MASTER

A clinical decision support system by using wrist-worn smartphone tremor measurements at the Maastricht University Medical Center

Zamora, G.

Award date:
2017

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
A clinical decision support system by using wrist-worn smartphone tremor measurements

at the Maastricht University Medical Center
by Guillaume Zamora

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
Pieter Kubben, Maastricht University Medical Center
Uzay Kaymak, Technical University of Eindhoven (1st supervisor)
Anna Wilbik, Technical University of Eindhoven (2nd supervisor)
Pieter van Gorp, Technical University of Eindhoven (3rd assessor)
Pierre Lemaire, The Industrial Engineering & Management School at Grenoble Institute of Technology

Eindhoven, April 2017
Abstract

**Background:** Tremor related diseases affect millions of people around the world, hindering various everyday life tasks, such as holding a glass of water. Tremor severity assessment is an important element for the diagnosis and treatment decision making process. For decades, subjective clinical rating scales were mostly performed. Recently, remarkable attention around computerized tremor analysis has grown. While dedicated devices are expensive and not practical for the everyday use, smartphone applications are promising. Previous studies on Parkinson’s disease or Essential Tremor mostly classified the Fahn score rating scale. **Objective:** Using machine learning techniques to regress tremor severity observed by clinicians (ETRS) and patients (QUEST), and give the research accessibility and new insights that would later lead to decision making process improvements. **Methods:** Five wrist-worn different tests were performed on 20 Essential Tremor patients from the open-source TREMOR12 iPhone/iWatch compatible application. Linear displacements and joint rotations are measured from in-device accelerometer and gyroscope. From these signals, time, frequency and time-frequency domain tools are used to extract the following features: dominant frequency, dominant magnitude, signal RMS, signal period and the power growth during the test. **Results:** While the study demonstrates good predictive power, its feature extraction shows to bring improvements when compared to previous close setting studies. **Conclusion:** This study gives the research new directions and tools in order to perform further investigations around tremor severity evaluation. Smartphone sensors improvement in the following years, research on the best predicted variable to use and larger data collection may lead to very robust models, measuring rapidly, accurately and more objectively tremor severity than clinical rating scales.

**Key words:** movement disorder, tremor severity, smartphone, feature extraction, regression, essential tremor

Acknowledgement

For this master thesis project I have received valuable and essential support from various sources.

First of all, I would like to express my gratitude to my French Engineering School: Grenoble INP Génie Industriel, and the Eindhoven University of Technology who gave me the opportunity to experience a double degree in France and The Netherlands.

Then I would like to express my deepest appreciation to my academic first supervisor, prof.dr.ir. Uzay Kaymak and to Pieter Kubben that proposed me this master thesis project. Their support was essential for this thesis project completion.

Finally, I also want to thank Aurélie Degeneffe (former medical student Maastricht University and Nicole Bakker (case manager DBS), who made the data collection for this study.
Management summary

This thesis comes in support to recent important research around closely linked Parkinson’s disease and Essential Tremor assessment by computerized techniques, in order to evaluate tremor severity in a more objective and effective way. This study has data mining, research and process oriented goals. The main data mining goal is to regress with good accuracy the two outputs values: Essential Tremor Rating Scale (ETRS) and Quality of life in Essential Tremor Questionnaire (QUEST). Creating a base line model following close setting literature and compare it with improvements brought in this paper is the second data mining goals. The research goals are naturally coming from data mining goals, in order to help the actual literature. Indeed, new techniques, combinations of features, parameters and tests, and new predicted outputs are all elements bringing value to this specific research area. As the research and applications are growing, new tremor severity assessment will be possible, improving the process measurement objectivity and performances.

Based on previous data collection realized by Aurélie Degeneffe and Nicole Bakker in 2016, data was collected from twenty Essential Tremor affected patients using TREMOR12 app iPhone wrist-worn measurements created by (Kubben P., Kuijf, Ackermans, Leentjens, & Temel, 2016). In order to collect the most representative dataset possible, five tests per patient were performed. One rest, two postural and two kinetic tests were indeed performed. Each patient is tested twice per test, on each wrist (left and right). Each test is providing linear acceleration and rotation speed signals on 3 dimensions x, y and z.

Signals are cleaned. Sampling errors are removed from the time signal and noise removal is performed using a bandpass Equiripple Finite Impulse Response (FIR) filter between 7 to 12 Hz, corresponding to Essential Tremor frequencies. This way, gravitational and motion components are removed from the signals by filtering.

From each cleaned signal, various features are extracted in order to represent the time series in single values that can later be used as input values for machine learning modelization. Two time domain features are extracted, namely signal RMS representing signal strength and signal period representing the average period for each signal wave. Dominant magnitude and frequency are extracted from the frequency domain, giving information about the highest strength peak accumulated over the frequency range and the number of oscillations per second associated. Power growth is extracted from the time-frequency domain in order to represent power increase in time.

Tests, parameters and features are combined in fifty variables for each of the twenty patients. A Sequential Forward Selection algorithm using inner cross-validation based on the modelization technique used later is then implemented to extract the best feature subset possible. This subset admits to know which variables and combinations are most relevant.

Using a leave-one-out cross-validation, regression trees are used as modelization technique. Three metrics are used to evaluate the various models presented. In these models, two frequency domain feature extractions are compared. The Welch method, using Hamming 2.5 seconds and 50% overlapping windows, and a Wavelet extractor. Also, Daubechies 8 Wavelets with 6 levels are mainly used in order to extract a time-frequency domain feature, namely the power growth. Other models
focusing on each test type, parameter, feature and side are presented in order to give in-depth insights for researchers in relevant setting that can be used.

The base line model using close setting studies features with a feature selection algorithm gives a 27.58% error for ETRS and 80.42% for QUEST using the Mean Absolute Percentage Error as metric. In comparison, the improved model gives a 22.24% error for ETRS and 42.47% for QUEST. These results prove that the goals of this study are reached.
Introduction

Tremor, a “rhythmic and involuntary movements of any body part” (Elias & Shah, 2014), is a symptom resulting from a neurological disorder. In 1817, James Parkinson published An Essay on Shaking Palsy, raising awareness around this important progressive and degenerative disease. From 1980 to 2013, worldwide deaths attributed to Parkinson’s disease (PD) raised from 44 000 to 103 000 every year (Lancet, 2014). PD does not directly imply death, but has various complications shortening patient’s life expectancy. It’s most characteristic symptom is rest tremors. Essential Tremor (ET) is the most common tremor disorder, less known than PD since it does not have effect on lifespan or lead to other serious brain disorder. However, when ET triggers severe tremors, it affects various daily tasks as drinking a glass of water for the patient.

Clinical process improvement motivations: a better assessment of tremor severity

If ET and PD cannot be cured, several ways to treat the tremor exist. In this study, the focus is on Deep Brain Stimulation (DBS), a neurosurgical procedure, as described in section 1.1.1, which is used for severe tremor. Having various side effects and an important cost, this procedure should be taken seriously. This evaluation process is an aggregation of response to drugs, physical observations and in most cases nowadays a form filled by clinicians. These forms are key elements because they are used to decide whether a tremor is considered as “severe”. If clinicians are trained to this observation and has a common basis with objective limits, this process is being reconsidered in recent literature, in order to provide a less subjective metric for tremor severity. Data mining techniques are used in recent literature.

Maastricht University Medical Center and research motivations

Maastricht University Medical Center is a combination between a Faculty of Health, Medicine and Life, and a University Hospital. The main missions focus on health and prevention, with 715 beds, 7000 employees and 4000 students. In 2015, the hospital hosts 27000 patient hospital admissions, 25000 daycare procedures and 29000 Emergency Room consultations.

The neurosurgery department undertaking the DBS procedure is bilocated in Maastricht and Heerlen in the Netherlands. This department counts 12 neurosurgeons. In Maastricht, a total of 1900 procedures were performed, including 134 DBS, in 2015. These DBS procedures were taken for various indications: Parkinson’s disease, Essential Tremor, Dystonia, Epilepsy, Tourette syndrome and Obsessive Compulsive Disorder.

Pieter Kubben is my supervisor at MUMC+. He is a neurosurgeon staff member since 2014. His clinical focuses are on DBS, invasive epilepsy measurements and robotics. But he is also involved in several IT project as an application developer.

He started to be interested in this tremor severity need of objectivity. The recent PD and ET research field highlights sensors as accelerometers (linear acceleration) or gyroscopes (rotation speed) attached to the wrist as a good way of measuring tremor severity. However, this particular problem offers many research opportunities. This study shows different settings and goals when compared to the other existing studies in the field, see section 1.2.1.
In order to collect the data from the sensors, the researchers need a dedicated application. Existing applications are described later in this study. Pieter Kubben created an open source smartphone or smartwatch application named TREMOR12 (Kubben P., Kuijf, Ackermans, Leentjens, & Temel, 2016). He also collected data from 20 Essential Tremor patients and measured their tremor severity using the actual forms for Essential Tremor, rated by clinicians. The main research goal is to see whether the obtained parameters from the application can be linked to the actual forms for Essential Tremor (section 1.1.2) and predict them, in order to know if this research direction is good to go more in depth. This study is then focused on ET since we only have ET patients, but another objective is to keep in mind that PD would also be likely to be used later, so a flexible design is necessary for our application.

**Big data in the context of healthcare: building our data mining structure**

Big data has quickly developed into an important topic in recent years for academics, industry and any type of organization. There is no universal definition for big data. Gartner (Laney, 2001) created a 3Vs (dimensions) model defining big data: “Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation” (Gartner). This model is represented on Figure 1 below:

![Figure 1: Big data characteristics](source: www.dataasciencemrtal.com)

From this 3 dimensional model, other authors added two more dimensions (Jin, Wah, Cheng, & Wang, 2015), namely:

- Veracity, referring to data quality or messiness
- Value, referring to our ability and opportunities to turn data into value

As Figure 1 shows, big data has no precise limit but does increase over time in terms of Volume, Variety and Velocity. Facing this increasing complexity, it is often the case that Veracity is getting lower, and Value can be largely increased, following the new opportunities offered by the data itself, but also the technique improvement for handling these data. This is of course also true for the healthcare sector.
Various studies focus on big data in healthcare (Fang, Pouyanfar, Yang, Chen, & Iyengar, 2016) (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014). Some challenges and opportunities to improve the decision making process in healthcare units are discussed: personalized care, clinical operations, public health, genomic analytics, fraud detection and device/remote monitoring. This last one is of our interest. It was described (Raghupathi & V., 2014) as the opportunity to “Capture and analyze continuous healthcare data in huge amounts from wearable medical devices both in the hospital and at home, for monitoring of safety and prediction of adverse events”.

Healthcare is a particular field, and since we are working with humans we cannot only use results from big data techniques to choose whether this particular drug or chirurgical procedure is applicable. R. Fang et al. (Fang, Pouyanfar, Yang, Chen, & Iyengar, 2016) proposed a Health informatics processing pipeline as shown below in Figure 2:

The general health informatics processing pipeline elements depicted in Figure 2 can be described as:

- **Capturing / Storing / Sharing**: These steps were already completed when this Master thesis project started. The first step is choosing the raw data extraction method. The existing technologies and the chosen one will be explained in depth in this report. Then after the raw data being captured, it is necessary have to store and share it. In our case these steps are easy since the data volume is 300MB, which is quite low.

- **Analyzing**: This step is where most of these study efforts are located. The *preprocessing* includes cleaning the data, filtering the noise and treating the missing data. *Feature selection* is in fact separated in two steps (Fang, Pouyanfar, Yang, Chen, & Iyengar, 2016): feature extraction, where the initial data is transformed from a cleaned version of the raw data to a transformed and size-reduced data. These extracted variables try to represent the data in a way that provides more information in a smaller and machine learning
compatible dimension. The second step is the feature selection. Indeed, all the variables are maybe not meaningful to the model. Plus, they may have some mutual information that are not useful to put several time in the model, in order to avoid overfitting.

The *Machine Learning* step refers to the various possible techniques as Neural Networks, regression trees and Support Vector Machines, using the selected features as input variables. In this step, we can use the Machine Learning technique for three major kind of analysis:

- **Classification:** Is the process of predicting the category belonging of a particular dataset in predefined classes. Classification is a supervised learning task, i.e. we know the possible outputs. An example is tumor classification.
- **Clustering:** Is the process of predicting the different categories several samples can be classified in. Clustering is an unsupervised learning task, i.e. we do not know the possible outputs, but want to find them.
- **Regression:** Is the process of predicting a numerical discrete output variable. Regression is a supervised learning, the algorithm learn from the observation based on known past outputs. This is what this study will do.

- **Searching:** This step refers to searching new patterns in patient’s information. For example, if one is provided with a signal in the time dimension, you can look for various patterns, features to extract in the time or frequency domain. This step in an ongoing one that will have to be continued after this first study, because so many opportunities of finding useful information in the data exist.

- **Decision support:** This step is the final output that clinicians, doctors or physicians will use to replace their older practices and improve the decision making.
Methodology: CRISP-DM

The most used structuring method among Data Mining users (Harper & Pickett, 2006), CRISP-DM (Cross-Industry Standard Process for Data Mining) was used in this project. The chapters of this report are organized to follow the CRISP-DM model represented in Figure 3.

Figure 3: CRISP-DM model, (Shearer, 2000)

Each chapter of this report and its purpose, corresponding to a CRISP-DM step, is described here below:

- **Chapter 1: Business Understanding**

This step introduces in depth the background and goals of this Master Thesis. First, this chapter focusses on Essential Tremor and Parkinson’s disease tremor symptom, the actual tremor severity assessment techniques and a technological review introducing the various sensors and parameters that are used in the literature. Then, the project settings, including the Capturing method (see Figure 2), used is presented. Some of the possible feature extraction techniques for this particular type of data are then considered, comparing to the closest setting studies existing in this new field. Finally, the process, research and data mining goals are presented. Indeed, at the end of the introduction and the Business Understanding, the background knowledge aggregated should be converted to a data mining problem definition (Shearer, 2000).

- **Chapter 2: Data Understanding**

Once the background is known, the study settings and goal described, the data understanding naturally comes in order to describe the data collected. This chapter provides an overview of the patients characteristics, the raw signals coming from the patient wearable sensor data extracted using dedicated software and the two outputs that this study aims to regress. Then the data quality will be assessed, describing the missing data and its importance, the inconstancies created by the collecting methods and finally the noise implied by the study settings and the wearable sensor.
- **Chapter 3: Data Preparation**

Data preparation is detailed in two main parts. The first part regroups the answer this study gives to the data quality problems highlighted in the second chapter. This data cleaning is important for the extracted information to be meaningful and less biased as possible. The second part is probably where most of the value will be extracted for the research. Indeed, it extracts specific features from cleaned signals, using time, frequency and time-frequency domain tools. In order to avoid overfitting and be adaptable for Essential Tremor (this study) and Parkinson’s disease (future), the best features are selected using a feature selection algorithm. This major second step is called data transformation, and will use the Matlab academic license provided by the Technical University of Eindhoven, as for the whole study.

- **Chapter 4: Modeling**

The previous chapters resulted in a set of meaningful features extracted from the initial signals after transformations. Each feature describe patient’s tremor characteristics, and allow the machine learning models to be used. Based on what literature did, the best performing modelling technique is taken and described. This chapter also presents the metrics and assessment techniques used in order to measure and compare various model later.

- **Chapter 5: Results**

This chapter basically applies what the previous chapters did express. Indeed, the chosen modelling techniques is used in order to show how the improvements brought by this study are affecting the prediction accuracy. For this purpose, a base line model is evaluated, using literature representative features extracted, together with Windowed Fourier Transforms frequency domain tools. This base line model is then compared to other models, showing the study improvements. This results chapter will also compare each of the various physical test performed on patients and each feature separately.

- **Chapter 6: Conclusion, limitations and future developments**

This chapter will recap the main insights obtained thanks to this study, shows its limitations and future developments. As described by Figure 3, a global iteration for the whole process is observable. It highlights the fact that this project is just the first step to enter this problem solution on the DBS process improvement and more generally for tremor severity assessment.
Chapter 1: Business understanding

1.1 Tremor related review

1.1.1 Essential Tremor and Parkinson’s disease

“Rhythmic and involuntary movements of any body part”, this is how (Elias & Shah, 2014) described tremors. Tremor affects progressively every situation in life, creating disabilities at home and at the workplace. Emotional effects can be even more significant than the physical disabilities in some cases (Lorenz, Schwieger, Moises, & Deuschl, 2006), creating social embarrassment, depression, etc.

Tremor is the first neurologic sign, or symptom, that heads patients to consult a doctor (Uitti, 1998) when they think about PD. Most of the patients first think of Parkinson’s disease, but tremors can be implied by many conditions, including dystonia, Essential Tremor (ET), Parkinson’s disease (PD), drug-induced tremor (Elble, 2013). ET and PD are the most common. We are going to use ET tremor data in this project, but PD is kept in mind for future developments. Here below are described in parallel general characteristics and interesting points for the shakings that we are interested in this study.

a) General definition and figures: the most common tremor disorders

Parkinson’s disease (PD) is a chronic and progressive movement disorder, involving the malfunction and death of vital nerve cells in the brain, called neurons (PDF, 2016). This degeneration implies a dopamine diminution (deficit), which is a neurotransmitter, generating movement impairments. It was first discussed in 1817 by James Parkinson (Parkinson, 2002). 1 million Americans are approximately affected by Parkinson’s disease, and 10 million worldwide. Each year 60 000 new PD are diagnosed in the United States, and this is without the thousands of undetected cases. It is the most known movement disorder, because it indirectly causes death. PD is the 14th leading cause of death in US (NPF, 2015) and cost up to 25 billion dollars in the United States only. For example, the most important indirect death reason caused by PD is pneumonia and bronchitis (44%, (Iwasaki, Y., K., Iwasaki, & Takakusagi, 1990)) because Parkinson’s symptoms (not only tremor) cause such difficulties to swallow at some advanced stages that the patient can aspirate foot into the lungs.

Defining ET is not as easy. Indeed, researchers are still defining ET. Causes and symptoms discussions are sometimes diverging because of this unstable definition. But ET is often described as a clinical syndrome, a progressive neurologic condition but not a specific disease (Elble, 2013). For (Benito-Leon, 2008) and others, ET had been described as a benign movement disorder or even a “super-healthy condition”. The more recent literature indicates that ET is not monosymptomatic disorder but the most common neurological disorder among adults, with a degenerative feature. Essential Tremor is among the more prevalent neurological disorders (Louis & Ferreira, 2010) and is the most common movement disorder with more than 10 million Americans affected (Stephens & Stephanie, 2011). ET is widely considered as benign by patients or some researchers. However, this condition cause in some cases disabilities in the everyday life, as writing, eating, cleaning or during work life. The official figures are probably underrated because many affected people have not been visiting a doctor and wait their elderly age, when tremors become more evident.
b) Causes and risk factors: advanced age, progressive and ET genetic transmission

The ET etiology (i.e. set of causes) is not clearly defined and is described as a syndrome of idiopathic (i.e. is any disease with unknown pathogenesis or apparently spontaneous origin) tremulousness. Half of the cases seem to come from genetic mutations, directly coming from at least one parent (familial tremor). Age is described as a risk factor when the adolescence age is coming and when age is above 65 (Elias & Shah, 2014). PD is not affected by family history but its incidence increases with age, particularly after 60 years old. Men are one and a half times more likely to have it than women.

In a recent study, (Louis, Benito-Leon, & F., 2015) state that having ET increases the odds of developing PD by a factor of 4-5.

c) Symptoms: frequency, amplitude, distribution and phase of movement as tremor description

As presented earlier, Essential Tremor is the most common neurological disorder, with only one symptom, whereas Parkinson’s disease presents several symptoms. However, ET patients sometimes develop other neurological signs and symptoms, for example an unsteady gait (i.e. ataxia).

Bradykinesia (Berardelli, Rothwell, & Thompson, 2001) is an important feature of PD, characterized by reduced movement amplitude and spontaneous movements. Rigidity is also discussed in literature and described as the muscles disability to relax normally (Lee, 1989). This rigidity affects the patient ability to move freely. Aches or pain can appear from affected muscles. Impaired balance and coordination or postural instability, is generally not the worst symptom at early stage of the disease but becomes prevalent with disease progression. It is among the most disabling symptoms, causing falls and loss of independence (Kim, Allen, Canning, & V.S., 2013). We already introduced tremor, the main subject of this project, which is probably the most common and recognizable symptom of PD present in around 75% of cases (Hughes, Daniel, Blankson, & Lees, 1993).

In order to focus on both ET and PD tremor physical examination, a typical approach dispensed by (Deuschl, Bain, & Brin, 1998) is generally used, emphasis four indicators: distribution, phase of movement, frequency and amplitude

Distribution

Essential Tremor shakings are the most commonly present in the upper limbs, for at least 95% of patients (Poston, Rios, & Louis, 2009). Again we have some inconstancies in the cited tremor distribution because of the unstable definition of ET, with only hands, voice and mouth cited (Elias & Shah, 2014), (Poston, Rios, & Louis, 2009) or also include legs and trunk (Elble, 2013). PD tremors progressively affect hands, legs, chin and other parts of the body.

Phase of movement

ET upper limbs tremor occurs in a postural way firstly, when the person maintains a position against gravity, such as holding the arms outstretched, but also in kinetics phase of movement (Elias & Shah, 2014), i.e. occurs during movement of a body part. It also happens in resting positions in the severe cases. PD shakings are mostly resting tremors, meaning it occurs in a body part in which the muscles
are not being voluntarily contracted and which is completely supported against gravity e.g. the whole limb resting on a couch.

**Amplitude**

Amplitude is graded from 0 to 4, using the Fahn score (Fahn, Tolosa, & Marin, 1993):

- Grade 0 = none
- Grade 1 = slight = amplitude < 0.5 cm. May be intermittent
- Grade 2 = moderate = amplitude 0.5 to 1 cm. May be intermittent
- Grade 3 = marked = amplitude 1 to 2 cm
- Grade 4 = severe = amplitude > 2 cm

**Frequency**

According to (Fahn, Tolosa, & Marin, 1993), tremor frequency may be classified as follow: low (<4Hz, i.e. 4 oscillations per second), medium (4-7Hz) and high (>7Hz). The classic tremor of Parkinson disease is reported as a resting tremor of 4 Hz to 6 Hz with a severe amplitude and stops after on purpose movements (Deuschl, Raethien, & Lindemann, 2000), when the Essential Tremor shakings are measured at a 7 to 12 Hz frequency (Elias & Shah, 2014) with a low amplitude.

d) **Treatments:** no existing cure, Deep Brain Stimulation as the most common surgery intervention to reduce symptoms effect

Parkinson’s disease currently has no cure, while it is still possible to improve life quality by reducing the symptoms. These therapies are: Drug Therapy, Deep Brain Stimulation (DBS), Lesioning, Radiosurgery and Radiofrequency ablation (PDF, 2016).

Essential Tremor can be affected by medications, even though only 8 percent of the patient had been prescribed medication (Stephens & Stephanie, 2011). These medications are not always satisfying and are judged as useless in 30% of the cases. ET also has its own surgery therapies, with Thalamotomy and DBS.

In both cases, Deep Brain Stimulation (DBS) is possible and this project is motivated by it. Indeed, tremor severity is part of the DBS decision making process described in the section 1.2.1. DBS is used in the most severe cases for ET that are resistant to medical therapies, while it is the most common treatment for PD. The patient’s skull is opened and DBS electrodes are inserted into a specific area in the brain. The implanted device is a neurostimulator that sends electrical impulses. The targets are globus pallidus internus (GPI), ventral intermediate nucleus (Vim) and recently the posterior subthalamic area (PSA) were introduced (Umemura, Oyama, Shimo, & Hattori, 2013). A recent study on the side effects of DBS on ET tremor (Sandvik, Koskinen, Lundquist, & Blomstedt, 2011) resulted in a reduction of 48.4% by using Vim and 58.2% with PSA. The rating scale used for this study is described in the section 1.1.2.

DBS is the most advanced surgery treatment for the symptoms, but again it will not stop the disease progressive aspect nor cure it. Side effects occurs and targeting Vim as an example, Paraesthesia
(19%), Dysarthria (9%), headache (7%), disequilibrium (4%) or paresis (3%) are possible (Itakura, 2015).

e) ET-PD combination: a double strike

ET and PD combination has been a controversial topic in the literature. (Cleeves, Findley, & W., 1988) first said that no association or genetic link between these two tremor types. Recent researches indicate different perspectives on the possible association of ET and PD. The combination ET-PD has not been thoroughly investigated, while it shows to be a double disease strike (Louis & al., 2016), with significantly more tremor than just a PD or ET patient. Cognitive, mobility, balance and sleeps problems are also highlighted.

1.1.2 Actual tremor assessment techniques and technological review

There is no definitive test to assess ET or PD, but several methods exist for tremor assessment apart from the simple clinical observations (Hess & Pullman, 2012):

a) Standardized rating scales: the most used tremor quantification method

A number of different standardized rating scales exist and are used in practice and research purpose. This is the most used way for quantification of tremor. But we are going to introduce as examples Essential Tremor Rating Scale (ETRS) and Quality of Life in Essential Tremor questionnaire (QUEST), that were moreover measured for this project. These questionnaires are used to assess the tremors and help the clinical decisions. ETRS were often used in association with DBS (Sandvik, Koskinen, Lundquist, & Blomstedt, 2011).

ETRS is a 21 questions form (Appendix 1) filled by a trained clinician, so it is partially subjective. Each question has 5 possible answers, also named grades, from 0 to 4 (see the Fahn amplitude score in the previous section). The first 9 questions assess the tremor intensity in the different body part, from “none” (0) to “severe = amplitude > 2 cm” (4). The next 4 questions are handwriting and drawing tests, from what the clinician has to grade the drawing with indications as “Grade 1 = slightly tremulous. May cross lines occasionally”. The 8 last questions are about everyday life hinder on different tasks. Some questions have sub questions, for a 36 grades total (144 points is the maximum score, 36*4).

QUEST is a 30 questions form (Appendix 2) filled by the patient (without training), also with 5 possible answers from 0 to 4. Questions are listed by categories. This questionnaire is supposedly more subjective since it has no limits like the Fahn score but use more the senses and patient personal feelings.

Concerning Parkinson’s disease, the most widely used questionnaire is the Unified Parkinson’s disease rating scale (UPDRS), also based on the Fahn score. This form is filled based on 6 different
parts evaluated by interview and clinical observation (Shargel, Mutnick, Souney, & Swanson, 2009). Interviews are self-evaluation from the patient that are the same kind as the QUEST score.

b) **Objective assessment of drawn figures: a visual way of depicting tremors**

Figure 4 below is an drawn figure example used for objective assessment. In this case, the patient can draw Archimedean spirals on a digitizing tablet, and from the trials an average can be made and several characteristics (frequency, direction, amplitude, etc.) are quantified.

![Figure 4: Objective assessment of drawn figures example, reprinted from (Hess & Pullman, 2012)](image)

The literature describe the tremors as quasi-sinusoidal movements, and very convenient for mathematical analysis and modeling (Hess & Pullman, 2012). Electromyography (EMG), gyroscopic measurements and accelerometry are described in the literature. Techniques exist to detect complex synchronization and relationships between tremors.

Very important characteristics that we mentioned before, amplitude and frequency are measured with computerized tremor analysis. But some studies show that frequency is generally not so hindering for patient (Findley & Koller, 1984) (Marsden, 1984), whereas amplitude and any wave-like form is important (Pullman, Fahn, & Rueda, 1992). Amplitude is accurately assessed by accelerometers, when we use a miniature device or something that do not change the patient posture, kinetic or resting phase of movement. Angular displacement or linear degree of the measured body part (tremor amplitude) is usually measured in millimeters or degrees. Filters exist to block noise or interferences. Joint movement in the affected limb is also important, because tremor are a combination of linear and rotational effects (Grimaldi & Manto, 2010).
d) Technological review – computerized tremor analyses

This technological review is based on (Grimaldi & Manto, 2010), a review from the journal Sensors. It starts by an affirmation: “Although tremor can be estimated clinically, the non-stationary feature and the difficulties related to a pure clinical evaluation (with inherent subjectivity) make the use of sensitive, reliable and stable sensors mandatory”.

Technologies: various types of sensors

The four most widely used types of sensors will be described in this section. It includes electromyography (EMG), accelerometers, gyroscopes, and force sensors. **Accelerometers** are based on Newton’s second law: \( \textit{Force} = \textit{Mass} \times \textit{Acceleration} \). It measures linear acceleration (in g) on a specific axis. Data using accelerometers are a combination of linear acceleration, gravity and additive noise. No analytic model has been validated to separate gravity from linear acceleration. Low-pass filtering is used in most studies (Grimaldi & Manto, 2010). **Gyroscopes** are measurements based on the Coriolis force. The angular velocity is obtained from “a spinning wheel or disc in which the axis of rotation is free to assume any orientation by itself. When rotating, the orientation of this axis is unaffected by tilting or rotation of the mounting, according to the conservation of angular momentum” (Wikipedia, 2016). Thanks to these settings, gyroscopes are a good way of measuring rotation speed (radians per second) and rotation / orientation (radians). **EMG** is an electrical receptor of the muscle fibers activity. There are two types of EMG: surface EMG and intramuscular (needle) EMG. Surface EMG takes the electrical potential from muscles from above the skin, while needle EMG use a fine wire inserted into a muscle, adding a surface electrode as reference. Many laboratories are adopting arrays of SEMG, that can be integrated in textiles for wearable solutions. EMG was successfully used by (Breit, Spieker, Schulz, & Gasser, 2008) in order to differentiate PD from ET. **Force sensors** measure the torque in newton meter (N.m) and angular motion in radians per second. Torque being the force applied to rotate an object around a pivot, or axis. This type of sensor is described as promising but most of haptic devices are non-portable and expensive.

Finally, (Grimaldi & Manto, 2010) proposed Table 1 that compares these most used sensors. This table was adapted for our use. It mainly shows that if force sensors and EMG might have very good results in the future, they are also not always accepted by the patient, expensive or not easy to use. Accelerometers are probably the most used sensors, but also not the easiest way to avoid noise.

<table>
<thead>
<tr>
<th></th>
<th>Accelerometer</th>
<th>Gyroscope</th>
<th>EMG</th>
<th>Force sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gravitational component</strong></td>
<td>yes</td>
<td>No</td>
<td>No</td>
<td>no</td>
</tr>
<tr>
<td><strong>Signal to noise ratio</strong></td>
<td>low to high</td>
<td>High</td>
<td>High</td>
<td>high</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>small</td>
<td>Small</td>
<td>Small</td>
<td>large</td>
</tr>
<tr>
<td><strong>Easy to use</strong></td>
<td>yes</td>
<td>Yes</td>
<td>Variable</td>
<td>relatively difficult</td>
</tr>
<tr>
<td><strong>Painful</strong></td>
<td>no</td>
<td>No</td>
<td>yes (needle EMG)</td>
<td>no</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>cheap</td>
<td>Cheap</td>
<td>Cheap</td>
<td>expensive</td>
</tr>
</tbody>
</table>

Table 1: principal sensor types comparison, adapted from (Grimaldi & Manto, 2010)
1.2 A pilot project at Maastricht University Medical Center

This master thesis is a pilot project in Maastricht University Medical Center, being the first project on wearable sensors and using computerized technologies for measuring tremor severity in this organization in order to advance the research in this new and hot topic. First, the project setting is presented. Next, the close settings projects that can serve as a reference to make some first data mining choices are introduced. Then the long term goals, process oriented will be described. The short terms goals, the ones that this master thesis project directly support will also be presented.

1.2.1 Study setting and goals

a) Study setting: technologies used, parameters and patient tested

In this section of the report, TREMOR12 is first introduced. It is an iPhone or iWatch application developed by (Kubben P., Kuijf, Ackermans, Leentjens, & Temel, 2016). Then the device used and its sensors will be investigated. Finally, the data collection procedure is described.

TREMOR12

In July 2016, TREMOR12: An Open-Source Mobile App for Tremor Quantification was published in Stereotactic and Functional Surgery. This technological report was written by (Kubben P., Kuijf, Ackermans, Leentjes, & Temel, 2016) and was published in the Neurology and Psychiatry department from the Maastricht University Medical Center, in the Netherlands. This master thesis project aims to use data extracted from TREMOR12 application.

TREMOR12 is the first open-source mobile app designed for research settings, with several parameters to explore, liberty of choosing algorithms and expansion possibilities. The device (iPhone or iWatch) TREMOR12 can be strapped to the wrist for convenience and avoid tremor modifications. Some accelerometer-based applications already exist. The free apps either have limited number of parameters or limited expansion. The commercial versions admit no access to the code or already use premade algorithms. Both the free and the commercial apps are not giving access to the raw measurement data, which is really important to make the analysis. Plus, having control on the code is important on a research setting in order to understand and create appropriate algorithms used and have expansion horizons. Plus, simple descriptive analyses are not satisfying, while advanced algorithm can give additional information that we maybe do not even imagine at the first glance. TREMOR12 does not calculate anything on-device, but thanks to the raw data extraction, it gives freedom for various data manipulations. The app were made for research settings only, this is why all these points were important, to provide an app that fills the gap from the already existing others. A proof-of-concept on 3 patients suffering from ET was made in the same paper.
**iPhone 6, sensors and extracted parameters**

To collect the data, an iPhone 6 was used. This version uses two types of sensors previously described in section 1.1.2: an accelerometer and a gyroscope, both on 3 axes (x, y, z). The extracted parameters are:

- Acceleration (in g)
- Rotation speed (in radians per second)

**20 patients – tremor measurements explanations, 5 tests, 2 sides**

This iPhone 6 were strapped to the wrist of 20 Essential Tremor affected. The data content is explored in the Chapter 2 of this report.

Then, all of them did 5 different standard tests described below. Each test were performed on both left and right side. The data collection and design were realized by Aurélie Degeneffe (former medical student Maastricht University) in 2016 together with Nicole Bakker (case manager DBS).

- **Rest**: The rest tremor is being measured in a position were both forearms and hands are resting on a table. Tremor is measured on both sides during 1 minute.

- **Postural 1**: The postural tremor is being measured in a position were both arms are outstretched forward with palms facing down, as depicted by Figure 5. Tremor is measured on both sides during 1 minute.

![Figure 5: Postural 1 position used for measurements](image)

- **Postural 2**: The postural tremor is being measured in a position were both arms are outstretched forward with palms facing down, as depicted by Figure 6. Tremor is measured on both sides during 1 minute.
- **Glass**: The action tremor is measured by raising a cup filled with water from the table, bringing it towards the mouth and putting the cup back on the table. This measurement was repeated three times.

- **Finger nose test**: The action tremor is measured by bringing the index finger to the nose and subsequently bringing the index finger towards the finger of the examiner, as depicted by Figure 7. This movement was repeated three times.
b) Process, research and data mining goals

In the short term, data mining goals will lead to use accelerometric and gyroscopic data in order to measure patient’s tremor severity, then bringing new elements to the wearable sensors area of research. This area will grow and one day might produce enough advancements enabling the horizon of DBS decision making process improvement, in a long term view.

This section will give further explanations concerning what we can expect in the short and long term.

**Long term: process goals**

First, the decision making process is described. The current DBS decision process (operate the patient or not) is a combination of various elements regarding the patient. This decision is made up in a group meeting where neurologists, neurosurgeons, psychologists and psychiatrist are involved. Here below are detailed the elements taken into account by these experts for the decision making process:

- Clinical symptoms: Mostly described by the tremor severity, using the ETRS and QUEST forms in case of ET

- MRI imaging: Magnetic Resonance Imaging is a technique used to map a body part. It uses magnetic fields, radio waves and field gradients, generating the body part image

- Response to medication: the patient’s response to medication is also a factor. If given drugs are not satisfactory or even judged as useless for instance, chances that the patient need DBS is increased

- Absence of contra indications: therapeutic ultrasounds, etc.

Clinical symptoms represented by ETRS and QUEST forms are important part of the decision making process. But they are subjective. QUEST more than ETRS, since it is filled by the patient and not a trained clinician. The main long term process goal is to provide more objectivity by using direct physical measures from wearable sensors. This decision making process is very important for patient’s life, so it would give more legitimacy to the decision, if this new process is reliable. The second goal is to provide better performances, in the sense that we can imagine shifting to an instantaneous ETRS (or another type of measure) obtained after one or several standardized test. Indeed, the instantaneous notion plus the saved time (test itself, training) would reduce the processing time and costs.

As described above, this process can be improved in the long term. It is the case for MUMC, but also any other medical institute that uses Deep Brain Stimulation to reduce tremor severity. Before, the global research area of wearable sensors has to grow. This point will be further explored in the last Chapter of this report.
**Short term: research goals**

Three major research goals can be highlighted. Firstly, this master’s thesis project comes in support to the research for TREMOR12, in order to test it with a larger scale than the first proof-of-concept using only 3 patients suffering from ET. For this reason, in preparation for this master thesis project the article’s authors made measurements with 20 patients using a smartphone sensor/application. Thanks to the open-source nature of TREMOR12 software and this study, other researchers would get an easy access to work on and expand literature in this area.

The second goal is to provide a research direction regarding the regression goal, the outputs that can be used. Indeed, it is important to know if ETRS and/or QUEST can be linked and if they are meaningful. Since some other first studies already used data mining tools, this study also has to dig further the techniques used in order to provide our main answer and adding value to the area. Plus, the setting characteristics of this study are different from the others. While most studies rely on Fahn score, this study has new objectives. Some other studies using the same kind of techniques are focusing on classification problem, for instance on the ET vs PD differentiation. The diversification of the data mining goals is important for the research, since the discussion on the output to predict is not mature enough.

**Short term: data mining goals**

This study will aim at reaching three data mining goals. First creating a base line model based on actual literature (section 1.2.2). Next, the second goal is predicting Essential Tremor forms scores (ETRS and QUEST) from sensors data. This is the main data mining goal, also implying research goals as previously explained. The third data mining goal will focus on bringing value to increase forms predictability. The main elements to be improved are the feature extraction techniques, the feature extracted and the variable selection algorithm that were not existant.

<table>
<thead>
<tr>
<th>Goal number</th>
<th>Goal horizon</th>
<th>Goal type</th>
<th>Goal description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Long term</td>
<td>Process</td>
<td>Improve process measurement objectivity</td>
</tr>
<tr>
<td>2</td>
<td>Long term</td>
<td>Process</td>
<td>Improve process performances (time, cost)</td>
</tr>
<tr>
<td>3</td>
<td>Short term</td>
<td>Research</td>
<td>Validating TREMOR12 on a larger scale</td>
</tr>
<tr>
<td>4</td>
<td>Short term</td>
<td>Research</td>
<td>Providing a research direction by testing accelerometric and gyroscopic signals prediction power on ETRS and QUEST</td>
</tr>
<tr>
<td>5</td>
<td>Short term</td>
<td>Research</td>
<td>Providing research area with new insights in appropriate techniques and usable outputs (linked to goal 7 and 8)</td>
</tr>
<tr>
<td>6</td>
<td>Short term</td>
<td>Data mining</td>
<td>Create a base line model following the actual literature</td>
</tr>
<tr>
<td>7</td>
<td>Short term</td>
<td>Data mining</td>
<td>Predict Essential Tremor forms scores (ETRS and QUEST) from sensors data</td>
</tr>
<tr>
<td>8</td>
<td>Short term</td>
<td>Data mining</td>
<td>Brought measured improvements to the base line model</td>
</tr>
</tbody>
</table>

Table 2: study goals summary
1.2.2 Close setting studies for objective tremor quantification and time series analysis

a) Close setting studies

In this section, the close setting studies will be explored. The goal is to provide the state-of-art basis in the actual close setting literature, to create a base line model. Then we will be able to optimize and dig on new data mining techniques provided from other sources that will be introduced in the chapters related to data preparation and modelling.

The chosen papers were those with the following characteristics:

- Phenomena observed: have a clear and separate analysis based on tremor
- Data samples: were using PD and/or ET patient samples
- Sensors: used accelerometer and/or gyroscope
- Task: regression or classification for measuring tremor severity.
- Rating scale (output) predicted: any of them. Indeed, none study focused on ETRS or QUEST before

Globally, all the studies were following Figure 2-like process described for healthcare applications. Appendix 3 gives the table containing all the relevant information the construction of our project needs in order to define what base-line model this study should first follow.

This Appendix 3 shows that the six retained references were presented between 2009 and 2015, confirming how recent this research area is. Some important elements are highlighted here. Three of the 5 studies focus only on Parkinson's disease dataset. Only one study has a larger dataset than we have. It is important to notice that only one of these studies is not from a healthcare related article, being a thesis on mechanical engineering. This fact is important, because most of the studies are not providing full details on their implementation. The base-line model will be introduced and measured in the Chapter 5: Results. This model will be based on what this literature shows to work best, even though some smaller parts of the previous studies are not fully detailed and will be chosen from other literature.

The literature taken for the base line model is completely focused on tremor measurement severity using accelerometers and gyroscopes as wearable sensors, most of them on the wrist. They are all using Machine Learning Techniques, either in regression or classification. It is possible for tremor severity to regress or classify since the score they all use is from 0 to 4. Some use five different classes, while one regress the score. This score is obtained after the features are extracted and this score is predicted using various kind of model, from simple non-linear model to Support Vector Machines or Hidden Markov model. To build the base line model, we will have to pay attention about the result significance. Indeed, some models have few data and very high accuracy. For example, (Dai et al, 2015) has only 7 PD patients but reach a 98% correlation, we can expect overfitting.
Time series analysis

All these studies are using time series analysis. Many techniques exist in this field, for extracting features from the data. They can mainly be divided in the time domain method, frequency domain and time-frequency methods.

- **Time domain**: includes various techniques based on the raw data, from a standard deviation to an autocorrelation measure. Some simple models can also be used as regressive and autoregressive models. But we are not interested in using these models since the close setting studies already used better algorithms.

- **Frequency domain**: the x-axis will no longer be the time, but the frequency. The frequency is the number of time we observe a periodic phenomenon. In this view, the associated variance with each periodicity, i.e. the variance profile over frequency, is called the power spectrum. The main frequency domain methods are based on periodic, cyclic signals. Frequency domain shows how signal’s energy is spread over a range of frequency. Time domain is also called temporal domain, while frequency domain is described as spectral. It is important to know that while going from the time domain to the frequency domain, we do not lose any information, we just look at the signal with a different view.

  The features are extracted from two different types of estimations: parametric and non-parametric. The parametric spectral estimations are based on a model (Moving Average, as instance). The non-parametric estimation is more adapted to this study, and is based on the Fourier Transforms. Wavelet transforms can also be used in this way. Indeed, the tremor is described as a repetitive and non-stationary phenomena (Grimaldi & Manto, 2010). A non-stationary signal sees its strength to get higher over time. More details are given in the Chapter 3.

- **Time-Frequency domain**: windowed Fourier Transforms and Wavelets are the two possible time-frequency methods. While frequency methods allow having frequency information on the whole time series, the time-frequency methods can localize the power information in time. This is very important for non-stationary phenomena. Indeed, for an increasing tremor over time while holding the position, the classic Fourier Transforms will only model “an averaged” signal and its features. Some information is then hidden.

The various choices made on which feature to extract from the signals will be further explained in the data preparation chapter.
Chapter 2: Data understanding

This chapter will provide data knowledge, in order to understand the data structure and quality. The goal is to be able to get where the data preparation has to bring corrections for later modelling the signals as best as possible.

2.1 Data description: an overview of the dataset

2.1.1 Patients characteristics

All the 20 patients were operated for Deep Brain Stimulation in the previous years, between 2003 and 2016. But the ETRS, QUEST and using TREMOR12 measurements were all tested between March and July 2016. Taking one patient separately, all its measurements were made the same day. Patients were aged of 66.85 years old on average (53-85) when 11 patients were men and 9 women were tested. The disease duration diversity in the dataset is important, from 3 to 56 years, with an average of 22.25 years. The ETRS score goes from 0 to 144 (integers) evaluated by a clinician, while QUEST goes from 0 to 120 (double) assessed by the patient. The dataset is not completely representative since patients ETRS scores are contained between 10 and 68, with an average of 35.15. The patients QUEST are scoring from 3 to 68.1, with an average of 26.8. The correlation coefficient between ETRS and QUEST scores for patients is 72.5 %, meaning that a change of +1 in the score given by the clinician (ETRS) based on practical observations results in a change of +0.725 for the score given by the patient himself for the everyday life consequences of its tremor.

2.1.2 Raw signals

The inputs variable are the raw signals extracted from TREMOR12. The study setting (section 1.2.1) can be summarized as follow:

- Tests = {Rest, Postural 1, Postural 2, Glass, Finger nose}
- Sides = {Left, Right}
- Parameters = {Acceleration, Rotation, Rotation speed, Gravity}
- Axis = {X,Y,Z}

For one patient tested, 10 files are extracted from TREMOR12 (5 tests)*(2 sides). And each file has 13 raws: the time stamps, and (4 parameters)*(3 axis) = 12 set of values. One signal being defined as a discrete set of points in time, we have 12 potential signal per file, and 120 for each patient.

Signals have a 10 milliseconds sample rate (100 Hz, or oscillations per second). As explained in section 1.2.1, the resting/postural1/postural2 positions are supposed to last a minute. But in fact some variations are observed. The signal length, by test type is detailed in Table 3 below. It contains patients and healthy volunteers together. As you can see, Glass and Finger Nose tests are taking far less time than the others. As instance, the Rest test was in average 5234 observations, meaning 52.34 seconds.
The signal representations will be introduced in the next section. Since 2400 signals can potentially be depicted against time (120 signals and 20 patients), this complexity and its specificities will be better explained while expressing data quality.

2.2 Data quality assessment: dataset exploration

2.2.1 Missing data and inconsistencies

   a) Missing tests (files)

Because “‘Big data’ very often means ‘dirty data’ and the fraction of data inaccuracies increases with data volume growth” (Mirkes, Coats, Levesley, & Gorban, 2016), this study is not an exception, and should be carefully inspected in this respect.

Having 20 Essential Tremor patients, each of them potentially having 10 tests recorded (5 different tests, 2 sides), a total of 200 files should be registered and ready for use. But 9 files or tests were not recorded (4.5%), mainly because they were the first wave of tests, all realized the same day. Indeed, the tests design evolved a bit after this first wave, from one kinetic test to two kinetic tests (glass and fingernose).

   b) Sampling errors

Another quality problem occurs, this time with the data sampling of the sensors. As depicted by Figure 8, some parts of the signal are alone before or after the recordings: 295 in total on the whole data set. Figure 8 shows 124 data points (1.24 seconds) of not expected recording at the signal hard right. Of course, it is potentially possible that the person taking the measurements registered a very short sample and forgot to erase it, but it is not most of cases. In fact, these little pieces of signals are separated in 6 on Figure 8, even though it is hard to see it. It explains the 295 sampling errors since each error is split in several little pieces.

<table>
<thead>
<tr>
<th>Test</th>
<th>Rest</th>
<th>Postural 1</th>
<th>Postural 2</th>
<th>Glass</th>
<th>Finger nose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounded Average signal length (number of observations)</td>
<td>5234</td>
<td>5176</td>
<td>5457</td>
<td>1967</td>
<td>1668</td>
</tr>
</tbody>
</table>

Table 3: Average signal length by test
To detect these small pieces, the signals were separated, cutting it when at least 100 milliseconds (check if not changed the 100 ms) separate them from the other observations. The sample rate is supposed to be 10 milliseconds. The 100 milliseconds were taken because it was a good balance between detecting the sampling errors and not cutting the regular signals. Indeed, it also happens that the regular signal (on the left on Figure 8) has sampling errors. For example, instead of being 10 milliseconds, it could be 20 milliseconds when a sample is missed by the sensor. For instance, the time shift between two observations is around 10% of the time larger or equal to 20 milliseconds out of Patient 7 – postural 2 – left side signal.

c) Clicking “start” and “stop” button problem

In TREMOR12, when the measurements were collected, starting and stopping the measures recording imply problems in the data. Indeed, since the patient has to click the button with its other hand, it creates some little inconstancies and peaks. Plus, some patients may not have taken the exact starting position, creating unexpected little movements a bit after clicking. Figure 9 illustrates this problem in the time domain.
2.2.2 Noise

Noise during recording the data is an important problem. As we previously introduced in the tremor medical review and the sensor review, noise occurs because of the gravity component (Grimaldi & Manto, 2010) in the accelerometer signals. Another source of noise is the voluntary motion (Querlioz, Thiriez, Jarraya, & Palfi, 2015) for kinetic tests: glass and finger nose test. Other symptoms, mainly when it comes to Parkinson’s disease are also important, such as dyskinesia (Dai, Zhang, & Lueth, 2015). These noises are creating a combination of noises, which is illustrated by Figure 10 and Figure 11, in the time and frequency domain respectively. In Figure 10, voluntary motions created by the Glass test are easy to see. Indeed, this test is made three times in a row, and 3 main peak areas seem to be repeated, when looking at acceleration.
Figure 10: Time domain noise from Patient 18 – glass test – right side – linear acceleration – axis X

Figure 11 shows major peaks in the low frequencies, where the noises we talked about are generally considered, when ET tremor are considered between 7 to 12 Hz. This noise is spread over the spectrum, strong at low frequencies, appears in every single frequency domain plot signal. More details will be given in the filtering part included in the next chapter: data preparation.

Figure 11: Frequency domain noise from Patient 18 – glass test – right side
Chapter 3: Data preparation

Data preparation is the major step in this study. It contains data cleaning and data transformation. Data transformation is split in feature extraction and feature selection, where most of this paper value is extracted.

3.1 Data cleaning
This first part of the chapter will resolve the data quality problems presented in chapter 2.2 Data quality assessment: dataset exploration. Data cleaning is an important step in order to have signals that give the most accurate possible information, representing best the physical phenomena measured by the wearable sensors. The four data quality concerns are:

- Sampling errors
- Starting/ending clicking problem
- Gravitational, motion and disease related noise
- Missing files
3.1.1 Inconstancies correction and general organization

This section will explain how the study handles the two first problems in the list above. This section on inconstancies correction cannot be compared to the close setting studies discussed in section 1.1.2, since the concerns completely depend on the study settings. Indeed, using another device, software or type of test would imply other problems.

Figure 12: Inconstancies cleaning process

Figure 12 illustrates inconstancies and cleaning management, from the initial 191 patient files. These files were imported into Matlab in a dataset named Data Raw, in a structure summarizing the important characteristics of the files and the signal. In between, the data three axis for linear acceleration and rotation speed are summed up. This is done in order to reduce the dimensionality, improve the generalization power by being less prone to orientation problem together with suppressing the gravity later on and. (Woods, 2012) is also reducing dimensionality in such a way. It created the dataset called data Merged. Then the sampling errors discussed earlier are treated. Having the time stamps, the time shift between the different observations is calculated. The signal is cut when the time shift is > 100 observations (1 second), obtaining the Data Separated dataset. 295 new signals are extracted from this operation. Most of them are small signals as depicted. Finally, 50
observations were suppressed at the end and beginning of every signal, to resolve the starting/ending clicking problem.

3.1.2 Filtering: tackling the noise
   a) Types of filters: FIR vs IR

Concerning discrete-time signal filtering in signal processing, there are two categories of filters: the recursive, also known as infinite impulse response (IIR), and the nonrecursive filters, generally called finite impulse response (FIR) (Schafer & Oppenheim, 1999). These names come from their properties. Indeed, the FIR creates outputs in a finite duration, while IIR has an internal loop and may respond indefinitely in some circumstances. These properties are best shown by observing their difference equation, showing how the output sequence is obtained:

\[ y[n]_{FIR} = \sum_{k=0}^{n} b_k \cdot x[n - k] \]

\[ y[n]_{IIR} = \sum_{k=0}^{n} b_k \cdot x[n - k] - \sum_{k=1}^{n} a_k \cdot y[n - k] \]

Where \( y \) is the outputs, \( x \) the inputs, \( a \) and \( b \) are weights and \( n \) the sequence of value index and \( k \) the addition index.

FIR equation obtains the output sequence \( y[n] \) by a weighted sum of the input sequence \( x[n] \), starting from the most recent ( \( x[n-k] \)). IIR has the same first element, to which is subtracted the internal loop based on the output, explaining the recursive dimension of this type of filter.

FIR and IIR can be compared on various criterions (Schafer & Oppenheim, 1999):

- **Computation time**: as explained, FIR has finite duration but required more time for the same level of performance

- **Stability**: IIR may have unstable behavior and lead the output to grow to infinity, while FIR are stable in any case

- **Design**: FIR filters are easier to design. In the case it is not known how to specify the filter parameters, there is a procedure to follow. First finding the type of filter that suits best the desired application. Then optimizing the parameters with an algorithm, like (Chan, Carson, Pun, Yeung, & Ho, 2002) did present using a least square approach. But this study aims not to optimize the answer given by a specific dataset, but to learn from this data in order to later predict ETRS or QUEST outputs, and to do it with a filter that represents best the real phenomena. Optimizing in this case could result in overfitting and to non-natural values

- **Phase**: IIR filters create less delay in the signal because they need a lower filter order for the same performance level, but they generally has non-linear phase, creating an additional noise in the output sequence. FIR can easily provide linear phase, but has a stronger delay
In the close setting studies highlighted in section 1.2.2, the filtering choices are not fully explained. (Dai, Zhang, & Lueth, 2015) paper states they used both IIR and FIR. (Rigas & Tzallas, 2012) used IIR without giving more reasons for their choice or which particular design they used. (Patel & Lorincz, 2009) used the elliptical IIR.

In this study, important characteristics for the filters are:

- **Bandpassing:** this study does need to keep the filters between 7 to 12 Hz, as Essential Tremor shaking is generally observed (Elias & Shah, 2014). Other papers state different frequencies for each kind of tremor (Grimaldi & Manto, 2010). Indeed, rest tremor is said to be typically in the 3-6 Hz, postural from 4 to 12 and kinetic from 2 to 7. But these values have to be taken carefully. For example rest tremor is mainly present for Parkinson’s disease, and the severe ET cases are really rare in comparison. Plus, experiments were led in order to see whether these limits make sense. It resulted that all tests separately were better representing the outputs with the 7 to 12 Hz limit. The assessment measures used for evaluating are described in the section 4.2.

- **Phase:** (Popovic, Sekara, & Popovic, 2010) states that the delay is not important. Indeed, studying tremor is about understanding the amplitude or tremor frequency from the features extracted. This delay is described as “negligible” on the overall performance. In the other hand the linear phase characteristic is important. A non-linear phase would change the tremor waveforms and mat result in very different features extracted. For example, if the extracted feature is the maximum magnitude of the tremor on a particular signal using frequency analysis, it would change the power of the signal over frequency.

Matlab signal processing toolbox gives access to both IIR and FIR design. While the Bandpassing requirement is probably better realized by an Elliptical IIR filter (Figure 13) that has sharper cutoff bands for less computing efforts, the fact that FIR filters can provide linear phase and that the delay is not important will make the decision. The study does not need the cutoff band to be as sharp as Figure 13 depicts, but it is definitely more important to keep the tremor waveforms as much realistic as possible and keep the extracted features representative of the physical phenomena by avoiding non-linear phase, also called ripple. This ripple is zoomed on Figure 14.
Figure 13: Example of bandpass Elliptical IIR design

Figure 14: zoom on Figure 13 ripple
c) **FIR filter method used: the Equiripple**

The widely used Parks-McClellan method created in 1972, also called Optimal, Minimax or Equiripple filter were used. In fact, it is a method for finding the optimal coefficients a and b, minimizing the error in the pass and stop bands.

Known as Equiripple in the Matlab library, this method is now a standard for designing FIR filters. To compare with the Elliptical filter design previously presented, Figure 15 and Figure 16 illustrate the linear ripple created by FIR filters and the less sharp cutoff bands. It is possible to make it sharper, against a greater computing time.

![Figure 15: Example of bandpass Equiripple FIR design](image)

- 36 / 76 -
After section 3.1.1 on inconstancies correction and merging the 3 axis, the filtering part helped again to improve signal representativeness. A good way to observe the change is by looking at the signal difference before and after these inconstancies and noise removal in the time domain (Figure 17), then to the frequency domain (Figure 18: Before/after signal cleaning in frequency domain on Patient 9 - postural - right – rotation speed). Figure 17 shows all the benefits extracted from this signal cleaning in the time domain. The signal started with an important sampling error on the hard left without meaning, finished by an important peak and had some noise because of gravity and motions. On the extreme left of the bandpassed signal (red), the filter delay is shown. As previously said, it is said to be negligible on the overall performance by the literature.

Figure 18 is in the frequency representation of the same signal (Figure 17). This figure shows that all frequencies not in the bandpass were filtered. Additional noises in the bandpass were removed by the inconstancies removed. (Querlioz, Thiriez, Jarraya, & Palfi, 2015) considered that the signal part under 3 Hz (3 oscillations per second) as representing these gravity, voluntary motion and dyskinesia noises.

Figure 19 illustrate the noise produced by an ET patient on a kinetic test (Fingernose), this time based on the rotation speed (axis X). It shows that filtering here detrend the signal and center it around 0 value.
Figure 17: Before/after signal cleaning in time domain on Patient 9 – postural 2 - right – rotation speed

Figure 18: Before/after signal cleaning in frequency domain on Patient 9 - postural - right – rotation speed
3.1.3 Missing data imputation

Missing data imply a choice between deleting and imputing the data. While deleting techniques (complete-case analysis) imply both bias and a lower statistical power, imputing techniques only imply a bias (Cismondi, et al., 2013). In our case, collecting the data is time demanding, and only a small part of the data is missing for the concerned patients or volunteers: kinetic positions. In case of deletion, the statistical power reduction would be too important. Indeed, while only 9 tests (on 200) and 4.5% of the dataset is missing. It concerns 4 patients, which is 20% of the data set, then deleting patients data would mean losing too much information. Plus, it mainly affects kinetic positions when ET is said to be more about postural positions by the literature. Imputation is then preferred for this study.

Several imputation techniques are described in the literature (Pigott, 2001):

- **Single-Value imputation:** the missing value is filled with a plausible one, the mean of the available cases being a simple and widely used example. This method generally implies underestimated variance (Little, 1992) and is always biased. Other single value imputation techniques imply interpolation. This interpolation can be based on linear, spline, on the next or previous non-missing value

- **Model-based methods:** these so-called methods use models in order to give the input variables as much predictive power as possible on the outcomes. The two main methods are the maximum likelihood and its Expectation-Maximization (EM) algorithm and the multiple imputation technique. Using these methods implies two assumptions: the missing data mechanism is ignorable, meaning the missingness is conditional on another variable (i.e. MAR: Missing At Random). The second assumption is that the joint distribution of all variables (inputs and outputs) is multivariate normal. It means that the input variables must be normally distributed conditional on the output variables
If model-based methods are described as more promising, our application does not respect both assumptions. In order to use a better single-value imputation than using the mean value of each input variable, a spline interpolation available in Matlab library is realized to fill the missing data. Figure 20 illustrates what is the effect of such an implementation on a two data point missing values (in red) out of the 40 samples (20 patients and two sides). It fills a missing value with the point that would exist if a spline continuous function would exist to represent the input values.

*Figure 20: Missing data imputation (*) marks using interpolation technique on a specific dominant magnitude input values (horizontal axis) against the patients ETRS scores (vertical axis)*
3.2 Data transformation

Once preprocessed, features or attributes have to be extracted from the data. Indeed, machine learning models cannot use complete time series signals in order to predict a known output, in the supervised learning case. Features are a new data representation that can be used as model inputs, binary, categorical or continuous values. A good data representation is a very specific and related to available measurements problem (Guyon & Elisseeff, 2006). In these problems, it is necessary to turn the data in a set of useful features.

This first step is complementary with the feature selection approach. Indeed, when the number of calculated features is important, selecting only the set of input variables providing the best results possible is a key element. It also improves model performances and reduces computation time (Chandrashekar & Sahin, 2014).

In this project, a very large number of features can be extracted from the preprocessed data while the data set size is rather small. In order to avoid overfitting and obtaining new insights in the predictive tests, parameters and features, such combination of feature extraction and feature selection is desirable. The data transformation will then provide a relevant set of features to be used as input variables.

3.2.1 Feature extraction

Some close setting studies earlier made were described in section 1.1.2. These studies on the tremor severity quantification did extract a limited number of features (Appendix 3), without the use of feature selection algorithm. Some of these features are used in this study.

This section will express the representation needs, the answer provided to these needs using some of the previous studies features extracted and new features also. The various tools used will also be described.

a) Signal representation required

Many methods for time series analysis exist. For this reason, it is important to clearly identify which characteristics of the observed phenomenon it is important to observe.

The medical review introduced in chapter 1 gave the four main characteristics for tremor description: frequency, amplitude, phase of movement and distribution. Frequency and amplitude are directly measurable from the signal. Having acceleration and rotation speed is very important to bring value to our study. Only (Dai, Zhang, & Lueth, 2015) with a small data set (7 observations) previously used both acceleration and rotation speed, while both parameters are very complementary, representing better the physical phenomenon with a linear measure and a joint rotation measure. This combination is a good tremor movement description (Grimaldi & Manto, 2010). Concerning the phase of movement, it was not possible to split the 5 different tests from patients as 5 different observations since phase of movement brings complementary information on the patient tremor...
severity. The distribution observed is here limited since only hands tremor is observed. This measure is assumed to be a representative since 95% of the ET patients have hands tremor (Poston, Rios, & Louis, 2009).

In order to represent the tremor as best as possible using the time series (signals) provided by experimentations, time domain tools are not enough. The frequency domain tools are the most important tools in this study to complement the information extracted from the signals.

Time-frequency features will be useful in order to have a measure of the non-stationarity of the signals that could potentially vary depending on the disease advancement. But all existing time-frequency features do not appear as necessarily useful tool in order to extract the required information pointed out since the signal is also filtered between two precise frequencies (7 to 12 Hz). Indeed, most features like energy distribution (Omerhodzic, Avdakovic, Nuhanovic, & Dizdarevic, 2010) are measuring the energy ratio between the different frequency bands obtained by using the wavelets (see wavelet description below for more explanations). In our case, the signals are filter bandpassed during the pre-processing because we do need to use a specific bandwidth (Elias & Shah, 2014). For some applications, observing the signal difference in the various scales is meaningful, but Essential Tremor and Parkinson’s disease do create particular tremor frequencies, the other frequency bands being noise.

b) Fourier Transforms: frequency domain tool

Fourier transforms are the technical basis used for the Windowed Fourier Transforms. Fourier Transforms is a frequency domain tool summing up sinusoids (see Figure 21) in order to represent the initial signal in time. Frequency or spectral domain representation shows how signal’s energy is spread over a range of frequency. Going from the time domain to the frequency domain is the transforms job. By doing so no information is lost, the signal is only observed from a different perspective. The spectral 2 dimensional representation is called the power spectrum. The x-axis is the frequency and the y-axis is the magnitude (power). Then, the power spectrum shows the power composing every frequency bin. An example is given on Figure 18. This information is obtained by the use of the summed up sinusoids (Figure 21).

Figure 21: Summing up two sine waves example
On this Figure 21, the sine waves are summed up for each time bin. Their specific forms are created thanks to 3 parameters:

- **Amplitude a**: The amplitude is the maximal range observed on the y-axis. On this example, the amplitude is 1 in sine waves 1 and 0.5 in sines waves 2. Sine waves 2 are filtering sine waves 1 in terms of amplitudes.
- **Frequency f**: frequency is the number of complete waves in the time observed. Frequency is 2 in both cases here above.
- **Phase p**: The phase is the horizontal offset or lag at the starting point. Sine waves 1 have no phase, while sines waves 2 are lagged from π (pi), i.e. half a complete wave.

Of course, we can create more complex waves from these combinations by adding more waves with different features. Each sine has the following equation:

\[ y(t) = a \times \sin(2 \pi \frac{f}{T} t + p) \]

Where \( t \) is the time variable and \( T \) is the time range, 10 on Figure 21.

This function \( y(t) \) is subsumed is the following equation, in order to transform the signal into frequency:

\[ X(f) = \int_{-\infty}^{+\infty} y(t)e^{-i\pi f t} dt \]

Spectral analysis is a fundamental tool for the understanding and use of many signal and image treatment techniques. It is extensively used in numerous filtering applications such as vocal recognition, musical filtering, radio and televisual diffusion or the biomedical sciences.

At the first glance, the Fourier Transforms seem to be the perfect match since this study does need global feature for the whole signal and not specific events in time. But the information is not fully accurate in the case of nonstationary signals, like tremor is (Grimaldi & Manto, 2010). A (weakly) stationary assumption (Shumway & Stoffer, 2011) means the process has:

- Has a finite variance process
- Has a constant mean value function \( \mu_t = E[x_t] \) over time
- The autocovariance function \( \gamma(s,t) \) depends on \( s \) and \( t \) by their difference \(|s-t|\)

These conditions are not met in the case of tremor, because the tremor variance and mean is most of the time increasing during the test. For this reason, a model where sinusoids are spread the same way on the whole time stamp is not enough to provide the whole information. After the best match, highlighting the need for time-frequency transforms, such as wavelets.
c) Windowed Fourier Transforms

The FT definition was: \( X(f) = \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft} dt \), where \( y(t) \) is the sine function used to represent a signal in time and \( f \) the frequency. The Fourier transform is the scalar product at the point \( f \), between \( y(t) \) and \( e^{-i2\pi ft} \). Additions are resulting from this scalar product. And additions are commutative, meaning that the order we sum up the different products is not important, meaning that all points from the function are weighted the same way. Then, information is spread out and may be hidden.

From that point, the STFT uses function \( w(t) \) to weight up the summations and giving an importance to the order it is processed:

\[
X(f, \tau) = \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft} w(t - \tau) dt
\]

Where \( w(t) \) is the window function, that we have to choose and \( \tau \) is the window translation parameter (time). This is a convolution. In fact, the regular Fourier Transform uses a rectangular window where all points are one-valued. This basic window type is not the best match for most applications, since it does create important side lobes for each frequency component, which are smaller repetitions of the “real” peak (Cerna & Harvey, 2000). As previously stated, some information is lost (i.e. spectral leakage), because these side lobes exist for every frequency component and then are disturbing the power spectral distribution over frequency. This is also the reason why filtering was an important task, while reading only the desired frequency band from the computed power spectrum was possible.

Many other window functions are available, such as triangular, Hann, Hamming, etc. Hamming function was selected as previous close setting studies did (Appendix 3). This window type is well fitted for tremor waves (Grimaldi & Manto, 2010) due to its shape, and was empirically tested as the best.

d) Welch method: reducing windowed Fourier transform noise

Welch method (Welch, 1967) is an improved method that uses windowed Fourier transforms, based on an averaging process. Indeed, it separates the pre-processed signal in \( n \) equal length sized signals, windowed and then averaged. It reduces the frequency resolution but reduces the noise observed on the power spectrum by averaging. Welch method also reduces the computation time and is said to be convenient for nonstationary signals. The method assumes that each signals are a composition of the natural tremor and its noise, repeated \( n \) times equally. Among the close setting studies, only (Querlioz, et al., 2014) is clearly stating the use of this method, but most of them are indicating the window length and overlapping segment length used. Indeed, in this method, some parts of the neighbor signals are replicated in order to enhance the averaging effect. No clear rule is existing concerning the overlap %, but is generally less than 70% in order to avoid that too much averaged signal parts are the same. If none of the signal window is overlapping, it refers to the Bartlett method.
In order to enhance the naturally reduced frequency resolution by the Welch method, it is possible to use zero-padding. This is a simple computational trick, adding zeros at the end of the signal (here of each signal), and naturally improving the frequency resolution (distance between two adjacent frequency points in the power spectrum axis) as it grows with larger sized signals. The frequency resolution can simply be calculated by $F_s/N$, $F_s$ being the sampling rate and $N$ being the number of data points.

\[ \text{Frequency resolution} = \frac{F_s}{N} \]

\( e) \text{ Wavelet transforms: time-frequency domain} \)

\( \text{Decomposition process: Mother wavelets and sons} \)

To process a wavelet analysis, we first have to choose the “mother wavelets” (or function) and “analyzing wavelet” (Graps, 1995). The original signal observed can be represented by a linear combination of the mother function and then allow use to use only the resultant coefficients for the data computation (calculus). These coefficient results in a scaled version of the mother wavelet, that is also translated. Figure 22 illustrates what it means and how it is different from the Fourier transform.

\[ \text{Figure 22: Fourier (left) and Wavelets (right) basis functions modifications, reprinted from (Ha & Romberg, 2013)} \]

Using the Fourier basis function, we adjust some parameters on the whole interval, as more detailed in the previous section of this report, in order to obtain information. For wavelets, we segment the time in multiple smaller scale wavelets, all with their own function coefficients derived from the mother function. It then allows us to look locally scaled-translated wavelet by wavelet, having different windows to look in (Figure 23). The basis/mother function used on Figure 22 is one of the many possibilities that are offered by the wavelet analysis, whereas Fourier transforms are only based on sine and cosine.
As previously mentioned, wavelets can be used from many different mother functions in order to find the one that suit best a particular application. These functions are easily invertible. A wave is a oscillating function with a zero average value, called $\psi$, with a certain regularity degree and a finite support (explaining the “wavelet” name, meaning little wave). Some examples are given in Figure 24.

We will briefly introduce how these basis functions are used in order to calculate the coefficients, using the simplest basic function: the Haar function (Haar, 1910).

Wavelets coefficients are the results of a scalar product between the signal and the different wavelets $\psi_{uls}$, where $u$ is the time parameter and $s$ the scale parameter. Wavelets $\psi_{uls}$ are derived from the mother wavelet $\psi$ translated by $u$ and scaled by $s$ (dilated when $s>1$ and contracted when $s<1$).

**Graphic representation**

Two representations are given with Haar wavelet, using a sine function and a more complex chirp function. Both are represented using time, frequency (Fourier transform) and time-frequency (wavelet transforms) domains. To understand this graph, we need to cope with the following elements:
- **Time:** the X-basis, decomposed in several time frames as previously explained.
- **Coefficient representation:** black, grey and white are respectively the coefficient encoded colors of positives, zeros and negatives values (no other value in case of Haar wavelet, see Figure 24)
- **Resolution (or scale):** is often depicted as the frequency reciprocal. For example, a black point at time 600 and resolution 100 specifies that the function variates around the frequency 1/100 at time 600. Frequency can also be directly depicted on the y-axis

![Image of time, frequency, and time-frequency representations](image)

*Figure 25: Representation of the complex chirp function in time, frequency (double-sided) and time-frequency domain, reprinted from (Charles C., 2011)*

The sine function drew on this Figure 25, is a simple function with amplitude 2, frequency 2 and no phase of movement. The time-frequency domain (third raw) representation on this first line shows that the value is going from positive to negative values (from black to white) gradually, without concern for the resolution/frequency. Which is logical since this function is stationary.

The chirp function depicted on the second line can be read using the time-frequency representation. Indeed, we can see a large black and white colored space in the top left corner. It means that at the start, this signal has a relative low frequency and switching (oscillating) from positive to negative values. These positive and negative coefficients are then gradually in time going to a lower resolution (high frequency) while their strength is decreasing. The periodogram would not provide the same level of information. Indeed, we can only say from the second raw graph that we have many high magnitude low frequency cycles, and the more we get away from low frequency, the more magnitude is low. But we do not know at anything about its repartition in time.

*Applications and Computational efficiency: wavelets are faster*

Plus, wavelet transforms is also known to be faster. The Fast Fourier transforms require $O(n \log n)$ computational operations (Strang, 1993), with $n$ being the length of the spectrum. The Fast Wavelet Transform calculates the required coefficients using $O(n)$ operations, where $n$ is the number of decomposition we did.
f) Features extracted

Table 4 summarizes the extracted features. Each feature is extracted from the 5 different tests (rest, postural 1, etc.) and the 2 different parameters (linear acceleration and rotation speed), so that each feature creates 10 possible input variables. These 50 potential input variables are normalized, between 0 and 1.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal RMS</td>
<td>Time</td>
</tr>
<tr>
<td>Signal Period</td>
<td>Time</td>
</tr>
<tr>
<td>Dominant magnitude</td>
<td>Frequency</td>
</tr>
<tr>
<td>Dominant frequency</td>
<td>Frequency</td>
</tr>
<tr>
<td>Power growth</td>
<td>Time-frequency</td>
</tr>
</tbody>
</table>

Table 4: Features extracted

- **Signal RMS**: the signal Root Mean Square were used by two of the previous close setting studies (Darnall, et al., 2012) (Patel & Lorincz, 2009) in order to represent the signal strength using the time domain. This could also be realized using as instance the energy of the discrete signal, calculated by $E = x_1^2 + x_2^2 + x_{n-1}^2 + x_n^2$. But all the measures having a different signal’s size, it is not appropriate for comparing them. A very popular and reliable measure is the signal RMS (Caetano & Rodet, 2011), calculated as follow:

$$E = \frac{1}{N} (x_1^2 + x_2^2 + x_{n-1}^2 + x_n^2) = \frac{E}{N}.$$  

- **Signal period**: the signal period is calculated by the mean difference between the positive peaks in the signal using in the time domain. Of course, this measure is proportional to the mean frequency of the signal, giving the average number of oscillation per second. The signal period $T$ is indeed used to calculate the mean frequency $(1/T)$.

- **Dominant magnitude and frequency**: this feature is obtained from the frequency domain. The highest magnitude peak over frequency is said to be the dominant or peak magnitude, while its associated frequency is the dominant or peak frequency. It does give which frequency is more present in the signal, at which power. These measures are used in all the six others close setting studies presented in Appendix 3.

- **Power growth**: in order to give insights in how much the tremor amplitude is growing during the test and to represent the non-stationarity (Grimaldi & Manto, 2010), the power growth is calculated. The calculation is based on the difference between the signal power spread on the signal second half, minus the signal power spread on the first half. This measure can be obtained by using the wavelet spectrum only, where the power is described over frequency at every point in time.
3.2.2 Feature selection

Feature selection is an important step in this study, because a very large number of features can be extracted from the potential 120 signals per patient. As (Chandrashekar & Sahin, 2014) said, “Feature selection techniques show that more information is not always good in machine learning applications”. Plus, having a small number of patients in this application, a large number of features selected as inputs for the models also mean the risk of overfitting is important. In this application, a space dimensionality reduction will also help to find where the relevant information is hidden in the signals. Finally, for a future Parkinson’s disease application, the relevant information is not expected to be the same as the Essential Tremor this study will present, giving the researchers the flexibility of changing the input data as well as the output regressed and still find a relevant model.

Feature selection methods can be divided into two main categories (Kohavi & John, 1997):

- **Filter methods**: Filter methods are looking for the best features based on criterion like distance, correlation, consistency, mutual information, etc. They rank the features and filter the “less relevant ones”, based on the criterion. These methods are said are computationally fast and can be very good in practice using particular criterion.

- **Wrapper methods**: try to find the best features that suit the outcome variable from the training data, based on a performance criterion. These methods are in general computationally expensive but provide good results.

In case of classification, a mutual information filter might be the best choice (Peng, Long, & Ding, 2005) as it generally provides very good results and is computed in a few seconds. But as (Frénay, Doquire, & Verleysen, 2013) noticed, the results can be different in case of regression. Mutual information were tested and compared with a wrapper approach and performed less in term of accuracy. The wrapper method used is longer to process, but the computational cost is not the main focus in this study.

This wrapper method is the Sequential Forward Selection (SFS). This method is based on an incremental algorithm. At the first iteration, the best feature depending on a performance criterion (see Assessment methods in section 4.2) is selected. From that point, every iteration try to add one by one the features not already selected and choose to add the one giving the best subset. Then the algorithm stops when no more gain is obtained from adding a feature. A tolerance can be added to stop the algorithm when the improvement is too small.

The best feature is chosen by the use of the modelling technique for which the feature selection is used, together with the Mean Absolute Value and leave-one-out Cross Validation described in section 4.2.

This method has the advantage of keeping as much feature as necessary to obtain the “best result”, integrating Cross Validation to avoid overfitting. SFS is a good technique to provide insights in which tests, parameters, features combinations are relevant.
Chapter 4: Modelling

After presenting the extracted features from the pre-processed signals, this chapter focusses on modelling. While the first part treats the model algorithms used, the second part is about the assessment methods.

4.1 Modelling technique

As described in the close setting studies review, (Darnall, et al., 2012) presented a comparative study showing that Random Forest and Regression trees equally represented best the signals among six classifiers: Random Forest, Decision Tree, Nearest Neighbor, Bayes Multilayer Perceptron, and Support Vector Machine. In this study, a regression tree technique is used. Random forests are based on regression trees, using a double randomness to enhance regression trees lack of robustness and generalization power (Breiman, 2001). The first randomness used is bagging or boostrapping, coupled with a squared error measurement. Leave-one-out cross validation is preferred, since the small dataset size allows to try all combinations with only one measurement out of the training set, in order to more accurately measure and generalize what would be the error in case of a prediction based on one patient’s measurements. Also, other measurements are preferred that makes more sense for the error understanding, as the next section will describe it. The second randomness is also not useful in this study case. Indeed, it is at each node split of the regression tree a random selection of variables. In this study, a feature selection algorithm was implemented, largely reducing the number of useful features in order to predict best the outputs. Indeed, too few variables would decrease the obtained robustness by the randomization. Indeed, trees too highly correlated tend to overfit (Louppe, 2014). Not using random forest in also a good thing for the feature selection model. Indeed, the result randomness for each adding variable iteration would lead to unstable results, and very long processing time.

A regression tree (Breiman, Friedman, Olshen, & Stone, 1984) is a machine learning algorithm where the feature dimensions are split in a binary way, as illustrated by Figure 26, creating a structure visually similar to trees, with nodes (or leaves). This figure illustrates a tree example with two input variables (i=2) or features $X_i$, in order to predict $Y_i$. All values are normalized between 0 and 1 in this example.
This step-by-step construction is directed by looking for the best split in a branch that minimizes the error. Indeed, creating a new node suppose finding the best split value $S$ separating $X_{ij} \in \mathbb{R}$, $i$ being the observation index for each feature or dimension $j$, that satisfies the following equation:

$$\min \sum_{i : X_{ij} > S} (y - y_i)^2 + \sum_{i : X_{ij} \leq S} (y - y_i)^2$$

The equation is in two parts, being the left and right sides of the splits. Each side is also call a cell. The realization of this equation splits the node in two more leaves, providing two new research spaces as Figure 27 illustrates it on a two dimensional graph including only one feature for simplicity. The algorithm is recursive, a certain number of remaining points in the cell most of the time being the stopping criteria. Each cell contains a certain number of points $X_{ij}$, together with a simple model attached to it, generally a mean of the small number of points (constant model).
4.2 Assessment method
The modelling techniques used were described in the previous section. In addition, assessment
metrics are required, as well as a general assessment structure in order to reduce overfitting and
have good generalization power.

4.2.1 Measurements used: MAE, MAPE and MASE
Choosing evaluation metrics is an important task for regression. Here below, the chosen metrics are
exposed and compared to other alternatives, using recommendations found in (Hyndman & Koehler,
2005). Within the regression paradigm, metrics are exposing the error that can be expected from the
model. This error $e_t$ is simply calculated by the difference between the predicted value by the model
$F_t$ and the real value at time $t$, $Y_t$: $e_t = Y_t - F_t$.

- **Mean Absolute Error**: The first metric is the Mean Absolute Error, classified as a scale-
dependent measure, as the Root Mean Square Error, the Mean Square Error or the Median
Absolute Error. The MAE is chosen in this application since it gives a real not biased view of
"what is the average error can be expected from the model". $MAE = mean(|e_t|)$

- **Mean Absolute Percentage Error**: The MAPE is one of the percentage based measurement
derived from the scale-dependent measures. It is used to give a global performance view of
the metric. MAPE is calculated by the following formula: $MAPE = mean\left(\frac{100*|e_t|}{Y_t}\right)$. This
metric as the disadvantage of being sensitive to the scale. Indeed, an error $e_1 = 2$ on a value
$Y_1=5$ would results in a global error of 40%; while the same error using $Y_2=500$ results in a
global error of 0.4%.

- **Mean Absolute Scaled Error**: The MASE is a metric proposed by (Hyndman & Koehler, 2005).
It takes advantage of MAPE drawbacks by comparing $e_t$ with an in-sample one-step naïve
forecast calculated as follow: $q_t = mean(|q_t|)$ ; Where $q_t$ is the scaled error calculated by:
$q_t = \frac{1}{n-1}\sum_{i=2}^{n}|F_{i-1} - Y_{i-1}|$. This measure is answering to the question “how much the method
performs in average against a one-step naïve forecast computed in-sample”. The model is
said better than this naïve prediction when $MASE < 1$.

4.2.2 Leave-one-out Cross Validation: enhance generalization power
As (Guyon & Elisseeff, 2006) say, “The danger of “overfitting” is to find features that “explain well”
the training data, but have no real relevance or no predictive power”. In order to reduce overfitting
and have good generalization capabilities, leave-one-out Cross Validation is used. Each sample is
taken separately and the error $e_t$ is calculated based on what the model built with all the other
observations is predicting for this particular sample. This particular Cross Validation setting is feasible
since this study as a small dataset. In K-fold Cross Validation, the calculated metrics depends on the
random split executed, while this version does not have this problem (Hamel, 2016). Plus, it will allow
us to see precisely how much predictive power has the model for every ETRS or QUEST value. As
instance, the central values are expected to be easier to predict than the values present at the tail.
Chapter 5: Results

This chapter will firstly introduce description and comparison between the base line model obtained from literature and the improved model.

5.1 Base line model and brought improvements

a) The input variable matrix is composed of:

- 40 observations or lines: 2 per patient, each hand side tested being an observation and having 20 patients. The 2 samples per patients are later averaged to obtain the final results
- 10*X columns or features: X being the number of feature (Dominant frequency, signal RMS, etc.) being extracted. 10 inputs are extracted per patient side since 5 tests and 2 parameters exist, corresponding to 10 signals

b) Feature extraction methods

Time domain feature (signal RMS and signal Period) are extracted directly on the time signal pre-processed. Two different frequency domain tools were used for the extraction of Dominant Frequency and Dominant Magnitude. The first one is the Welch methods with Hamming windows length of 2.5 seconds and 50% overlapping to extract frequency domain features. Hamming function was selected as previous close setting studies did (Appendix 3). This window type is well fitted for tremor waves (Grimaldi & Manto, 2010) due to its shape, and was empirically tested as the best. The second frequency domain extractor was the wavelet. In general used for time-frequency features, it can be used by summing up the time scale over magnitudes. A time-frequency feature, the power growth, is extracted by wavelet analysis. Daubechies 8 wavelet with 6 levels is used. This waveform is used after important testing by (Ai, 2008) on tremor. Empirical tests were also performed in this study and confirmed the very good performances of this waveform. In Table 5 second column, the choice between Welch and Wavelet for frequency method extractor is depicted.

c) Features extracted

Table 5 third column, shows the different features extracted. Each combination of tests (rest, postural 1, etc.) and parameters (acceleration or rotation speed) corresponds to a signal pre-processed. For each signal, 3 or 5 features are extracted. The 3 features are Dominant Frequency, Dominant Magnitude and SignalRMS, that were already existent in the literature. Signal Period and Power Growth are added when 5 features are mentioned in Table 5.

d) Test 0 and feature selection

Test 0 is the base line model. It uses the Welch method as most previous close setting studies did (see section 1.2.1) and their most used and compatible to this study features. It is the only model presented in this report that do not use any feature selection algorithm. Next, all the others use Sequential Forward Selection, in order to directly compare the predictive power of the tools and features used.
### e) Results and interpretation

<table>
<thead>
<tr>
<th>Test n°</th>
<th>Frequency method</th>
<th>Features extracted from each signal</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Welch</td>
<td>Dominant Frequency Dominant Magnitude Signal RMS no Feature Selection</td>
<td>10.23</td>
<td>34.91</td>
<td>0.64</td>
<td>18.96</td>
<td>119.18</td>
<td>1.04</td>
</tr>
<tr>
<td>1</td>
<td>Welch</td>
<td>Dominant Frequency Dominant Magnitude Signal RMS</td>
<td>7.53</td>
<td>27.58</td>
<td>0.47</td>
<td>9.43</td>
<td>80.42</td>
<td>0.52</td>
</tr>
<tr>
<td>2</td>
<td>Welch</td>
<td>Dominant Frequency Dominant Magnitude Signal RMS Signal Period Power Growth</td>
<td>7.51</td>
<td>27.14</td>
<td>0.47</td>
<td>8.65</td>
<td>44.25</td>
<td>0.47</td>
</tr>
<tr>
<td>3</td>
<td>Wavelet</td>
<td>Dominant Frequency Dominant Magnitude Signal RMS</td>
<td>7.01</td>
<td>26.26</td>
<td>0.44</td>
<td>9.67</td>
<td>51.12</td>
<td>0.53</td>
</tr>
<tr>
<td>4</td>
<td>Wavelet</td>
<td>Dominant Frequency Dominant Magnitude Signal RMS Signal Period Power Growth</td>
<td>6.41</td>
<td>22.24</td>
<td>0.39</td>
<td>9.01</td>
<td>42.47</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Table 5: Base line model (test 1) and improved models comparisons

Table 5 shows 5 different tests. Comparing feature selection usage between test 0 and 1 shows that it brings much value. Indeed, it can be seen that the MAPE and MASE QUEST shows test 0 not to be useful, i.e. it does not predict the QUEST output better than a random model. Test 1 is a better basis to compare improvement brought by the frequency extraction method and by Signal Period and Power Growth. Test 1 shows acceptable results for ETRS, with 27.58 in Mean Absolute Percentage Error and a MASE clearly below 1, showing the good predictability power. In the other hand, QUEST MASE shows to be better than a one-step naïve forecast computed in-sample (0.52), but the high MAPE value (80.42) highlights that the model precision is bad and not reliable.

Using test 2, ETRS MAPE (27.14%) is slightly enhanced by the use of Signal Period and Power Growth in addition to the others more classical features. The difference is really important when it comes to QUEST indicators, showing that an acceptable predictability can be approached (44.25 MAPE) using more diverse features.

Test 3 shows that the use of wavelet as frequency extraction method is also bringing value. This time, more added values are extracted for ETRS (26.26 MAPE) than for the QUEST (51.12 MAPE). Indeed, the wavelet frequency extraction seems not to be well suited to QUEST score prediction.

Wavelet for frequency extraction, and new features combinations (test 4) is the best obtained model for this study. A 22.24 MAPE is reached for ETRS and 42.47 concerning QUEST. Also, MASE values are clearly under 1 and show the model to be better than a one-step naïve forecast computed in-sample. Then, ETRS shows to be more predictable using this model than QUEST, which is better together with test 2.
QUEST real score values are in general smaller than ETRS scores, implying more sensitivity when it comes to the MAPE indicator as explained in the previous chapter. But even looking at the Mean Absolute Error, it has to be highlighted that ETRS has a larger scale but does predict in average at 6.41 ETRS points the real results, when QUEST is predicted in average at 9.01 QUEST points. The Mean Absolute Scaled Error is the best measure to show this scale difference in predictability results. This fact is probably mainly explained because QUEST score is a more subjective measure based on patient feeling, when ETRS is rated by trained clinician observations. QUEST score predictive power is quite similar for test 2 and 4. Test 2 is chosen as the reference for the next analyses since the MASE score is the less range error repartition dependent. Indeed, some MAPE values might not look consistent with the others. Again, it is because of the small scale and explained by the range error repartition, e.g. if a big prediction error is made on a small real value, the MAPE is more affected than if this prediction error is made on a big real value.

b) Error repartition over the range and side merging

The best models obtained, test 4 for ETRS and test 2 for QUEST are here explored in order to understand where in the ETRS and QUEST range one can expect good accuracy. This purpose will be reached thanks to Figure 28 and Figure 29 below.

Figure 28: ETRS score error repartition over the range for “test n°4”
c) Feature selection

<table>
<thead>
<tr>
<th>Variables kept for ETRS prediction (test 4)</th>
<th>Variables kept for QUEST prediction (test 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>'rest_acc_powerGrowth'</td>
<td>'postural1_rotSpeed_dominantFrequency'</td>
</tr>
<tr>
<td>'postural1_acc_dominantMagnitude'</td>
<td>'postural2_acc_powerGrowth'</td>
</tr>
<tr>
<td>'postural1_rotSpeed_signalPeriod'</td>
<td></td>
</tr>
<tr>
<td>'postural2_acc_dominantFrequency'</td>
<td></td>
</tr>
<tr>
<td>'postural2_rotSpeed_dominantMagnitude'</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Variables kept for best obtained the ETRS and QUEST prediction model

Table 6 shows that the Sequential Forward selection algorithm used did keep 5 variables for ETRS and only 2 concerning QUEST. The variables are kept by importance order. ETRS and QUEST variables kept highlight that postural tests are the most represented variables extracted. This is easily explained by the literature that presents the Essential Tremor shakings to be mainly about postural phase of movement (Elias & Shah, 2014), except when it comes to severe tremor. Knowing that none of the patients in this data set is in a severe case (section 2.1.1), this assumption completely makes sense. Each test predictive power will be further investigated in the next section on this report.

Rotation speed and linear acceleration are really used in combination by the feature selector, showing the need for both linear displacement and joint rotation measures (Grimaldi & Manto, 2010). Except signal RMS, all features extracted are represented in the best variable selection. Each variable predictive power will be studied in the next section of this report. It is interesting to see that power growth, the only time-frequency feature extracted is clearly a very important feature for both
ETRS and QUEST score prediction. The important information from Table 6 is the number of variables. Having 20 patients, 5 variables seems for ETRS seems to be reasonable. But for a better generalization power, more variables might be added to QUEST prediction in the future. Indeed, it is possible to ask the Sequential Forward Selection to result in the best feature subset having a fixed feature number. Though, having a larger scale in QUEST scores would probably naturally lead to a larger number of variables extracted.

5.2 Separate analyses
In this section, the various tests, parameters, features and sides are explored one by one using test n°4 model, since it is the best one for predicting ETRS and QUEST in overall.

a) Tests

<table>
<thead>
<tr>
<th>Test n°</th>
<th>Test</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
<th>MAE</th>
<th>MAPE</th>
<th>MASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Rest</td>
<td>9.11</td>
<td>33.92</td>
<td>0.57</td>
<td>12.87</td>
<td>102.82</td>
<td>0.71</td>
</tr>
<tr>
<td>6</td>
<td>Postural 1</td>
<td>7.2</td>
<td>26.14</td>
<td>0.45</td>
<td>12.99</td>
<td>103.92</td>
<td>0.71</td>
</tr>
<tr>
<td>7</td>
<td>Postural 2</td>
<td>5.97</td>
<td>22.98</td>
<td>0.37</td>
<td>10.09</td>
<td>46.62</td>
<td>0.55</td>
</tr>
<tr>
<td>8</td>
<td>Glass</td>
<td>8.96</td>
<td>34.13</td>
<td>0.56</td>
<td>12.12</td>
<td>65.88</td>
<td>0.66</td>
</tr>
<tr>
<td>9</td>
<td>Finger nose</td>
<td>8.90</td>
<td>35.08</td>
<td>0.55</td>
<td>11.14</td>
<td>108.72</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 7: Testing each test separately results table

Table 7 present situations where only one test type or phase of movement is taken into account. As expected, postural tests have more predictive power than others. Postural 2 and its test n°7 is almost an alternative by itself to test n°4. Though, the data actual representability is not very good in this study, since patients ETRS scores are from 10 to 68, when it can go from 0 to 144. Indeed, no severe cases are presented, and these severe cases are where other than postural tests are supposed to be more relevant (Elias & Shah, 2014) and its combinations with postural tests should provide the best set of information. It is interesting to see that this test n°7 shows a better MAE and MASE. Even if the Sequential Forward Selection is taking the best set of feature, it has to be said again that the final test average both left and right hand predicted ETRS (and QUEST) scores. However, these inconstancies are really marginal.


**b) Parameters**

<table>
<thead>
<tr>
<th>Test n°</th>
<th>Parameter</th>
<th>MAE ETRS</th>
<th>MAPE ETRS</th>
<th>MASE ETRS</th>
<th>MAE QUEST</th>
<th>MAPE QUEST</th>
<th>MASE QUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Linear acceleration</td>
<td>7.27</td>
<td>27.59</td>
<td>0.45</td>
<td>10.09</td>
<td>46.62</td>
<td>0.55</td>
</tr>
<tr>
<td>11</td>
<td>Rotation speed</td>
<td>9.11</td>
<td>33.92</td>
<td>0.57</td>
<td>10.98</td>
<td>62.87</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 8: Testing each parameter separately results table

Table 8 shows linear acceleration as a better predictor than rotation speed for both outputs, even if they could in theory be used separately. As shown by Table 5, parameter combination is recommended and provides better results, as it represents linear displacement and joint rotation, giving a more complete description of tremor.

**c) Features**

<table>
<thead>
<tr>
<th>Test n°</th>
<th>Feature</th>
<th>MAE ETRS</th>
<th>MAPE ETRS</th>
<th>MASE ETRS</th>
<th>MAE QUEST</th>
<th>MAPE QUEST</th>
<th>MASE QUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Dominant frequency</td>
<td>8.72</td>
<td>33.66</td>
<td>0.54</td>
<td>9.67</td>
<td>51.12</td>
<td>0.53</td>
</tr>
<tr>
<td>13</td>
<td>Dominant magnitude</td>
<td>7.23</td>
<td>27.24</td>
<td>0.45</td>
<td>12.07</td>
<td>69.72</td>
<td>0.66</td>
</tr>
<tr>
<td>14</td>
<td>Signal RMS</td>
<td>6.77</td>
<td>26.54</td>
<td>0.42</td>
<td>12.60</td>
<td>96.49</td>
<td>0.69</td>
</tr>
<tr>
<td>15</td>
<td>Signal Period</td>
<td>7.79</td>
<td>30.75</td>
<td>0.49</td>
<td>8.77</td>
<td>59.40</td>
<td>0.48</td>
</tr>
<tr>
<td>16</td>
<td>Power growth</td>
<td>9.92</td>
<td>34.85</td>
<td>0.62</td>
<td>9.79</td>
<td>85.68</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 9: Testing each feature separately results table

Table 9 shows interesting insights about feature combinations. Signal RMS taken alone is the best feature to extract value, but is not present in the best variables kept by the Sequential Forward Selection. Also, it is probable that Signal RMS has some mutual information with Dominant Magnitude. Again, the Power Growth is globally the worst feature for ETRS but does provide more information using specific test or parameter (see Table 6) and with combination with other features. These combinations characteristics are hard to predict in the actual literature state, giving much interest for feature selection algorithm, which allows more insights and better results.

**d) Sides**

<table>
<thead>
<tr>
<th>Side</th>
<th>MAE ETRS</th>
<th>MAPE ETRS</th>
<th>MASE ETRS</th>
<th>MAE QUEST</th>
<th>MAPE QUEST</th>
<th>MASE QUEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>6.84</td>
<td>25.27</td>
<td>0.42</td>
<td>14.24</td>
<td>107.16</td>
<td>0.78</td>
</tr>
<tr>
<td>right</td>
<td>8.99</td>
<td>37.88</td>
<td>0.56</td>
<td>10.46</td>
<td>105.23</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Table 10: Testing each side separately results table

Table 10 shows that using only one side (or sample) for each patient is not the best setting when it comes to results stability. For this reason, repeatability tests and increased datasets and sample per patient is strongly recommend in future studies.
Chapter 6: Conclusion, limitations and future developments

6.1 Main conclusions
Tremor wearable data is completely different from most clinical data (Kubota, Jason, Chen, & Little, 2016), with a patient instantaneous status picture, following one or several predefined test. This paper shows an alternative method for dealing with tremor severity signal measurements on 20 Essential Tremor patients. It is the first time ETRS and QUEST total scores are regressed from accelerometer and gyroscopic measures using wrist-worn smartphone in order to help the research around decision support for Deep Brain Stimulation. Wavelet Time-frequency domain tool and a wrapper feature selection for optimal accuracy. Leave-one-out cross-validated regression tree was used as previous study shown its accuracy in other circumstances, in order to measure new tools improvements. Both outputs were regressed with good accuracy, though better for ETRS. This study provides a positive answer to data mining goals, gives both new insights and new directions (see next section) to follow for the researchers and shows that ETRS and QUEST are interesting dependent variables to predict in order to later improve the process measurement objectivity and performances.

6.2 Limitations and future developments

a) Technological considerations and developments

ETRS and QUEST appear to be good dependent variable. In the future, sensor accuracy improvements will bring even more precision to the measures. In fact, the actual residual noise after cleaning and filtering is still a constraint in order to be very precise in such a wide scale (0-144), which is a much more complicated task than classifying between 5 classes in case of Fahn score (0-4). Depending on why the score was regressed, Fahn score might be enough to take a decision and is anyway an interesting indicator for researchers. But the sensor and technical improvements years after years will let the smartphone to be used as more and more cheap, reliable and precise tools for tremor quantification.

In order to assess the sensor advancement, evaluating more in depth the repeatability at the same level of tremor intensity would be interesting. Indeed, results as shown that even on the same patient at a very close time interval, the patient can provide quite different feature extracted using the left or the right hand. Another factor to take into account when it comes to sensors is the fact that “material fatigue and aging under long-term repeated cycling load may cause potential device failure, which in turn decreases the device reliability” (Grimaldi & Manto, 2010).

Power growth measured in this study shows that some signals have a power growth under 0 before normalization. It means that the signal power is more important in the first half of the signal than the second half, which is not supposed to happen. Indeed, tremor are described in literature to grow over time. This is surprising when most measures are approximately a minute. It shows that noise is still important, even after filtering. Of course, some residual noise is also spread in the frequency bands. In this study, Savitzky–Golay and median noise filters were tested without success. More advanced noise reductor might be interesting to test. It is important to note that removing manually
or automatically by detection in time a peak is not recommended, since it does modify and disturb the frequency distribution.

b) Feature extraction

Feature extraction is an on-going research. Indeed, all the technological improvements cited before will raise the feature extraction accuracy, and could even make some nowadays useless predictors very powerful. An interesting track to follow could be the 3 dimensional techniques. All smartphone sensors are tri-dimensional and mechanical literature might be a plus for this literature. In this direction, (Sprdlík, Hurak, Hoskovcová, Ulmanová, & Ruzicka, 2010) obtained comparable results to other studies by decomposing acceleration into gravity and inertial acceleration for detecting if patients have ET or not, based on quaternion representation of kinematics. Also, other time-frequency features and tools exist and may provide useful information for the signal. However, the signal noisy character for now indicates that sensors improvements will be more and more useful since the accuracy will be enhanced over time.

(Woods, 2012) paper shows the level of attention and distraction to be a good input feature for classifying a patient between PD and ET using Wavelet feature extraction on a smartphone based mobile medical application. This kind of feature could also be investigated and added to the actual test design and might offer new improvement perspectives for predicting tremor severity.

c) Modelization

In this study, regression trees were taken as modelization method. The reasons are that it is a simple and presented as the best by (Darnall, et al., 2012) in predicting tremor severity by feature extraction compared to 5 others modelling techniques. However, this comparative study used classification and not regression as this present study do so. Then, even if Support Vector Machines were tried and shown to be clearly less accurate than regression trees, other techniques might be interesting to use.

d) Big data and dataset representativeness

Of course a larger dataset would increase the accuracy. In this study, the dataset is small but most of all, it’s representativeness could be better. While this study goal is to provide a tremor severity assessment, patients tested do no have what is called a “severe” tremor. Indeed, the maximum ETRS score being 144, when the highest patient’s ETRS in this study is only 68. Then, approximately half of the complete scale is not represented. This is the major limitation concerning this study. The same device, or at least the same sensor has to be used for the measurements.

e) Regarding Parkinson’s disease

This study presented a flexible model that could fit for Parkinson’s disease patients. Indeed, no major changes has to be done for switching the type of patient. Collecting a new dataset, a new and separate model can easily be created for Parkinson’s disease patients. Indeed, the only thing would be changing the filter design by modifying ET frequency bands (7 to 12Hz) to PD frequency bands (4-6Hz) (Elias & Shah, 2014). Though, if limits were tested for all the tests in this case, it might be interesting to test other limits follow (Elias & Shah, 2014) guidelines in case of PD. Indeed, rest
tremor is said to be typically in the 3-6 Hz, postural from 4 to 12 and kinetic from 2 to 7 (Grimaldi & Manto, 2010).

6.3 Managerial implications
The Matlab code used will be shared. Together with the free and open source TREMOR12 application, anyone with a matlab license can do some new measures and advance the research, using this paper as reference point. The previous section discusses few guidelines to orient the research.

At Maastricht University Medical Center the piece of code can be used either for research purpose as discussed in the previous section and for predictions. Based on the improved models provided (see Chapter 5), any new patient ETRS and QUEST scores can be predicted. Also, patients can be quickly evaluated using only one test on both right and left sides. Then, the measure takes only 2 minutes using as instance the postural 2 test (best results using only one test) and the predictions few seconds. But again, the ETRS and QUEST patient scores range used in this paper is rather limited. Indeed, mainly small tremored patients can be tested based on these models. For more precisions in general and on patients with strong Essential Tremor, it is best to first add new data to the actual dataset, as discussed in the previous section.

After enhancing data representativeness and new assessment using the provided code, the code will be more robust and can be used as part of the Deep Brain Stimulation decision making process. Indeed, the assessment performed in this paper revealed good accuracy. The research going more and more advanced in allows to think that the measurements objectivity and performance will be enhanced soon.
Bibliography


Chan, S., Carson, Pun, K., Yeung, K., & Ho, K. (2002). On the design and implementation of FIR and IIR digital filters with variable frequency characteristics. Retrieved from The HKU Scholars Hub - The University of Hong Kong: http://hdl.handle.net/10722/42940


Popovic, L., Sekara, T., & Popovic, M. (2010). Adaptive band-pass filter (ABPF) for tremor extraction from inertial sensor data. *Elsevier*.


Appendix

Appendix 1: ETRS form .......................................................................................................................... 69
Appendix 2: QUEST form ....................................................................................................................... 74
Appendix 3: Close settings studies summary ........................................................................................ 75
Appendix 1: ETRS form

**ESSENTIAL TREMOR RATING SCALE**

- REST: at REST (in repose). For head and trunk, when lying down
- POST: with POSTure holding
  - UE: arms outstretched, wrists mildly extended, fingers spread apart
  - LE: legs flexed at hips and knees; foot dorsal-flexed
  - TONGUE: when protruded
  - HEAD & TRUNK: when sitting or standing
With ACTION and INTention:
  - UE: finger to nose and other actions
  - LE: toe to finger in flexed posture
Except for tongue tremor and voice tremor the following scoring applies.
Grade 0 = none
Grade 1 = slight = amplitude < 0.5 cm. May be intermittent
Grade 2 = moderate = amplitude 0.6-1 cm. May be intermittent
Grade 3 = marked = amplitude 1-2 cm
Grade 4 = severe = amplitude > 2 cm

<table>
<thead>
<tr>
<th>TREMOR SITE</th>
<th>GRADE 0</th>
<th>GRADE 1</th>
<th>GRADE 2</th>
<th>GRADE 3</th>
<th>GRADE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Face</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Tongue</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Voice</td>
<td>ACT/INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Head</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Right upper extremity</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACT/INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Left upper extremity</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACT/INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Trunk</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Right lower extremity</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACT/INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Left lower extremity</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACT/INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stimulation: X on, + off
10. Handwriting. Have patient write the sentence: "This is a sample of my best handwriting". Sign his or her name and write the date.

0 = normal
1 = mildly abnormal. Slightly untidy, tremulous
2 = moderately abnormal. Legible, but with considerable tremor
3 = marked abnormal. Illegible
4 = severely abnormal. Unable to keep pencil/pen on paper without holding hand down with other hand

Ask the patient to join both points of the various drawings without crossing the lines. Test each hand, beginning with the lesser involved hand, without leaning the hand or arm on the table.

<table>
<thead>
<tr>
<th>DRAWING</th>
<th>GRADE 0</th>
<th>GRADE 1</th>
<th>GRADE 2</th>
<th>GRADE 3</th>
<th>GRADE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Fig A LEFT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>RIGHT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>12. Fig B LEFT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>RIGHT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>13. Fig C LEFT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>RIGHT HAND</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Grade 0 = normal
Grade 1 = slightly tremulous. May cross lines occasionally
Grade 2 = moderately tremulous or crosses lines frequently
Grade 3 = accomplishes the task with great difficulty. Many errors
Grade 4 = unable to complete drawing

14. Pouring. Use firm plastic cups, about 8 cm tall, filled with water to 1 cm from top. Ask patient to pour water from one cup to another. Test each hand separately.

0 = normal
1 = more careful than a person without tremor, but no water is spilled
2 = spills a small amount of water (up to 10% of total amount)
3 = spills a considerable amount of water (> 10-50%)
4 = unable to pour water without spilling most of the water.

15. Speaking. This includes spastic dysphonia if present.

0 = normal
1 = mild voice tremulousness when 'nervous' only
2 = mild voice tremor, constant
3 = moderate voice tremor
4 = severe voice tremor. Some words difficult to understand

16. Feeding other than liquids.

0 = normal
1 = mildly abnormal. Can bring all solids to mouth, spilling only rarely
2 = moderately abnormal. Frequent spills of peas and similar foods. May bring head halfway to meet food
3 = markedly abnormal. Unable to cut or use 2 hands to feed
4 = severely abnormal. Needs help to feed


17. Bringing liquids to mouth.
   0 = normal
   1 = mildly abnormal. Can still use a spoon, but not if it is completely full
   2 = moderately abnormal. Unable to use spoon; uses cup or glass
   3 = markedly abnormal. Can drink from cup or glass, but needs two hands
   4 = severely abnormal. Must use a straw

   0 = normal
   1 = mildly abnormal. Able to do everything, but is more careful than the average person
   2 = moderately abnormal. Able to do everything, but with errors; uses electric razor because of tremor
   3 = markedly abnormal. Unable to do most fine tasks, such as putting on lipstick or shaving
       (even with electric razor), unless using two hands
   4 = severely abnormal. Unable to do any fine-movement tasks

19. Dressing.
   0 = normal
   1 = mildly abnormal. Able to do everything, but is more careful than the average person
   2 = moderately abnormal. Able to do everything, but with errors
   3 = markedly abnormal. Needs some assistance with buttoning or other activities, such as tying shoelaces
   4 = severely abnormal, requires assistance even for gross activities

20. Writing.
   0 = normal
   1 = mildly abnormal. Legible. Continues to write letters
   2 = moderately abnormal. Legible, but no longer writes letters
   3 = markedly abnormal. Illegible
   4 = severely abnormal. Unable to sign checks or other documents requiring a signature

   0 = tremor does not interfere with job
   1 = able to work, but needs to be more careful than the average person
   2 = able to do everything, but with errors. Poorer than usual performance because of tremor
   3 = unable to do regular job. May have changed to a different job because of tremor. Tremor
       limits housework, such as ironing
   4 = unable to do any outside job; housework very limited
   5 = not applicable

Stimulate: on off
Score [ ] [ ] [ ]
HANDWRITING: This is a sample of my best handwriting.

Signature: ________________________________________

Date: ____________________________________________

Drawings A, B, and C are made with the ______ left hand
    ______ right hand

DRAWING A       DRAWING B

DRAWING C
NON-DOMINANT HAND

Handwriting: This is a sample of my best handwriting

Signature:

Date:

Drawings A, B, and C are made with the ______ left hand
______ right hand

DRAWING A

DRAWING B

DRAWING C
Appendix 2: QUEST form

**QUEST Scoring**

If a question is Not Applicable, "X" through NA and leave blank—do not assign a score of 0.

<table>
<thead>
<tr>
<th>Question</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My tremor interferes with my ability to communicate with others.</td>
<td></td>
</tr>
<tr>
<td>2. My tremor interferes with my ability to maintain conversations with others.</td>
<td></td>
</tr>
<tr>
<td>3. It is difficult for others to understand my speech because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>4. My tremor interferes with my job or profession.</td>
<td></td>
</tr>
<tr>
<td>5. I have had to change jobs because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>6. I had to retire or take early retirement because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>7. I am only working part time because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>8. I have had to use special aids or accommodations in order to continue my job due to my tremor.</td>
<td></td>
</tr>
<tr>
<td>9. My tremor has led to financial problems or concerns.</td>
<td></td>
</tr>
<tr>
<td>10. I have lost interest in my hobbies because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>11. I have quit some of my hobbies because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>12. I have had to change or develop new hobbies because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>13. My tremor interferes with my ability to write (for example, writing letters, completing forms).</td>
<td></td>
</tr>
<tr>
<td>14. My tremor interferes with my ability to use a typewriter or computer.</td>
<td></td>
</tr>
<tr>
<td>15. My tremor interferes with my ability to use the telephone (for example, dialing, holding the phone).</td>
<td></td>
</tr>
<tr>
<td>16. My tremor interferes with my ability to fix small things around the house (for example, change light bulbs, minor plumbing, fixing household appliances, fixing broken items).</td>
<td></td>
</tr>
<tr>
<td>17. My tremor interferes with dressing (for example, buttoning, zipping, tying shoes).</td>
<td></td>
</tr>
<tr>
<td>18. My tremor interferes with brushing or flossing my teeth.</td>
<td></td>
</tr>
<tr>
<td>19. My tremor interferes with eating (for example, bringing food to mouth, spilling).</td>
<td></td>
</tr>
<tr>
<td>20. My tremor interferes with drinking liquids (for example, bringing to mouth, spilling, pouring).</td>
<td></td>
</tr>
<tr>
<td>21. My tremor interferes with reading or holding reading material.</td>
<td></td>
</tr>
<tr>
<td>22. My tremor interferes with my relationships with others (for example, my family, friends, coworkers).</td>
<td></td>
</tr>
<tr>
<td>23. My tremor makes me feel negative about myself.</td>
<td></td>
</tr>
<tr>
<td>24. I am embarrassed about my tremor.</td>
<td></td>
</tr>
<tr>
<td>25. I am depressed because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>26. I feel isolated or lonely because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>27. I worry about the future due to my tremor.</td>
<td></td>
</tr>
<tr>
<td>28. I am nervous or anxious.</td>
<td></td>
</tr>
<tr>
<td>29. I use alcohol more frequently than I would like to because of my tremor.</td>
<td></td>
</tr>
<tr>
<td>30. I have difficulty concentrating because of my tremor.</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix 3: Close settings studies summary

<table>
<thead>
<tr>
<th>#</th>
<th>Ref</th>
<th>Source type</th>
<th>Study setting</th>
<th>Phenomena observed</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patel et al, 2009</td>
<td>IT biomedicine</td>
<td>12 PD patients</td>
<td>Tremor, bradykinesia and dyskinesia</td>
<td>accelerometer (=linear acceleration)</td>
</tr>
<tr>
<td>2</td>
<td>Darnall et al, 2012</td>
<td>Thesis Mechanical and Materials Engineering</td>
<td>3 ET, 5 PD, 2 ET-PD</td>
<td>Tremor</td>
<td>gyroscope (=rotation speed)</td>
</tr>
<tr>
<td>3</td>
<td>Rigas et al, 2012</td>
<td>IT biomedicine</td>
<td>18 PD, 5 controls</td>
<td>Tremor</td>
<td>accelerometer</td>
</tr>
<tr>
<td>4</td>
<td>Quelioz et al, 2015</td>
<td>Neurosurgery journal</td>
<td>8 ET patients</td>
<td>Tremor</td>
<td>accelerometer (smartphone)</td>
</tr>
<tr>
<td>5</td>
<td>Dai et al, 2015</td>
<td>Sensors journal</td>
<td>7 PD patients</td>
<td>Tremor, dyskinesia</td>
<td>accelerometer and gyroscope</td>
</tr>
<tr>
<td>6</td>
<td>Pan et al, 2015</td>
<td>Wearable computing and domotics for health</td>
<td>40 PD patients</td>
<td>Tremor and gait difficulty</td>
<td>accelerometer (smartphone)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Body part tested</th>
<th>Task</th>
<th>Rating scale used / classification</th>
<th>Phase of movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Body sensor network</td>
<td>Regression</td>
<td>UPDRS (0-4)</td>
<td>Kinetic</td>
</tr>
<tr>
<td>2</td>
<td>Wrist</td>
<td>Classification</td>
<td>Fahn Rating Scale (0-4)</td>
<td>Rest, postural and kinetic</td>
</tr>
<tr>
<td>3</td>
<td>Several body segments</td>
<td>Classification</td>
<td>UPDRS (0-4)</td>
<td>Rest and postural</td>
</tr>
<tr>
<td>4</td>
<td>Wrist</td>
<td>Classification</td>
<td>Fahn Rating Scale (0-4)</td>
<td>Postural and kinetic</td>
</tr>
<tr>
<td>5</td>
<td>Wrist</td>
<td>Classification</td>
<td>UPDRS (0-4)</td>
<td>Rest, postural and kinetic</td>
</tr>
<tr>
<td>6</td>
<td>Hand</td>
<td>Regression</td>
<td>Hoehn &amp; Yahr score (1-5)</td>
<td>Rest</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Preparation</th>
<th>Feature methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bandpass filtered 3-8 Hz</td>
<td>time and frequency (5s window)</td>
</tr>
<tr>
<td>2</td>
<td>no info</td>
<td>Fast Fourier Transforms (5s window)</td>
</tr>
<tr>
<td>3</td>
<td>Low-pass finite-impulse-response (FIR) filter cutoff 3Hz, to remove gravity force and intentional movement + band-pass 3to12 Hz</td>
<td>Windowed FFT (3s window + 1.5s overlapping)</td>
</tr>
<tr>
<td>4</td>
<td>kinetic: high-pass filtered (one pole, 2.5-Hz cutoff) for voluntary motion ; Postural: 1-Hz cutoff to filter gravity</td>
<td>simple tools and FFT (Hamming windowed, unknown length)</td>
</tr>
<tr>
<td>5</td>
<td>Removed gravity and dyskinesia with band-pass filter 3.25to12Hz</td>
<td>FFT (windowed? Not said)</td>
</tr>
<tr>
<td>6</td>
<td>Low pass filter</td>
<td>FFT (windowed? Not said)</td>
</tr>
<tr>
<td>#</td>
<td>Feature extraction</td>
<td>Model</td>
</tr>
<tr>
<td>----</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>The range of amplitude of each channel, the root mean square (rms) value of each accelerometer signal, two cross-correlation-based features (i.e., the peak of the normalized cross-correlation function derived from pairs of accelerometer time series and the time lag corresponding to such peak value; the normalization of the cross-correlation function limited its values between −1 and 1), two frequency-based features (i.e., the dominant frequency component and the ratio of the energy associated with the dominant frequency component to the total energy), and the signal entropy</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>2</td>
<td>the peak frequency and magnitude ; the RMS value for each axis</td>
<td>Six were used: Random Forest, Decision Tree, Nearest Neighbor (NN), Bayes, Multilayer Perceptron (MLP), and Support Vector Machine (SVM)</td>
</tr>
<tr>
<td>3</td>
<td>For tremor severity only: dominant frequency, amplitude of the dominant frequency, energy of harmonic motion, spectrum entropy, LF and HF energy, ratio to high to total energy</td>
<td>2 Hidden Markov models</td>
</tr>
<tr>
<td>4</td>
<td>Average and maximum accelerations, time above 1 g of acceleration, peak frequency, typical magnitude of tremor</td>
<td>Nonlinear model (log)</td>
</tr>
<tr>
<td>5</td>
<td>Amplitude of tremor and dominant frequency of tremor</td>
<td>Nonlinear model (log)</td>
</tr>
<tr>
<td>6</td>
<td>Power between 4 and 6 Hz, Fraction of power between 4 and 6 Hz, Power ratio between 3.5-15 Hz and 0.15-3.5 Hz, total power from 0 to 20 Hz, Peak power, average acceleration in time signal</td>
<td>3 regression models</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th>#</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tremor estimated error: 3.4%</td>
</tr>
<tr>
<td>2</td>
<td>82% accuracy (on rating scale, for random forest &amp; decision tree)</td>
</tr>
<tr>
<td>3</td>
<td>87% accuracy (on rating scale)</td>
</tr>
<tr>
<td>4</td>
<td>Mean acceleration 92%/83%, maxi acceleration 89%/71%, tremor magnitude 86%/83%, time above 1 g of acceleration only relevant when tremor rating is 3 or 4/4</td>
</tr>
<tr>
<td>5</td>
<td>98% accuracy (on rating scale)</td>
</tr>
<tr>
<td>6</td>
<td>0.81 correlation coefficient</td>
</tr>
</tbody>
</table>