71% for classes '0', '1' and '2' respectively. In the case of the DDK, the SVM trained using the matrix of 16 MFCC x 80 (DDK test-1) samples and tested using the matrix of 16 MFCC x 80 (DDK test-2) samples, classified between the UPDRS-S severity levels with an accuracy of 66%, with TPR of 60%, 71% and 68% for classes '0', '1' and '2' respectively. Likewise, in the case of the SVP, the SVM trained using the matrix of 16 MFCC x 160 (SVP tests 1 and 2) samples and tested using the matrix of 16 MFCC x 160 (SVP tests 3 and 4) samples classified between the UPDRS-S severity levels with an accuracy of 68%, with a TPR of 62%, 67% and 74% for classes '0', '1' and '2' respectively.

In both these classification experiments, first using the 10-fold cross validation scheme and second using the training and testing sets, the MFCC from the TRS samples produced on average the highest accuracy rate (76%) in classifying between the 3 levels of the UPDRS-S as compared to the MFCC from the SVP and the DDK. Importantly, the classification rates produced by the TRS-MFCC in both these experiments were consistent.

ROC analysis:
In the two classification experiments, first using 10 fold cross validation and second using training-testing sets in the SVM, the TRS-MFCC produced the average AoC of 85% and 84% respectively, suggesting that this scheme can yield an 'excellent classification model' for categorizing speech symptom severity. Importantly, the AoC for all severity classes, in both experiments,

were above 80%, indicating the indubious distinction of samples in each severity class. Moreover, the values of the AoC for each severity class were consistent in both experiments.

By contrast, although the SVP-MFCC produced a high average AoC (88%) in the 10-fold cross validation scheme, this value decreased to 77% when testing-training sets were utilized, suggesting that the change in standard loudness levels in SVP tests may demote the performance of the MFCC in categorizing symptom severity. Additionally, although the values of the average AoC produced by the DDK-MFCC were consistent (77%) in both classification experiments, the results do not suggest that DDK-MFCC together with SVM can yield an 'excellent classification model' for separating symptom severity.

5.2 The CSD analysis of pathological speech

5.2.1 Test of validity

The CSD-based features δ_{CSD} and AP_{CSD} showed a strong statistically significant ($p<0.05$) correlation ($\mu_2>0.5$) with the UPDRS-S ratings in all of the TRS tests. Specifically, the δ_{CSD} showed the highest correlation with the speech ratings, i.e. 0.70, 0.74 and 0.78 in TRS tests 1, 2 and 3 respectively. Noticeably, this correlation improved with the increasing textual difficulty from TRS test 1 to test 3. An important characteristic of δ_{CSD} was that the standard errors were decreasing with the increasing textual difficulty, suggesting that the δ_{CSD} can be an even better discriminator of speech symptom severity levels if the subjects are given a more difficult text to read.

The comparison between the CSD and non-CSD features suggests that although some of the non-CSD features such as H1-H2 and means of the 3rd and 4th MFCC showed a high correlation with the speech ratings in some of the TRS tests, but only the mean of the 4th MFCC showed an increasing correlation relative to the increasing textual difficulty from TRS test 1 to test 3. Besides, the HNR was weakly correlated with the speech ratings in all of the TRS tests.

5.2.2 Test of reliability

According to the ICC analysis, the AP_{CSD} was the most consistent feature than other speech features and showed very high values (>0.94) of single and average ICC measurements. Importantly, the ICC was found higher (>0.97) in symptom severity class '2' proving that the AP_{CSD} remains consistent specifically when measuring severe speech symptoms. Likewise, the δ_{CSD} showed high values (>0.8) for single and average ICC measurements and these correlations were higher in severity class '2'.

By contrast, the ICC values of H1-H2 and HNR were low. Besides, the ICC values of the means of the 3rd and 4th MFCC were high among the

symptom classes. Nevertheless, the high ICC of CSD features in different test scenarios depicts that the CSD features remain consistent irrespective of the kind of TRS being used to assess speech symptoms.

5.2.3 Test of repeatability

The CSD features δ_{CSD} and AP_{CSD} remained efficient in discriminating between speech symptom severity groups in different text settings as is proved by the group mean values and standard errors in ANOVA. The feature values increased monotonically with the increasing UPDRS-S severity groups in all TRS tests, with statistical significance of having different group mean values ($p < 0.0001$). These group mean values were repeatable for the same severity group in different TRS tests.

5.3 Classification between UPDRS-S levels

5.3.1 Textual difficulty vs. classification performance

In the first classification test taking TRS test-1 into account, the SVM classified the samples into 3 severity levels with a classification accuracy of 63% and TPR of 63%, 40% and 81% for classes '0', '1' and '2' respectively. The low TPR for classes '0' and '1' indicate the difficulty of discriminating between the normal and mildly impaired speech samples. One reason could be that the textual difficulty in TRS test-1 paragraph was not strong enough to stress the mildly impaired subjects, who comfortably read the passage without displaying any symptom and were thus classified as normal. Additionally, the high TPR for class '2' indicates that the more severely impaired subjects exhibited reading difficulty even in this low-stress setting and revealed symptoms which were effectively quantified.

In the second classification test, the increased textual difficulty in TRS test-2 improved the classification accuracy to 70% and TPRs to 60%, 72% and 78% for classes '0', '1' and '2' respectively. Upgraded TPRs for classes '0' and '1' indicate the improved ability of features in discriminating between the samples in class '0' and '1'. This finding suggests that the mild symptoms which remained hidden in TRS test-1 were detected by the more demanding level of reading difficulty in TRS test-2. The findings were confirmed in the third classification test when the subjects were exposed to the highest textual difficulty. The speech samples were classified with a marked improvement in the classification accuracy (84%) and the TPRs for classes '0' (75%), '1' (76%) and '2' (97%) respectively.

In all the three classification tests, the averaged AoC remained above 75% ('good' model) and improved with the level of textual difficulty i.e. 76% ('good model'), 80% ('very good model') and 91% ('excellent model') in TRS tests 1, 2 and 3 respectively.

5.3.2 Classification performance in the complete dataset

5.3.2.1 Using 10-fold cross validation:

Data stratification with 10-fold cross-validation on the input vector with the dimensions of 13 (features) x 240 (total samples) produced a high classification rate of 83% with TPRs of 84%, 76% and 87% for classes '0', '1' and '2' respectively. The averaged AoC was 89%. Specifically, the marked distinction of samples belonging to group '2' supports that the selected measures are representative features of speech symptoms in PD.

5.3.2.2 Using training/testing sets:

A high accuracy (82%) was achieved by this scheme in classifying the test set between the 3 levels of UPDRS-S with an averaged AoC of 90%. Specifically the samples in class '2' were predicted again with a very high TPR (100%).

5.4 The computer-vision analysis of finger-tapping

5.4.1 Feature validation:

Nearly all the tapping features showed a strong correlation with the clinical ratings provided by rater-1. Specifically, the features representing arrhythmia and fatigue were strongly significantly ($p < 0.0001$) correlated ($\mu_2 < -0.8$). Additionally, the features representing speed and amplitude showed moderate correlation with rater-1.

In comparison to rater-1, although the representative features of arrhythmia and fatigue were moderately correlated with the clinical ratings provided by rater-2. However, the features representing speed and amplitude were weakly correlated.

Despite the differences between the raters, the fatigue and arrhythmia features were either very strongly correlated with rater-1 or moderately correlated with rater-2. Some of these features were perfectly correlated between each other. For example, tapping speed was perfectly correlated with opening and closing velocities. Likewise, some of the features of fatigue were perfectly correlated between each other. Redundancies were eliminated by choosing that feature which has the higher sum of absolute correlation values between rater-1 and rater-2. The remaining features were used in tapping classification.

5.4.2 Tapping classification:

5.4.2.1 Experiment 1: Classification between UPDRS-FT levels

Two different classification tests were performed using the clinical ratings from rater-1 and rater-2 as targets in test-1 and test-2 respectively. An accu-

racy of 88% was achieved in test-1 for classification between the 3-symptom severity levels of the UPDRS-FT using 10-fold cross validation. High TPR were achieved for classes '0' (90%), '1' (89%) and '2' (85%) respectively. Importantly, the AoC for severity classes '0', '1' and '2' were 95%, 88% and 97% respectively. The averaged AoC was 93%.

In contrast to test-1, the classification accuracy was low in test-2 (76%). Despite of the fact that the averaged AoC was high (83%), a low TPR (53%) was observed for the classification of samples in severity class '0'. One possible reason could be that, rater-2 had rated the samples with the knowledge that the patients were in the advanced stage of PD and with a low expectation that they could perform well on the tapping test.

5.4.2.2 Experiment 2: Classification between PD and HC samples

An accuracy of 95.8% was achieved for classification between PD and HC samples using 10-fold cross validation. The TPR were high for both HC (83%) and the patient group (98%).

5.4.2.3 Experiment 3: Comparison between SVM and LLGMN schemes

In different classification experiments, the proposed CV-based SVM scheme discriminated between PD and HC samples with an average accuracy of 94.5% which was higher than the 93.1% classification accuracy produced by the sensor-based LLGMN scheme.

5.5 The computer-vision analysis of gait

Several gait cycles of a group of four normal controls n_1... n_4 and three PD patients p_1, p_2 and p_3 were assessed using the described gait analysis tool. The gait patterns from the normal controls showed a high percentage of gait matching $G\%$ with the perfect gait pattern. The normal controls n_1, n_2 and n_3 exhibited low stride variability d_{avg} which resulted in higher $G\%$ values of 92%, 87% and 88% for n_1, n_2 and n_3 respectively. By contrast, the gait pattern of n_4 showed high stride variability d_{avg} which resulted in a low $G\%$ value of 71%. All of the normal controls n_1 … n_4 produced low values of posture lean l_{avg} and residual mean r_{avg}.

In the case of PD patients, the gait matching between the pathological gait patterns and the perfect gait pattern was fairly low i.e. $G\%$ was equal to 16%, 7% and 58% for p_1, p_2 and p_3 respectively. The subject p_3 showed a low value of lean in the posture ($l_{avg} = 0.033$) which resulted in gait matching of 58%. Preliminary analysis suggests that high accuracies (up to 100% for this dataset) in identifying pathological gait can be achieved by using an appropriate threshold value of G%.

6. Discussion, conclusions and future work

This chapter summarizes the RCs in this thesis and discusses the issues related to the justification of developing and using different FPDD models for assessment of motor disorders in PD. Conclusions drawn from this work are discussed together with the limitations and future potential directions in which this research can be further pursued.

The thesis focused on investigating and developing new AI methods for the classification of movement disorders related to PD. The study investigated the potential of using biomedical signal and image analysis to quantify and classify symptom severity as defined by the standard reference clinical metric UPDRS. Towards this aim, the primary emphasis was on developing and investigating the FPDD model for speech that can extract clinically useful information from recorded speech signals and identify symptom severity levels according to UPDRS-S. Apart from speech, it is also important to quantify the severity of gross and fine motor symptoms to accurately and completely determine the disease profile. For this reason, the CV algorithms based on the FPDD approach were developed to identify gross motor symptoms through gait analysis and fine motor symptoms through finger-tapping analysis.

In order to develop new methods, the literature reviews and discussions with specialists were conducted to study existing computer tools and clinical procedures for the assessment of PD symptoms. Here the author had the opportunity to interact with domain experts in the field that helped in understanding the clinical knowledge and content of the domain.

During the research, several RQs were addressed and consequently their solutions with new findings were achieved that contributed to an overall accomplishment of this thesis. The nature of selected PD symptoms was analyzed carefully in order to make an appropriate selection of first-principle parameters for accurate and faster convergence of data-driven decision making. Also, keeping in view the physical limitations of patients, unobtrusive techniques for data-acquisition (i.e. audio-video recording of patient) were adopted to capture the natural state of the patient.

6.1 Result related issues

The main objective of this research was to develop and evaluate novel computer methods that can assist clinicians in quantifying the severity of motor

symptoms in PD in order to support the clinicians in devising appropriate treatment plans for symptom alleviation. Considering the aspects of cost and logistics in treatment, and the physical limitations of patients, remote monitoring solutions were investigated by incorporating speech and video data analysis. Importantly, this thesis investigates FPDD models for clear interpretability of results and for fast and accurate convergence of clinical decision making. In order to attain this objective, the research was split into three different RQs as presented in table 1. The answers to these questions resulted in the following RCs:

RQ 1: What computer methods can be used to quantify motor speech, gross motor (gait) and fine motor (finger-tapping) symptoms in PD, and what are their limitations? To answer this question:

- A systematic literature review (section 4.1.1 [Paper 1]) was done to study methods and parameters that are generally used for measuring motor speech symptoms. The main finding of this review was that the previous algorithms used SVP for speech assessment due to which these algorithms were unable to capture important characteristics of speech intelligibility (section 2.3.4). Due to technical complexities in processing continuous speech signals (section 4.1.1), these algorithms avoided using TRS, which according to Zraick et al. (2003) is important for evaluating speech using the UPDRS motor examination. Specifically, since speech is multi-dimensional, the processing of TRS is essential to cover broader aspects of PD speech symptoms. Another finding of the review was that the SVM classifiers were used only to discriminate between healthy and pathological speech samples, and not to discriminate between speech symptom severity levels.

- Another review study was done (section 4.2.3 [Paper 6]) to study existing computer methods and tools used for evaluating gross motor symptoms in gait. The main finding of this review was that the CV based gait analysis is restricted to produce accurate symptom estimation due to artifacts in image segmentation. The Microsoft kinect can be an alternative solution; however in the current form, it cannot assess sideways body movements, which makes it unfeasible to estimate posture lean and stride variability in gait.

- A review of existing methods used to quantify fine motor symptoms in finger-tapping is presented in section 4.2.1. The main finding of this review was that the available CV methods were unable to compute tapping symptoms according to the UPDRS-FT, since these methods used platforms where subjects were to place their hands to perform tapping. Due to this limitation, these systems were unable to estimate PD symptoms

when a subject is in the state of dyskinesias, where he is unable to keep his hands consistently at one position.

In addition, meetings with clinicians were held to apprehend how the chosen symptoms are evaluated in clinical practice. Posters were presented at international conferences to discuss and share ideas with the domain experts, to understand the application domain, and to determine and analyze suitable computer methods that can quantify gross motor, fine motor and motor speech symptoms in PD.

Nevertheless, the study of previous computer applications revealed that most of these applications were restricted due to technical specifications, experimental setups and prior training for clinicians and patients to carry out clinical operations. The review findings suggest that there is a need to develop computer methods, compatible with clinical standards, as well as methods that satisfy some transparency in data analysis in order to allow easy interpretation of results by non-technical users, such as clinicians. In this thesis, on one hand, this transparency was achieved by computing first principle parameters from audio and video recordings, where clinicians can watch or hear the recorded data several times to compare the measured parameters. On the other hand, using the UPDRS as a biomarker allows maintaining the clinical standard in symptom evaluation.

RQ2: How can we develop novel computer methods that address the limitations identified in RQ-1 by feature extraction based on first principle models, to quantify motor speech, gross motor (gait) and fine motor (finger tapping) symptoms in PD, and compare the features to a standardized rating scale? To answer this question:

Considering the limitations identified in RQ-1, new methods employing feature extraction based on first principle models were developed to quantify motor speech (section 4.1.4 [Paper 3]), fine motor (finger-tapping: section 4.2.2 [Paper 5]) and gross motor (gait: section 4.2.4 [Paper 7]) symptoms in PD.

- In the case of speech, the CSD algorithm [Paper 3] utilizes a source-filter model of speech production (section 3.1.1) that enables first-principle quantification of speech symptoms. The main finding of this work was that, the CSD features were the most strongly significantly correlated features with the UPDRS-S as compared to the other non-CSD speech features. Since the CSD features AP_{CSD} and δ_{CSD} were used to represent symptoms of 'hoarseness' and 'harshness' in speech respectively that arise as a result of abnormal glottal constrictions. This infers that the obtained results are in agreement with Midi et al. (2008) who claimed that unintended glottal closure pattern is the most common ma-

nifestation in PD speech. Another important finding in this paper was that, although the symptom of 'hoarseness' ('soft speech') may not disturb the speech intelligibility with the same intensity as the symptom of 'harshness' (rough speech), the strong ICC values of AP_{CSD} suggests that 'hoarseness' is the most consistent attribute in PD speech. This finding is in line with Stewart & Selesnick (2010) who suggested that 'hoarseness' is the most constant chronic symptom in the speech of PD patients. An advantage of using CSD features is that they were developed to estimate speech symptoms using TRS which enables quantifying speech symptoms according to the standard UPDRS-S (a solution to the limitation identified in RQ-1).

- In the case of finger-tapping, the motion gradient algorithm (section 3.1.2) was utilized to compute first principle motion parameters of index finger to quantify fine motor symptoms in tapping. The main finding of this paper was that, the proposed CV algorithm can identify tapping symptoms as equally well as any marker based or sensor based system. Specifically, the practical experiments on wearing markers showed that a marker-free algorithm using motion detection can locate the moving index fingertip position with the same pixel accuracy as it can be located using a marker-based algorithm. The proposed system has another important advantage that it can evaluate tapping symptoms even if the patients are in the state of dyskinesias as a fact that the system is platform independent (a solution to the limitation identified in RQ-1). Moreover, the system utilizes recorded videos of finger-tapping motor examination that allows estimating tapping symptoms according to the standard UPDRS-FT. Importantly, the data analysis showed a strong statistically significant correlation between the proposed tapping features and the UPDRS-FT.

- In the case of gait analysis, a physiological human model was utilized to extract first-principle features of locomotion (section 3.1.3) to quantify gross motor symptoms in gait. The main finding of this work was that the extracted gait features representing stride variability and posture lean were capable of discriminating between pathological and normal gait with a high degree of accuracy. Importantly, the method was able to assess sideways body movements during gait (a solution to the limitation identified in RQ-1). However, the results need to be validated using a larger dataset of recorded gait videos. Moreover, the videos need to be rated by a clinician using the UPDRS motor examination of gait and posture, so that the ratings can be utilized to train the algorithm to classify the severity of gait and posture symptoms.

Nevertheless, the strong significant Guttman correlation between the proposed features of speech and finger-tapping, and the UPDRS motor examination suggests that these features are valid representatives of PD symptoms. Importantly, the data analysis using ICC, ANOVA and chi-squared statistics suggests that the feature quantities were consistent and reproducible in different test occasions and experimental settings.

RQ3: How can we develop and evaluate computer systems that allow automatic classification of PD symptom severities using first-principle data-driven models and unobtrusive data acquisition techniques? To answer this question:

- A system based on first-principle features of speech and data-driven SVM was developed (section 4.1.5 [Paper 4]) for classification of symptom severity according to the UPDRS-S. An important finding of this paper was that the accuracy in classifying symptom severity increases relative to the increasing textual difficulty in paragraphs given to the subjects for recitation. Importantly, the algorithm was able to compute deficits in sub-systems of speech that allow the establishing of a relationship between the malfunctioning of specific speech organs and the severity of speech symptoms. Moreover, the results using n-fold cross-validation and training/testing experiments suggest that the chosen (first-principle) features and the (data-driven) SVM model have a strong ability to discriminate between UPDRS-S severity levels. Besides, the audio recording of speech allows unobtrusive data acquisition.

- A system based on first-principle motion parameters of index-finger and data-driven SVM was developed (section 4.2.2 [Paper 5]) for classification of tapping symptom severity according to the UPDRS-FT. The algorithm was able to classify symptom severity levels with a high degree of accuracy. Importantly, the selected features were able to represent all of the symptoms on the UPDRS-FT scale including slowness in speed, amplitude reduction, fatigue and arrhythmia. Since some of these features were previously used by a sensor-based system (Shima et al., 2009), our experiments showed that the same features when computed using the proposed CV system could classify between UPDRS-FT levels with the same classification accuracy as the sensor-based system. Besides, the utilization of a CV approach allows unobtrusive data collection. An important finding was that the features of speed and amplitude were ranked high by the chi-squared algorithm when they were selected to classify between HC and PD samples. By contrast, the same features were ranked low when they were selected to classify between UPDRS-FT severity levels. On the other hand, rhythm and fatigue features were ranked high when they were selected to classify between UPDRS-FT le-

vels and were ranked low when they were selected to discriminate between HC and PD samples. These findings are in agreement with Kimber et al. (1999) and Bazner et al. (2005) who suggested that decay in speed and amplitude in tapping is a cardinal clinical feature of bradykinesia that principally distinguish PD subjects from HC. Besides, when discriminating between symptom severity levels, the high chi-squared values of rhythm and fatigue features are in line with the UPDRS motor examination of finger-tapping according to which the rhythm and fatigue symptoms are more evident in the advanced clinical stage, and therefore are evaluated in the 'moderate' and 'severe' levels of the UPDRS-FT, but not in the 'normal' and 'mild' levels.

- A comparative analysis was done using MFCC computed from recordings of TRS, SVP and DDK tests, to assess their feasibility to be used in examining speech (section 4.1.3 [Paper 2]). Supported by the claims (Kwon & Lee, 2004; Stevens & Volkmann, 1940) that the mel-filters provide a simulation of human auditory system (section 3.1.1.2), that enables quantifying the (clinician's) perception of (patient's) speech. The findings in this paper suggest that a larger amount of impairment related information can be captured using MFCC from TRS signals, rather than by using MFCC from the SVP and DDK. Our analysis has shown that MFCC from TRS remain more consistent and reproducible in different test settings and have a stronger statistical correlation with the UPDRS-S as compared to the MFCC from the SVP and DDK. The accuracy of classifying speech symptom severity using the TRS-MFCC was higher and this accuracy remained consistent when a different type of paragraph was utilized for recitation. The findings serve as a basis for the utilization of TRS for computerized assessment of clinical speech symptoms. This paper contributes to two RQs. (1) It contributes to RQ-1, since the MFCC, previously known for their capability to estimate speech dysfunctions, were studied and practically applied to assess speech symptoms in PD. (2) It contributes to RQ-3, since, prior to this research, MFCC had never been computed from TRS signals to classify speech symptom severity in PD.

Nevertheless, the high values of AoC achieved by the proposed algorithms strongly support the feasibility of the selected pool of features and the SVM model for classification of symptom severity in PD.

6.2 Significance of work

According to the previous research, the symptoms in speech, gross and fine motor functions have the strongest influence on the quality of life of patients with PD (Gomez et al., 2007; Behari et al., 2005; Damiano et al., 1999;

Schrag et al., 2000; Brozova et al., 2009). In this thesis, this served as the main motivation for selecting the PD dimensions of speech, gross and fine motor functions for research and development. The secondary motivation of selecting these dimensions is that the computer algorithms that can assess speech, gross and fine motor functions have a high general applicability and have the ability to solve problems in domains other than in medicine.

6.2.1 Significance of speech algorithms

Sources suggest that impairment in speech has the worst impact on the quality of life, since the patients become less able to communicate verbally to the caregiver, to ask for general needs and to express their emotions. In a study by Hartelius and Svensson (1994), 29% of the patients consider vocal impairment as the worst restriction associated with the disease, which reduces their social interaction. In another study, Jankovic (2008) suggested that speech dysfunction in patients with PD can be even more disabling than the cardinal symptoms. In a further study by Gomez et al. (2007), when the patient-assessed PDQ-39 (PD questionnaires) scores were correlated to the sub-scores of UPDRS part III to determine which UPDRS III items contribute most to the worsened quality of life, the speech item 18 was found to be the second most correlated to the PDQ-39 scores (Pearson's $r=0.43$, $p<0.001$) after gait (Pearson's $r=0.52$, $p<0.001$).

Keeping in view that the speech processing algorithms presented in paper 2 (section 4.1.3), 3 (section 4.1.4) and 4 (section 4.1.5) have shown a strong ability to quantify and classify speech symptoms, these algorithms can be useful to the clinicians for tracking speech symptoms and to help them directing appropriate treatment. Specifically by involving a FPDD model with an ability to locate symptoms in the subsystems of speech, clinicians can adopt appropriate methods of speech therapy to help patients regain their communication skills which will enable them to express their wants, needs and ideas. This will consequently improve their quality of life.

6.2.1.1 General applications of speech algorithms

The study of the vibration of vocal fold patterns in speech algorithms, in particular CSD, includes the evaluation of biomechanical parameters that are susceptible to malfunction in the presence of pathology. As discussed in section 2.3.2, the glottal constrictions and the abnormal increase in aspiration noise can be a result of vocal fold tissue problems such as nodules, or can be symptoms of vocal fold cancer (Dubuisson, 2011). This suggests that the proposed CSD algorithm can be used to diagnose and evaluate other serious diseases, such as cancer. Some other areas of the application of speech algorithms include speech recognition systems and speech compression for transmission purposes etc.

6.2.2 Significance of finger-tapping algorithm

One of the cardinal symptoms of PD is bradykinesia (slowness of movement) that greatly affects the quality of life of patients with PD. The patients find it difficult to dress themselves, to pick up utensils, to hold a pen or to play a piano etc. In one study (Rahman et al., 2008), the 'slowness of movement' and 'difficulty in dressing' were significantly correlated ($p<0.01$) with the mean quality of life scores in PDQ-39. Likewise, in the study by Gomez et al. (2007), the UPDRS III score of bradykinesia measured using the finger-tapping test was significantly correlated (Pearson's $r=0.40$, $p<0.001$) with the mean PDQ-39 scores. In this regard, the proposed finger-tapping algorithm [paper 5] can support clinicians in evaluating and quantifying fine motor symptoms in tapping.

6.2.2.1 General applications of finger-tapping algorithm

Apart from PD (Volkow et al., 1998), the finger-tapping test is widely utilized to assess neuromuscular coordination in many other disorders including acute strokes (Heller et al., 1987), the testing of subjects with Korsakoff's alcoholic syndrome (Welch et al., 1997), the characterization of upper limb motor functions (Giovannoni et al., 1999), and in particular to quantify neurological disorders such as Alzheimer's disease (Ott et al., 1995), ataxia (Notermans et al., 1994) and schizophrenia (Green et al., 2004). Hollingworth (1914) used a tapping test to characterize the influence of menstruation in female subjects. Tsujimura et al. (2013) assessed tapping performance to estimate the level of constipation in PD patients. One study (Jansen, 2013) suggests that finger-tapping can be used to assess neuromuscular abilities in cancer. All of these works strongly support that the proposed CV-based finger-tapping algorithm can be utilized to assess many other diseases, other than PD.

6.2.3 Significance of gait algorithm

Many studies strongly support that gait and posture disturbances have serious implications on the quality of life of patients with PD (Brozova et al., 2009; Moore et al., 2007; Schrag et al., 2000). For instance, in the study performed by Gomez et al. (2007), the gait item in part III of the UPDRS showed a strong statistical correlation (Pearson's $r=0.52$, $p<0.001$) with the mean quality of life scores in PDQ-39. Specifically, some studies (Brozova et al., 2009; Moore et al., 2007) suggest that freezing of gait has a significant effect on the quality of life scores in PDQ-39 (Pearson's $r=0.37$, $p < 0.0015$) far beyond other gait components. While in another study, it was found that the postural stability has the greatest influence on the quality of life in PD and its improvement should be an important target in the treatment of disease (Shrag et al., 2000).

Considering the complications in gait, the proposed CV-based algorithm for gait and posture analysis can be useful to estimate gait symptoms in PD to devise appropriate therapies and treatment. Specifically, the real-time assessment of posture lean using web-cameras can reduce the risk of falling. Alarms can be triggered if the posture lean angle is below a certain threshold and this can help in avoiding falls, and consequently can help in reducing the fatality rate due to falls (Stevens, 2005).

6.2.3.1 General applications of the gait algorithm

The algorithm can be applied to solve several problems other than identifying gait symptoms in PD. For example, it can be used for person identification based on a subject's gait patterns. It can be used in physical training. It can be used for rehabilitation of gait after surgeries, for detection of falls of elderly people etc.

6.3 Conclusions

The investigation of the PD dimensions of speech, gross and fine motor functions, and the comparisons between the proposed first-principle data-driven methods and other existing computer methods for assessing motor symptoms in PD suggest that high classification accuracies in discriminating the severity of symptoms can be achieved by employing first principle features. Importantly, findings revealed that the first-principle features show a strong significant correlation with the clinicians' perspective of symptom severity, particularly because these features estimate the physical characteristics of the disease that are intuitively obvious to the clinicians. Moreover, the statistical analysis of these features supports that the feature quantities were consistent and reproducible in different test occasions and experiments.

The proposed algorithms were able to follow clinical standards, which suggest that these algorithms are able to provide a second-opinion in examining motor symptoms according to the UPDRS. Further, the utilization of audio and video recordings of the motor examination allows the clinicians to watch and hear the recorded data several times in order to compare between the subjective and objective assessment of symptoms. This can help in reducing bias in rating the symptom severity. Importantly, the unobtrusiveness, portability and simplicity of the equipment favor the usability of these methods for the assessment of speech, gross and fine motor symptoms from the patient's home environment.

6.4 Limitations and future work

In the case of speech analysis:

- One of the limitations was that the recordings of speech examinations were made in a silent room. If speech were recorded in the home environment, the data would be susceptible to noise in the background. Then, it would be important to employ noise suppression algorithms, capable of discriminating between impairment related noise and environmental noise, so that the environmental noise can be eliminated without affecting the symptom information in speech signals.
- The speech algorithms need to be validated on languages other than English.
- An optimum level of textual difficulty needs to be investigated for devising more effective paragraphs for recitation.

In the case of gait analysis:

- A larger clinically rated database is required to evaluate the proposed model of gait to quantify the level of symptom severity according to the UPDRS. More dynamic algorithms need to be incorporated in the gait model to solve problems in image segmentation. One alternative is to use depth-imaging technology (Calderita et al., 2013) that uses pixel-depth data for image segmentation that allows avoiding artifacts in illumination and color.

In the case of finger-tapping analysis:

- A biomechanical model of a hand can be developed by incorporating 3D recordings of a hand. This will enable precise estimation of joint angles and flexions of the index finger and thumb, which will enhance the clinical analysis; however it is not required in the UPDRS assessment of finger-tapping.

Apart from these limitations, another limitation in this research was that the data was collected and acquired from different sources and patient groups. If the data could be collected altogether for speech, gait and finger-tapping from the same patient group, this could have allowed finding a correlation between the proposed methods. In addition, a longitudinal study needs to be performed to determine the behavior of symptoms and effect of medication on feature quantities in the long run.

Currently the proposed algorithms are stand-alone functions for speech, gait and finger-tapping analyses. Specifically, the speech algorithms can play an important role as they can be integrated in a mobile device where patients can perform other fine motor tasks such as spiral drawings and stylus tapping (Westin et al., 2012). Additionally, the same mobile device can be used to record videos of gait and finger-tapping.

An important advantage of using speech and CV algorithms and the FPDD approach for symptom assessment is that, the proposed methods can be optimized to allow quantification of other items present in the UPDRS part-III. For example, 'Facial Expression' (item 19) can be quantified on the scale '0' to '4' using some modifications in the face detection module used in the CV method for finger-tapping analysis [Paper 5]. Moreover, body tremors including 'Tremor at Rest' (item 20), 'Action or postural tremor of hands' (item 21), 'Rigidity (item 22), 'Hand Movements' (item 24), 'Rapid Alternating Movement of Hands' (item 25) and Leg Agility (item 26) require assessment of movement symptoms that can be quantified using the motion gradient algorithm used in the finger-tapping analysis [Paper 5]. Likewise, the 'Postural Stability' (item 30) can be quantified using the proposed gait analysis method [Paper 7]. Further, item 27 ('Arising from Chair') and item 31 ('Body Bradykinesia and Hypokinesia') can be quantified by replacing the physiological model proposed in Paper 7 with a more dynamic model capable of representing different human poses and activities (Dubois & Charpillet, 2013) such as sitting in a chair, rising from the chair, lying on a bed, squatting, falling, bending etc. The motion gradient algorithm can be used in conjunction with this dynamic human model to enable estimating the movement symptoms including bradykinesia and hypokinesia.

Hence, all of the items on the UPDRS part-III, from item 18 ('Speech') to item 31 can be quantified using the proposed methods [Papers 1-7], if used either in parts or as a whole, with some changes. This would allow developing a complete package for the computerized assessment of the UPDRS part-III, which can be integrated in a mobile device to support remote monitoring and quantification of motor examination. Importantly, on one hand the quantification of all of the UPDRS III items would provide a holistic view of the disease profile that would improve clinical decision making on the patient's status. Besides, the use of mobile devices would provide the patients with an unobtrusive, cost-effective, portable and easy to use solution for symptom assessment.

The current plan is to accommodate the proposed algorithms into a test battery system for PD assessment (Westin et al., 2012) that at present employs patient-assessed questionnaires (Nyholm et al., 2004), and stylus tapping and spiral drawing tests for capturing upper limb motor abnormalities. The inclusion of new algorithms would enable the battery to capture motor speech, gross and fine motor dysfunctions, along with abnormalities in the upper-limbs, which would expand the coverage of the PD profile by the battery. Importantly, the improved battery functions would provide a feasible solution for clinicians to track disease progression and treatment intervention, as well as benefit patients who have physical restrictions by allowing at-home monitoring.

References (A-Z)

A

Adali, T., Bakal, B., Sönmez, M. K., Fakory, R., & Tsaoi, C. O. (1997). Modeling nuclear reactor core dynamics with recurrent neural networks. Neurocomputing, 15(3), 363-381.

Agostino, R., Curra, A., Giovannelli, M., Modugno, N., Manfredi, M., & Berardelli, A., (2003). Impairment of individual finger movements in Parkinson's disease. Movement disorders, 18(5), 560-565.

Ancillao, A., Galli, M., Vimercati, S. L., & Albertini, G. (2013). An optoelectronic based approach for handwriting capture. Computer methods and programs in biomedicine, 111(2), 357–365

Angaran, D. M. (1999). Telemedicine and tele-pharmacy: current status and future implications. American journal of health-system pharmacy, 56(14), 1405-1426.

Apps, M. C., Sheaff, P. C., Ingram, D. A., Kennard, C., & Empey, D. W. (1985). Respiration and sleep in Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry, 48(12), 1240-1245.

Avanzini, F. (2008). Simulation of vocal fold oscillation with a pseudo-one-mass physical model. Speech Communication, 50(2):95–108.

B

Baldereschi, M., Di Carlo, A., Rocca, W. A., Vanni, P., Maggi, S., Perissinotto, E., Inzitari, D., et al. (2000). Parkinson's disease and parkinsonism in a longitudinal study Two-fold higher incidence in men. Neurology, 55(9), 1358-1363.

Bayulkem, K., & Lopez, G. (2010). Non motor fluctuations in Parkinson's disease: clinical spectrum and classification. Journal of the neurological sciences, 289(1), 89-92.

Bäzner, H., Schanz, J., Blahak, C., Grips, E., Wöhrle, J. C., & Hennerici, M. (2005). Differential pattern of hand-tapping compromise in vascular versus idiopathic parkinsonism: A study based on computerized movement analysis.Movement disorders, 20(4), 504-508.

Behari, M., Srivastava, A. K., & Pandey, R. M. (2005). Quality of life in patients with Parkinson's disease. Parkinsonism & related disorders, 11(4), 221-226.

Bellazzi, R., Montani, S., Riva, A., & Stefanelli, M. (2001). Web-based telemedicine systems for home-care: technical issues and experiences. Computer Methods and Programs in Biomedicine, 64(3), 175-187.

Bemmel, J.H.V & Musen, M.A. (1997). Handbook of Medical informatics. Springer. ISBN 3-450-63351-0

Beranek, L. L. (1986). Acoustics. American Institute of Physics. New York, 128-129.

Berardelli, A., Rothwell, J. C., Thompson, P. D., & Hallett, M. (2001). Pathophysiology of bradykinesia in Parkinson's disease. Brain, 124(11), 2131-2146.

Berger, Y. G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. Biometrika, 94(4), 953-964.

Bloem, B. R., Grimbergen, Y. A., Cramer, M., Willemsen, M., & Zwinderman, A. H. (2001). Prospective assessment of falls in Parkinson's disease. Journal of neurology, 248(11), 950-958.

Bogart, B. P., Healy, M. L. R., & Tukey, J. W. (1963). The quefrency analysis of time-series for echoes. In Proc Symp. Time Series Analysis, Wiley, NY (pp. 209-243).

Bolton, S. (1997). Analysis of variance. Pharmaceutical Statistics: Practical and Clinical Applications. New York, Marcel Dekker, 737, 265-325.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211-252.

Bradski, G. R., & Davis, J. W. (2002). Motion segmentation and pose recognition with motion history gradients. Machine Vision and Applications,13(3), 174-184.

Brozova, H., Stochl, J., Roth, J., & Ruzicka, E. (2009). Fear of falling has greater influence than other aspects of gait disorders on quality of life in patients with Parkinson's disease. Neuro Endocrinol Lett, 30(4), 453-457.
Brubaker, M. A. (2012). Physical Models of Human Motion for Estimation and Scene Analysis (Doctoral dissertation).
Burkle, T. Z., Kewley-Port, D., Humes, L., & Lee, J. H. (2004). Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners. Journal of the Acoustical Society of America,115(5), 2601.

C

Calais, E., Legrand, L., Voisin, Y., & Diou, A. (2002). Contribution to a marker-free system for human motion analysis. Journal of Electronic Imaging, 11(3), 381-392.
Calderita, L. V., Bandera, J. P., Bustos, P., & Skiadopoulos, A. (2013). Model-Based Reinforcement of Kinect Depth Data for Human Motion Capture Applications. Sensors, 13(7), 8835-8855.
Campenhausen, S., Bornschein, B., Wick, R., Bötzel, K., Sampaio, C., Poewe, W., Dodel, R. et al. (2005). Prevalence and incidence of Parkinson's disease in Europe. European Neuropsychopharmacology, 15(4), 473-490.
Carlsson, A. (1974). The" on-off" effect. Advances in neurology, 5, 367.
Carlsson, A. (2002). Treatment of Parkinson's with L-DOPA. The early discovery phase, and a comment on current problems. Journal of neural transmission, 109(5), 777-787.
Catford, J. C. (1964). Phonation types: the classification of some laryngeal components of speech production. In honour of Daniel Jones, 26, 37.
Chen, B. R., Patel, S., Buckley, T., Rednic, R., McClure, D. J., Shih, L., Bonato, P, et al. (2011). A Web-based system for home monitoring of patients with Parkinson's disease using wearable sensors. Biomedical Engineering, IEEE Transactions on, 58(3), 831-836.
Cho, C. W., Chao, W. H., Lin, S. H., & Chen, Y. Y. (2009). A vision-based analysis system for gait recognition in patients with Parkinson's disease.Expert Systems with applications, 36(3), 7033-7039.
Clark, J., Yallop, C., and Fletcher, J. (2007). An Introduction to Phonetics and Phonology. Blackwell Publisher Ltd., Oxford, UK, 3th edition.
Collyer, C. E., Broadbent, H. A., & Church, R. M. (1994). Preferred rates of repetitive tapping and categorical time production. Perception & Psychophysics,55(4), 443-453.
Criss, K., & McNames, J. (2011, August). Video assessment of finger tapping for Parkinson's disease and other movement disorders. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 7123-7126). IEEE.
Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
Cummins, N., Sethu, V., Joshi, J., Goecke, R., Dhall, A., & Epps, J. (2013). Diagnosis of depression by behavioural signals: a multimodal approach.depression, 11, 31.
Cunado, D., Nixon, M. S., & Carter, J. N. (1999). Automatic gait recognition via model-based evidence gathering.
Cutkosky, M. R., & Howe, R. D. (1990). Human grasp choice and robotic grasp analysis. In Dextrous robot hands (pp. 5-31). Springer New York.
Czop, P., Kost, G., Sławik, D., & Wszołek, G. (2011). Formulation and identification of First-Principle Data-Driven models. Journal of Achievements in Materials and Manufacturing Engineering, 44(2), 179-186.

D

Damiano, A. M., Snyder, C., Strausser, B., & Willian, M. K. (1999). A review of health-related quality-of-life concepts and measures for Parkinson's disease.Quality of Life Research, 8(3), 235-243.
Dauer, W., Przedborski, S., (2003). Parkinson's disease: mechanisms and models. Neuron, 39(6), 889-909.

Davis, S. B., & Mermelstein, P. (1978). Evaluation of acoustic parameters for monosyllabic word identification. The Journal of the Acoustical Society of America, 64, S180.
De Looze, C., Ghio, A., Scherer, S., Pouchoulin, G., & Viallet, F. (2012). Automatic Analysis of the Prosodic Variations in Parkinsonian Read and Semi-Spontaneous Speech. In Speech Prosody 2012.
De Rijk, M. C., Launer, L. J., Berger, K., Breteler, M. M., Dartigues, J. F., Baldereschi, M., Hofman, A. et al. (1999). Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology, 54(11 Suppl 5), S21-3.
Decoste, D., & Schölkopf, B. (2002). Training invariant support vector machines. Machine Learning, 46(1-3), 161-190.
Doretto, G., & Soatto, S. (2006). Dynamic shape and appearance models.Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(12).
Dubois, A., & Charpillet, F. (2013, July). Human activities recognition with RGB-Depth camera using HMM. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE (pp. 4666-4669). IEEE.
Dubuisson, T. (2011). Glottal Source Estimation and Automatic Detection of Dysphonic Speakers.
Dzitkowski, T., & Dymarek, A. (2008). Design and examining sensitivity of machine driving systems with required frequency spectrum. Journal of Achievements in Materials and Manufacturing Engineering, 26(1), 49-56.

E

Edward, V., Windischberger, C., Cunnington, R., Erdler, M., Lanzenberger, R., Mayer, D., ... & Beisteiner, R. (2000). Quantification of fMRI artifact reduction by a novel plaster cast head holder. Human brain mapping, 11(3), 207-213.
Elbaz, A., Bower, J. H., Maraganore, D. M., McDonnell, S. K., Peterson, B. J., Ahlskog, J. E., Rocca, W. A. et al. (2002). Risk tables for parkinsonism and Parkinson's disease. Journal of clinical epidemiology, 55(1), 25-31.
Elble, R. J. (2005). Gravitational artifact in accelerometric measurements of tremor. Clinical Neurophysiology, 116(7), 1638-1643.
E-MOTIONS, (2010-2012). Online: http://www.du.se/sv/Forskning/Forskningsprojekt/?&pid=HDA2010-00041.

F

Factor, S. A., (2008). Parkinson's disease: diagnosis and clinical management. W. J. Weiner (Ed.). New York: Demos.
Fahn, S.R.L.E., Elton, R., & UPDRS Development Committee., (1987). Unified Parkinson's disease rating scale. Recent developments in Parkinson's disease 2, 153-163.
Fant, G. (1960). Accoustic theory of speech production: with calculations based on X-ray studies of Russian articulations. Mouton & Company.
Findley, L.J., (2007). The economic impact of Parkinson's disease. Parkinsonism & Related Disorders 13, S8-S12.
Flanagan, J.L., Ishizaka, K., & Shipley, K.L., (1975). Synthesis of speech from a dynamic model of the vocal cords and vocal tract. Bell System Technical Journal 54(3), 485-506.
Fry, D. B. (1979). The physics of speech. Cambridge University Press.

G

Garud, S. S., Derle, D. V., & Derle, N. D. (2013) A ROLE OF MODELS IN QUALITY BY DESIGN (QbD) PARADIGM-A REVIEW. International Journal of Biopharmaceutics, 4(3), 180-189.
Gelzinis, A., Verikas, A., & Bacauskiene, M. (2008). Automated speech analysis applied to laryngeal disease categorization. Computer Methods and Programs in Biomedicine, 91(1), 36-47.

Geng, X., Liu, T. Y., Qin, T., & Li, H. (2007, July). Feature selection for ranking. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 407-414). ACM.

Ghassemi, M., Lemieux, S., Jog, M., Edwards, R., & Duval, C. (2006). Bradykinesia in patients with Parkinson's disease having levodopa-induced dyskinesias. Brain research bulletin, 69(5), 512-518.

Giovannoni, G., Van Schalkwyk, J., Fritz, V. U., & Lees, A. J. (1999). Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. Journal of Neurology, Neurosurgery & Psychiatry, 67(5), 624-629.

Godino-Llorente, J. I., Ritchings, R. T., & Berry, C. (2003). The effects of inter and intra speaker variability on pathological voice quality assessment. In 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications.

Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Fahn, S., ... & Seidl, L. (2003). The unified Parkinson's disease rating scale (UPDRS): status and recommendations. Movement Disorders, 18(7), 738-750.

Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Counsell, C., ... & Seidl, L. (2004). Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson's disease.Movement disorders, 19(9), 1020-1028.

Goetz, C. G., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stebbins, G. T., ... & LaPelle, N. (2007). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. Movement Disorders, 22(1), 41-47.

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., ... & LaPelle, N. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Movement Disorders, 23(15), 2129-2170.

Goetz, C.G., Stebbins, G.T., Wolff, D., DeLeeuw, W., Bronte-Stewart, H., Elble, R., & Taylor, C. B., (2009). Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device. Movement Disorders 24(4), 551-556.

Goffredo, M., Seely, R. D., Carter, J. N., & Nixon, M. S. (2008, September). Markerless view independent gait analysis with self-camera calibration. InAutomatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on (pp. 1-6). IEEE.

Gómez-Esteban, J. C., Zarranz, J. J., Lezcano, E., Tijero, B., Luna, A., Velasco, F., ... & Garamendi, I. (2007). Influence of motor symptoms upon the quality of life of patients with Parkinson's disease. European Neurology, 57(3), 161-165.

Gorry, G. A. (1973). Computer-assisted clinical decision making. Methods of Information in Medicine, 12(45-51), 11.

Green, M. F., Kern, R. S., & Heaton, R. K. (2004). Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS.Schizophrenia research, 72(1), 41-51.

Green, R. D., Guan, L., & Burne, J. A. (2000). Video analysis of gait for diagnosing movement disorders. Journal of Electronic Imaging, 9(1), 16-21.

Guan, W. (2011). New support vector machine formulations and algorithms with application to biomedical data analysis.

Guttman, L. (1944). A basis for scaling qualitative data. American sociological review, 9(2), 139-150.

Guttman, L., Stouffer, S. A., Suchman, E. A., & Lazarsfeld, P. F. (1950). The American Soldier, Vol. 4: Measurement and Prediction.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46(1-3), 389-422.

Guyton, A. C., & Hall, J. E. (2006). Textbook of medical physiology. Elsevier Saunders. Philadelphia, PA, 764-5.

H

Haas, L. F. (2001). Jean Martin Charcot (1825–93) and Jean Baptiste Charcot (1867–1936). Journal of Neurology, Neurosurgery & Psychiatry, 71(4), 524-524.

Haaxma, C. A., Bloem, B. R., Borm, G. F., Oyen, W. J., Leenders, K. L., Eshuis, S., Horstink, M. W. et al. (2007). Gender differences in Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry, 78(8), 819 824.

Hagell, P. (2000). Timed tests in the clinical assessment of motor function in Parkinson's disease. Journal of neuroscience nursing, 32(6), 331-336.

Hanson, D. G., Gerratt, B. R., & Ward, P. H. (1984). Cinegraphic observations of laryngeal function in Parkinson's disease. The Laryngoscope, 94(3), 348-353.

Harel, B., Cannizzaro, M., & Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. Brain and cognition, 56(1), 24-29.

Hartelius, L., & Svensson, P. (1994). Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. Folia Phoniatrica et Logopaedica, 46(1), 9-17.

Haubenberger, D., Kalowitz, D., Nahab, F. B., Toro, C., Ippolito, D., Luckenbaugh, D. A., ... & Hallett, M. (2011). Validation of digital spiral analysis as outcome parameter for clinical trials in essential tremor. Movement Disorders, 26(11), 2073-2080.

Heller, A. N. D. R. E. W., Wade, D. T., Wood, V. A., Sunderland, A., Hewer, R. L., & Ward, E. (1987). Arm function after stroke: measurement and recovery over the first three months. Journal of Neurology, Neurosurgery & Psychiatry,50(6), 714-719.

Hixon, T. J., Weismer, G., & Hoit, J. D. (2008). Preclinical speech science: anatomy, physiology, acoustics, and perception. Plural Pubications.

Hjelm, N. M. (2005). Benefits and drawbacks of telemedicine. Journal of telemedicine and telecare, 11(2), 60-70.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease.Behavioural neurology, 11(3), 131-137.

Hoehn, M. M., & Yahr, M. D. (1998). Parkinsonism: onset, progression, and mortality. Neurology, 50(2), 318-318.

Hollingworth LS. Functional Periodicity. In: Classics in the History of Psychology; 1914.

Holmes, R., M Oates, J., J Phyland, D., & J Hughes, A. (2000). Voice characteristics in the progression of Parkinson's disease. International Journal of Language & Communication Disorders, 35(3), 407-418.

Hong, Y., & Bartlett., R. (Eds.), (2010). Routledge handbook of biomechanics and human movement science. Abingdon, UK: Routledge.

Huang, P. S., Harris, C. J., & Nixon, M. S. (1999). Recognising humans by gait via parametric canonical space. Artificial Intelligence in Engineering, 13(4), 359-366.

Hughes, A. J., Daniel, S. E., Blankson, S., & Lees, A. J. (1993). A clinicopathologic study of 100 cases of Parkinson's disease. Archives of Neurology, 50(2), 140.

I

International Phonetic Association (Ed.). (1999). Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press.

Isacson, D., Bingefors, K., Kristiansen, I. S., & Nyholm, D., (2008). Fluctuating functions related to quality of life in advanced Parkinson disease: effects of duodenal levodopa infusion. Acta Neurologica Scandinavica, 118(6), 379-386.

J

Jafari, A. (2013). CLASSIFICATION OF PARKINSON'S DISEASE PATIENTS USING NONLINEAR PHONETIC FEATURES AND MEL-FREQUENCY CEPSTRAL ANALYSIS. Biomedical Engineering: Applications, Basis and Communications.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis.Journal of Neurology, Neurosurgery & Psychiatry, 79(4), 368-376.

Jansen, C. E. (2013, November). Cognitive Changes Associated with Cancer and Cancer Therapy: Patient Assessment and Education. In Seminars in oncology nursing (Vol. 29, No. 4, pp. 270-279). WB Saunders.

Jobbagy, A., Furnée, H., Harcos, P., Tárczy, M., Krekule, I., & Komjáthi, L. (1997). Analysis of movement patterns aids the early detection of Parkinson's disease. In Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE (Vol. 4, pp. 1760-1763). IEEE.

Jones, H. N. (2009). Prosody in Parkinson's disease. Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 19(3), 77-82.

K

Kandori, A., Yokoe, M., Sakoda, S., Abe, K., Miyashita, T., Oe, H., ... & Tsukada, K. (2004). Quantitative magnetic detection of finger movements in patients with Parkinson's disease. Neuroscience research, 49(2), 253-260.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential.Journal of communication disorders, 32(3), 141-186.

Khan, T., Westin, J., & Dougherty, M. (2013). Classification of speech intelligibility in Parkinson's disease. Biocybernetics and Biomedical Engineering.

Kim, S., Eriksson, T., & Kang, H. G. (2004, October). On the time variability of vocal tract for speaker recognition. In INTERSPEECH.

Kimber, T. E., Tsai, C. S., Semmler, J., Brophy, B. P., & Thompson, P. D. (1999). Voluntary movement after pallidotomy in severe Parkinson's disease.Brain, 122(5), 895-906.

Kjeldsen, T. H. (2000). A contextualized historical analysis of the Kuhn–Tucker Theorem in nonlinear programming: The impact of World War II. Historia Mathematica, 27(4), 331-361.

Klingholtz, F., 1990. Acoustic recognition of voice disorders: A comparative study of running speech versus sustained vowels. The Journal of the Acoustical Society of America 87, 2218-2224.

Koop, M. M. (2011). Targeting Hypokinesia in Parkinson's Disease with Quantitative Measures and a Computational Model to Determine Its Severity, Its Improvement from Therapy, and Explore Underlying Pathophysiological Mechanisms. Stanford University.

Krom, G. D. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. Journal of Speech, Language and Hearing Research, 36(2), 254.

Krom, G.D., 1995. Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. Journal of Speech, Language and Hearing Research 38(4), 794.

Kupryjanow, A., Kunka, B., & Kostek, B. (2010, August). UPDRS Tests for Diagnosis of Parkinson's Disease Employing Virtual-Touchpad. In Database and Expert Systems Applications (DEXA), 2010 Workshop on (pp. 132-136). IEEE.

Kwon, O. W., & Lee, T. W. (2004). Phoneme recognition using ICA-based feature extraction and transformation. Signal Processing, 84(6), 1005-1019.

L

LaMar, M. D., Qi, Y., and Xin, J. (2003). Modeling vocal fold motion with a hydrodynamic semicontinuum model. Journal of the Acoustical Society of America, 114(1):455–464.

Lang, A. E., & Lozano, A. M. (1998). Parkinson's disease. New England Journal of Medicine, 339(15), 1044-1053.

Lees, A. J. (1989). The on-off phenomenon. Journal of Neurology, Neurosurgery & Psychiatry, 52(Suppl), 29-37.
Little. M.A, (2006). Biomechanically Informed Nonlinear Speech Signal Processing, D.Phil. Thesis, University of Oxford.
Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Biomedical Engineering, IEEE Transactions on, 56(4), 1015-1022.
Livio, M. (2008). The golden ratio: The story of phi, the world's most astonishing number. Random House Digital, Inc..
Llorente, J. I., Fraile, R., Sáenz-Lechón, N., Osma-Ruiz, V., & Gómez-Vilda, P. (2009). Automatic detection of voice impairments from text-dependent running speech. Biomedical Signal Processing and Control, 4(3), 176-182.
Llorente, J. I., Ritchings, R. T., & Berry, C. (2003). The effects of inter and intra speaker variability on pathological voice quality assessment. In 3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications.
Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders,43(1), 47.
Londono, J. D., Godino-Llorente, J. I., Sáenz-Lechón, N., Osma-Ruiz, V., & Castellanos-Dominguez, G. (2011). Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients. Biomedical Engineering, IEEE Transactions on, 58(2), 370-379.
Louis, E. D., Gillman, A., Boschung, S., Hess, C. W., Yu, Q., & Pullman, S. L. (2012). High width variability during spiral drawing: further evidence of cerebellar dysfunction in essential tremor. The Cerebellum, 11(4), 872-879.

M

Mayo clinic (2013), online: http://www.mayo.edu/ [Last accessed: 14th December 2013].
McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In Philosophical Logic and Artificial Intelligence (pp. 161-190). Springer Netherlands.
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. Psychological methods, 1(1), 30.
Melnick, M. E., Radtka, S., & Piper, M. (2002). Gait analysis and Parkinson's disease. Rehab management, 15(6), 46-46.
Memedi, M., Westin, J., Nyholm, D., Dougherty, & M., Groth, T., (2011). A web application for follow-up of results from a mobile device test battery for Parkinson's disease patients. Computer methods and programs in biomedicine, 104(2), 219-226.
Metz, C. E. (1978, October). Basic principles of ROC analysis. In Seminars in nuclear medicine (Vol. 8, No. 4, pp. 283-298). WB Saunders.
Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M.A., & Gunal, D.I., (2008). Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. Acta Neurologica Scandinavica 117(1), 26-34.
Miller, R. A., & Antoine Geissbuhler MD, F. A. C. M. I. (2007). Diagnostic decision support systems. In Clinical Decision Support Systems (pp. 99-125). Springer New York.
Mitchell, S. L., Harper, D. W., Lau, A., & Bhalla, R. (2000). Patterns of outcome measurement in Parkinson's disease clinical trials.Neuroepidemiology, 19(2), 100-108.
Moore, O., Peretz, C., & Giladi, N. (2007). Freezing of gait affects quality of life of peoples with Parkinson's disease beyond its relationships with mobility and gait. Movement disorders, 22(15), 2192-2195.
Morse, P. M. (1986). Theoretical acoustics. Princeton University Press.
Mouradian, M. M., Junecos, J. L., Fabbrini, G., Schlegel, J., Bartko, J. J., & Chase, T. N. (1988). Motor fluctuations in Parkinson's disease: central pathophysiological mechanisms, part II. Annals of neurology, 24(3), 372-378.

Muir, S. R., Jones, R. D., Andreae, J. H., & Donaldson, I. M. (1995). Measurement and analysis of single and multiple finger tapping in normal and Parkinsonian subjects. Parkinsonism & Related Disorders, 1(2), 89-96.
Murdoch, B. E. (Ed.). (1998). Dysarthria: a physiological approach to assessment and treatment. Nelson Thornes.
Murphy, P. (2008). Source-filter comparison of measurements of fundamental frequency perturbation and amplitude perturbation for synthesized voice signals.Journal of Voice, 22(2), 125-137.
Murray, M. P., Kory, R. C., Clarkson, B. H., & Sepic, S. B. (1966). Comparison of free and fast speed walking patterns of normal men. American Journal of Physical Medicine & Rehabilitation, 45(1), 8-24.
Myers, J. L., Well, A. D., & Lorch, R. F. (2010). Research design and statistical analysis. Routledge.

N

Nebylitsyn, V. D. (1972). Fundamental properties of the human nervous system. G. L. Mangan (Ed.). New York: Plenum Press.
Nieuwboer, A. (2006). Freezing of Gait: Problem analysis and rehabilitation strategies. Parkinsonism & Related Disorders, 12, S53-S54.
Nooteboom, S. G., & Coden, A. (1984). Het proces van spreken en verstaan, een nieuwe inleidingin de experimentele fonetiek. Van Gorcum, Assen, The Netherlands.
Norman, K. E., & Héroux, M. E. (2013). Measures of fine motor skills in people with tremor disorders: appraisal and interpretation. Frontiers in neurology, 4(50), In Press.
Notermans, N. C., Van Dijk, G. W., Van der Graaf, Y., Van Gijn, J., & Wokke, J. H. (1994). Measuring ataxia: quantification based on the standard neurological examination. Journal of Neurology, Neurosurgery & Psychiatry,57(1), 22-26.
Nyholm, D., Askmark, H., Gomes-Trolin, C., Knutson, T., Lennernäs, H., Nyström, C., & Aquilonius, S. M. (2003). Optimizing levodopa pharmacokinetics: intestinal infusion versus oral sustained-release tablets.Clinical neuropharmacology, 26(3), 156-163.
Nyholm, D., Kowalski, J., & Aquilonius, S. M. (2004). Wireless real-time electronic data capture for self-assessment of motor function and quality of life in Parkinson's disease. Movement Disorders, 19(4), 446-451.
Nyholm, D., Remahl, A. N., Dizdar, N., Constantinescu, R., Holmberg, B., Jansson, R., Askmark, H., et al. (2005). Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson disease. Neurology, 64(2), 216-223.

O

Olanow, C.W., Stern, M.B., & Sethi, K., (2009). The scientific and clinical basis for the treatment of Parkinson. Neurology, 72(21 Supplement 4), S1-S136.
Okuno, R., Yokoe, M., Akazawa, K., Abe, K., & Sakoda, S. (2006, August). Finger taps movement acceleration Measurement System for Quantitative Diagnosis of Parkinson's disease. In Proceedings of the twenty-eighth annual international conference of the IEEE Engineering in Medicine and Biology Society (pp. 6623-6626).
Okuno, R., Yokoe, M., Fukawa, K., Sakoda, S., & Akazawa, K. (2007, August). Measurement system of finger-tapping contact force for quantitative diagnosis of Parkinson's disease. In Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE(pp. 1354-1357). IEEE.
Organisation mondiale de la santé. (2006). Neurological disorders: public health challenges. World Health Organization.
Ott, B. R., Ellias, S. A., & Lannon, M. C. (1995). Quantitative assessment of movement in Alzheimer's disease. Journal of geriatric psychiatry and neurology.
Ottesen, J. T., Olufsen, M. S., & Larsen, J. K., (2004). Applied mathematical models in human physiology Philadelphia: Society for Industrial and Applied Mathematics.

P

Parkinson, J. (1817). An essay on the shaking palsy. Printed by Whittingham and Rowland for Sherwood, Neely, and Jones.

Paja, M. O. S., & Falk, T. H. (2012, September). Automated Dysarthria Severity Classification for Improved Objective Intelligibility Assessment of Spastic Dysarthric Speech. In INTERSPEECH.

PAULINA (2013-2015). Online: http://www.du.se/en/Forskning/Forskningsprojekt/?&pid=HDA2013-00032

Pasquina, P. F., & Cooper, R. A. (Eds.). (2009). Care of the combat amputee. Government Printing Office.

Peng, Y., & Reggia, J. A. (1986). Plausibility of Diagnostic Hypotheses: The Nature of Simplicity. In AAAI (Vol. 86, pp. 140-145).

Pinto, S., Ozsancak, C., Tripoliti, E., Thobois, S., Limousin-Dowsey, P., & Auzou, P. (2004). Treatments for dysarthria in Parkinson's disease. The Lancet Neurology, 3(9), 547-556.

Plackett, R. L. (1983). Karl Pearson and the chi-squared test. International Statistical Review/Revue Internationale de Statistique, 59-72.

R

Rahman, S., Griffin, H. J., Quinn, N. P., & Jahanshahi, M. (2008). Quality of life in Parkinson's disease: the relative importance of the symptoms. Movement Disorders, 23(10), 1428-1434.

Rajput, A. H., Rozdilsky, B., & Rajput, A. (1991). Accuracy of clinical diagnosis in parkinsonism--a prospective study. The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques, 18(3), 275-278.

Rajput, A., Rajput, M., & Rajput, A. H., (2007). Old age and Parkinson's disease. Handbook of Clinical Neurology, 84, 427-444.

Ramaker, C., Marinus, J., Stiggelbout, A. M., & van Hilten, B. J. (2002). Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Movement Disorders, 17(5), 867-876.

Rao, G., Fisch, L., Srinivasan, S., D'Amico, F., Okada, T., Eaton, C., & Robbins, C. (2003). Does this patient have Parkinson disease?. JAMA: the journal of the American Medical Association, 289(3), 347-353.

Reiter, R., (1987). A theory of diagnosis from first principles. Artificial intelligence, 32(1), 57-95.

Riva, G., Wiederhold, B. K., & Molinari, E. (Eds.). (1998). Virtual environments in clinical psychology and neuroscience: Methods and techniques in advanced patient therapist interaction (Vol. 58). IOS press.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. The Journal of the Acoustical Society of America, 129(1), 350-367.

S

Salhi, L., Talbi, M., & Cherif, A. (2008). Voice disorders identification using hybrid approach: Wavelet analysis and multilayer neural networks. InProceedings of World Academy of Science, Engineering and Technology (Vol. 35, pp. 2070-374).

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural computation, 13(7), 1443-1471.

Schrag, A., Jahanshahi, M., & Quinn, N. (2000). What contributes to quality of life in patients with Parkinson's disease?. Journal of Neurology, Neurosurgery & Psychiatry, 69(3), 308-312.

Shima, K., Tsuji, T., Kan, E., Kandori, A., Yokoe, M., & Sakoda, S. (2008, August). Measurement and evaluation of finger tapping movements using magnetic sensors.

In Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE (pp. 5628-5631). IEEE.
Shima, K., Tsuji, T., Kandori, A., Yokoe, M., & Sakoda, S. (2009). Measurement and evaluation of finger tapping movements using log-linearized Gaussian mixture networks. Sensors, 9(3), 2187-2201.
Shimoyama, I., Ninchoji, T., & Uemura, K. (1990). The finger-tapping test: a quantitative analysis. Archives of neurology, 47(6), 681.
Shortliffe, E. (Ed.). (1976). Computer-based medical consultations: MYCIN. Elsevier.
Shortliffe, E. H. (1987). Computer programs to support clinical decision making. JAMA, 258(1), 61-6.
Shulman, L. M., Pretzer-Aboff, I., Anderson, K. E., Stevenson, R., Vaughan, C. G., Gruber-Baldini, A. L., ... & Weiner, W. J. (2006). Subjective report versus objective measurement of activities of daily living in Parkinson's disease.Movement disorders, 21(6), 794-799.
Siddiqui, M. F., Rast, S., Lynn, M. J., Auchus, A. P., & Pfeiffer, R. F., (2002). Autonomic dysfunction in Parkinson's disease: a comprehensive symptom survey. Parkinsonism & Related Disorders, 8(4), 277-284.
Silbert, N., & de Jong, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. The Journal of the Acoustical Society of America, 123, 2769.
Skodda, S., Rinsche, H., & Schlegel, U. (2009). Progression of dysprosody in Parkinson's disease over time—a longitudinal study. Movement Disorders,24(5), 716-722.
Stevens, J. A. (2005). Falls among older adults—risk factors and prevention strategies. Journal of Safety Research, 36(4), 409-411.
Stevens, S. S., & Volkmann, J. O. H. N. (1940). The relation of pitch to frequency: A revised scale. The American Journal of Psychology, 53(3), 329-353.
Stewart, M., & Selesnick, S. (Eds.). (2010). Differential Diagnosis in Otolaryngology: Head and Neck Surgery. Thieme.
Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. Psychological review, 117(2), 541.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society. Series B (Methodological), 111-147.

T

Tafazzoli, F., & Safabakhsh, R. (2010). Model-based human gait recognition using leg and arm movements. Engineering applications of artificial intelligence,23(8), 1237-1246.
Thompson, L., Lin, E., & Robb, M. P. (2011). The impact of breathiness on the intelligibility of speech.
Titze, I. R. (1994). Principles of voice production (pp. 279-306). Englewood Cliffs: Prentice Hall.
Titze, I. R. (2000). Principles of Voice Production (National Center for Voice and Speech, Iowa City, IA:). Chap, 6, 149-184.
Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. The Journal of the Acoustical Society of America, 123, 2733.
Tsanas, A. (2012). Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning, D.Phil. Thesis, University of Oxford.
Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests.Biomedical Engineering, IEEE Transactions on, 57(4), 884-893.
Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2011). Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of aver-

age Parkinson's disease symptom severity. Journal of the Royal Society Interface, 8(59), 842-855.
Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. Biomedical Engineering, IEEE Transactions on, 59(5), 1264-1271.
Tsujimura, A., Yamamoto, Y., Sakoda, S., Okuda, H., Yamamoto, K., Fukuhara, S., ... & Nonomura, N. (2013). Finger taps and constipation are closely related to symptoms of overactive bladder in male patients with Parkinson's disease. International Journal of Urology.

U

Üstün, B., Melssen, W. J., & Buydens, L. M. C. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems,81(1), 29-40.

V

Van Laar, T. (2003). Levodopa-induced response fluctuations in patients with Parkinson's disease. CNS drugs, 17(7), 475-489.
Vapnik, V. N. (1992). Principles of risk minimization for learning theory.Advances in neural information processing systems, 4, 831-838.
Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.
Volkow, N. D., Gur, R. C., Wang, G. J., Fowler, J. S., Moberg, P. J., Ding, Y. S., ... & Logan, J. (1998). Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals.American Journal of Psychiatry, 155(3), 344-349.
von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering, 60, 86-101.

W

Wang, F., Stone, E., Dai, W., Banerjee, T., Giger, J., Krampe, J., ... & Skubic, M. (2009, September). Testing an in-home gait assessment tool for older adults. In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE (pp. 6147-6150). IEEE.
Weaver, F., Follett, K., Hur, K., Ippolito, D., & Stern, M., (2005). Deep brain stimulation in Parkinson disease: a metaanalysis of patient outcomes. Journal of neurosurgery, 103(6), 956-967.
Weiss, S. M., Kulikowski, C. A., Amarel, S., & Safir, A. (1978). A model-based method for computer-aided medical decision-making. Artificial intelligence,11(1), 145-172.
Welch, L. W., Cunningham, A. T., Eckardt, M. J., & Martin, P. R. (1997). Fine motor speed deficits in alcoholic Korsakoff's syndrome. Alcoholism: Clinical and Experimental Research, 21(1), 134-138.
Westin, J., (2010). Decision support for treatment of patients with advanced Parkinson's disease. D.Phil Thesis, University of Uppsala.
Westin, J., Schiavella, M., Memedi, M., Nyholm, D., Dougherty, M., & Antonini, A., (2012). Validation of a home environment test battery for supporting assessments in advanced Parkinson's disease. Neurological Sciences, 33(4), 831-838.
Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. Journal of Phonetics, 23(3), 349-366.
Wickwar, V. B., & Carlson, H. C. (1999). Ionospheric and thermospheric couplings: vertical, latitudinal and longitudinal. Journal of atmospheric and solar-terrestrial physics, 61(1), 141-152.

Winter, D.A., (1990). Biomechanics and Motor Control of Human Movement 3rd edition John Wiley and Sons.
Woollacott, M., & Shumway-Cook, A. (2002). Attention and the control of posture and gait: a review of an emerging area of research. Gait & posture,16(1), 1-14.
Wolters, E. C., van Laar, T., & Berendse, H. W. (Eds.). 2007. Parkinsonism and related disorders. VU University Press.
Wu, T. D. 1990. Efficient Diagnosis of Multiple Disorders Based on a Symptom Clustering Approach. In AAAI 357-364.

Y

Ye, J., Johnson, M. T., & Povinelli, R. J. (2003). Study of attractor variation in the reconstructed phase space of speech signals. In ISCA Tutorial and Research Workshop on Non-Linear Speech Processing.
Yoder, N., Peak Finger (2011), Online available at: http://www.mathworks.com/matlabcentral/fileexchange/25500, [last accessed 20th October 2012].

Z

Zhou, H., Green, P. R., & Wallace, A. M. (2004, May). Efficient motion tracking using gait analysis. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on (Vol. 3, pp. iii-601). IEEE.
Zraick, R. I., Dennie, T. M., Tabbal, S. D., Hutton, T. J., Hicks, G. M., & O'Sullivan, P. S. (2003). Reliability of speech intelligibility ratings using the Unified Parkinson Disease Rating Scale. Journal of Medical Speech-Language Pathology, 11(4), 227-240.