

Development of an automatic drowsiness monitoring system using the electrocardiogram

Auteur : Bourghelle, Florent

Promoteur(s) : Verly, Jacques

Faculté : Faculté des Sciences appliquées

Diplôme : Master en ingénieur civil électricien, à finalité approfondie

Année académique : 2015-2016

URI/URL : <http://hdl.handle.net/2268.2/1451>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.



University of Liège
Faculty of Applied Sciences
Department of Electrical Engineering and Computer Science

Development of an automatic drowsiness monitoring system using the electrocardiogram

Master thesis submitted for the degree of Master in Electrical Engineering
by **Florent Bourghelle**

Supervised by Prof. **Jacques Verly**

Academic year 2015-2016

Development of an automatic drowsiness monitoring system using the electrocardiogram

Florent Bourghelle
Supervised by Prof. Jacques Verly

Electrical Engineering
University of Liège

Academic year 2015-2016

Abstract

The aim of this thesis consists of the development of an automatic drowsiness monitoring system based on the electrocardiogram (ECG). Moreover, as the feasibility of this physiological signal to detect drowsiness is still not proved, this thesis also investigates its feasibility.

This thesis is based on an experiment where subjects were sleep deprived during 28 hours. At 3 specific moments of sleep deprivation, subjects performed psychomotor vigilance task (PVT). During these tasks, different physiological signals whose electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) were recorded. Based on the EEG and EOG signals, which are the references to assess drowsiness, the true state of each subject is known on the Karolinska Drowsiness Scale and can be defined as awake or drowsy given a defined threshold.

First, this thesis performs a review of the literature to find the possible parameters indicative of drowsiness computed from the ECG. Then, a complete processing chain of the ECG signal is implemented to be able to compute these parameters in the time and statistical domains, the non-linear domain, and finally in the frequency domain from the raw ECG of the subjects. As the respiratory signal can be derived from the ECG (ECG-Derived Respiration signal), this thesis also incorporates parameters from the respiratory domain in order to see if this domain can be used to detect drowsiness.

Once these parameters are computed, a machine learning phase is developed. During this phase, the issue of the variability of the features between the subjects was highlighted. Several techniques to compensate this variability have been tested but none improved the results obtained. This variability makes the system developed to be not reliable enough on all the subjects of the experiment to only use the ECG to predict drowsiness.

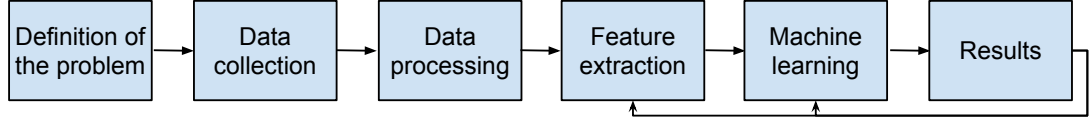


Figure 1: Work flow performed in this thesis.

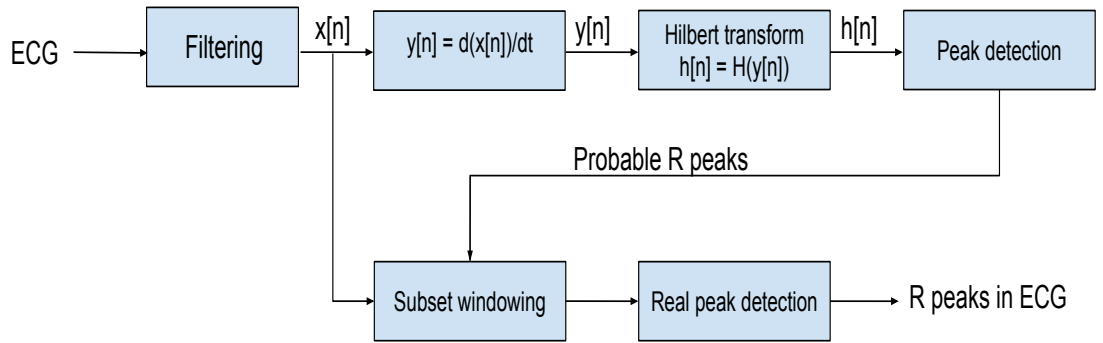


Figure 2: Summary of the algorithm developed for the R peaks detection based on [1].

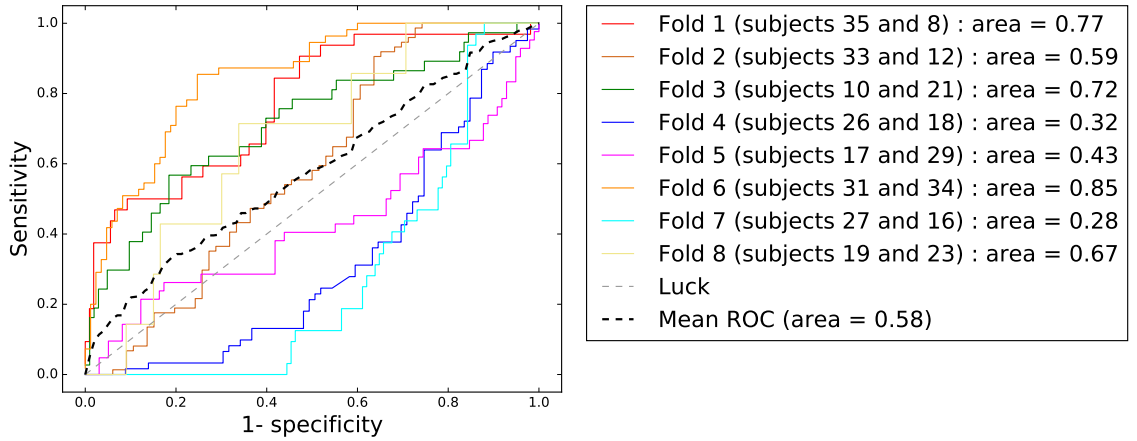


Figure 3: The ROC curve of one on the model developed during this thesis during a cross-validation. Each color corresponds to a fold. Note that observations of the same subject are never split between the different fold. One can observe the issue of the variability between the subjects encountered during this thesis.

Acknowledgements

First of all, I would like to thank Clémentine François for proposing to me to work on this important and interesting subject. I am very grateful for all the supervision, and precious advice she gave to me during all this project. I want to thank her for all the support and motivation she put during this thesis.

I also want to express my thanks to Professor Verly for the time he gave during the proofreading of this thesis.

I am also grateful to Professor Pierre Geurts and Antonio Sutera for helping me during the machine learning part of this thesis.

My thoughts also go to my parents, sister, and friends who always support me during all the progress of this work.

Finally, I thank all those who have contributed, in one way or another, to the research in the drowsiness sector, and who are not mentioned here.

Contents

1	Introduction	1
2	Drowsiness and electrocardiogram (ECG): does there exist a possible link?	4
2.1	Drowsiness	4
2.1.1	Definition	4
2.1.2	How to assess drowsiness ?	5
2.2	The electrocardiogram	6
2.2.1	A few words about the physiology of the heart	6
2.2.2	Action of the nervous system on the heart	7
2.2.3	Lead configuration	8
2.2.4	Heart rate variability (HRV)	10
2.3	State of the art of the use of ECG to detect drowsiness	11
2.4	Flowchart of the work performed in this thesis	13
3	Feature extraction	14
3.1	R peaks detection algorithm	14
3.1.1	Filtering	15
3.1.2	Derivative	21
3.1.3	Hilbert transform	21
3.1.4	First R peak detector and subset windowing	23
3.1.5	Second peak detector	24
3.1.6	Result of the R peaks detection algorithm	24
3.2	Computation of the features	24
3.2.1	Creation of the tachogram	24
3.2.2	Time duration from which the features are computed	27
3.2.3	The HRV features in the different domains	29
3.2.4	The respiratory features	35
4	Machine learning	38
4.1	Introduction and theory about machine learning	38
4.2	The output of the model	39
4.2.1	The KDS score	39
4.2.2	Choice between regression and classification	40
4.3	Machine learning model developed	42
4.3.1	The final dataset	42
4.3.2	The cross-validation	43
4.3.3	Different estimators used to build the model	44
4.3.4	Performance measures	44
4.4	Conclusion of the chapter	46

5	Results obtained and discussion	47
5.1	Balanced and normalized dataset	47
5.2	Discussion of the results obtained for the normalized dataset	53
5.3	Feature selection	54
5.3.1	Use of the coefficient of correlation to select the features	55
5.3.2	The use of the random forest to rank the features	56
5.4	Discussion of the features computed	57
5.5	Discussion about the sleep deprivation of the subjects	58
5.6	Conclusion of the results	58
6	Conclusion	60
	Appendices	62
A	Features as function of the KDS score	63
B	Distribution of the KDS score per subject	67

List of Figures

2.1	Circadian cycle influence on sleep propensity	5
2.2	Link between the heart activity and the ECG pattern.	7
2.3	Cardiac conduction system	7
2.4	Universal 12-lead ECG	9
2.5	Bipolar leads forming the Einthoven's triangle with the heart on the middle	9
2.6	Example of a 12 lead ECG recording	9
2.7	Example of a tachogram computed during Chapter 3.	10
2.8	Flowchart of the work performed in this thesis.	13
3.1	Flowchart of the algorithm developed for the R peaks detection based on [1]	15
3.2	Noise effects occurring in different ECGs used in this project.	16
3.3	Block diagram of an FIR filter	17
3.4	Modulus of the frequency response of the ideal low pass filter	18
3.5	Impulse response of an ideal low pass filter (the cut-off frequency = f_c).	18
3.6	Frequency response and impulse response of the band pass FIR filter implemented in this work. The passband is between 8Hz and 20Hz.	20
3.7	Impact of the FIR filter on the different noise effects of Fig. 3.2. In blue, the original noisy signal. In green, the FIR filtered signal.	22
3.8	Results of the R peaks detector algorithm on the noisy ECGs of Fig. 3.2. The R peaks are shown by green crosses.	25
3.9	Hilbert transform of the three ECGs with noise shown in Fig. 3.2. The Hilbert transform is shown in blue. The threshold computed according to the level of noise ECG in that window is shown in black. The red crosses are peaks of the Hilbert transform above the threshold that are false peaks. The green crosses are the peaks found that are correct true peaks.	26
3.10	Successive steps in the generation of the final tachogram. Top: raw tachogram just after the R peak detection algorithm. Middle: the result when only the NN intervals are kept. Bottom: the outliers are also removed.	28
3.11	Illustration of the principle of the moving window used to compute the features for the machine learning model.	29
3.12	Example of a Poincaré plot	31

3.13	Comparison of the PSD estimation using the Fourier transform and the Lomb method. Only the frequency in the LF and HF parts of the spectrum are represented as the PSD is based on a short term record of 120 seconds. (a): result of using the Fourier transform. Before applying this transform, the signal was re-sampled by using cubic spline interpolation at 8 Hz. (b): result of the Lomb method applied to the unevenly sampled signal.	33
3.14	BB intervals computed from the ECG-Derived Respiration signal for a subject during a PVT test.	37
4.1	Example of a male subject that was removed from the dataset due to an acquisition problem. The ECG is totally corrupted and full of noise. No useful information can be derived from a signal with so much noise.	43
4.2	Example of a ROC curve. The diagonal dashed line corresponds to the random result.	46
5.1	The three ROC curves obtained during a cross-validation. Each model is used with all the features and the optimal parameters of the estimators are used. The same cross-validation is used for the 3 models. The area under the ROC curve is also indicated. The gray diagonal dotted line corresponds to random predictions.	51
5.2	The three ROC curves as in Fig. 5.1 during a cross-validation for another set of folds. Each model is used with all the features and the optimal parameters of the estimators are used. The same cross-validation is used for the 3 models. The area under the ROC curve is also indicated. The gray diagonal dotted line corresponds to random predictions.	52
5.3	Mean ROC curves for the 3 models obtained on all the combinations of the subjects of the experiment. Each color corresponds to a model using a specific estimator. The area under each ROC curve is also indicated. The gray diagonal dotted line corresponds to the random predictions called luck in the legend.	53
5.4	Ranking of the 20 features obtained by using extra trees on a eight folds cross-validation.	57
5.5	Evolution of the ROC curve as a function of the number features selected from the ranking of Fig. 5.4.	57
A.1	Plots of the mean and the standard deviation of each feature of interest for all the subjects of the experiment as a function of the KDS score. The mean is the blue dotted curve. For each value, the standard deviation (std) is represented by the vertical bar. Each vertical bar has a length of two std. This vertical bar gives information about the error on that value of the feature. A long vertical bar means that there are big variations of that feature for the same KDS score between the subjects. One can find the full name of the acronyms used in Tables 3.1, 3.2, 3.3, and 3.4 of Chapter 3.	66
B.1	Distribution of the KDS scores as a function of the subjects of the experiment.	67

List of Tables

3.1	Table summarizing the names, units, definitions, and the formulas of the features in the time and statistical domains.	30
3.2	Table summarizing the names, units, definitions, and the formulas of the features in the non linear domain.	31
3.3	Table summarizing the names, units, definitions, and the formulas of the features in the frequency domain.	32
3.4	Table summarizing the names, units, definitions, and the formulas of the features in the respiratory domain.	37
4.1	Example of the scoring sheet used to assess the level of drowsiness of a subject during a PVT test on the <i>Karolinska drowsiness scale</i>	41
4.2	Example of a confusion matrix used in binary classification.	45
5.1	Results of the sensitivity (Se), specificity (Spe), positive predicted value (PPV), and the negative predicted value (NPV) for the different machine learning models built with the 3 estimators used with their default parameters. The first block, called No balancing, consists of the raw dataset. The second block, called Balancing, consists of the balanced dataset. The third block, called Balancing + Normalization, consists of the balanced and normalized dataset.	48
5.2	Results of the sensitivity (Se), specificity (Spe), positive predicted value (PPV), and the negative predicted value (NPV) between the different machine learning models for the balanced LS. The estimators of the models are used with their best parameters found.	49
5.3	Results of the sensitivity (Se), specificity (Spe), positive predicted value (PPV), and the negative predicted value (NPV) between the different machine learning models on all the possible combinations. The estimators of the models are used with their best parameters found.	50
5.4	Comparison of the Sensitivity (Se), Specificity (Spe), Positive Predicted Value (PPV) and the Negative Predicted Value (NPV) for four cases. The three first cases (z-baseline, sub-baseline, and div-var-baseline) attempts to remove the variability between the subjects. The fourth case (no compensation) is the case with no variability compensation method.	55
5.5	Comparison of the performance measures obtained for the model using the SVM estimator with the linear kernel. In the first line all the features are used. In the second line only the 5 features found with an absolute coefficient of correlation above 0.25 are used. The performance measures are: the sensitivity (Se), specificity (Spe), positive predicted value (PPV) and the negative predicted value (NPV). . . .	56

List of abbreviations

ANS: Autonomic nervous system
BB: Breath to breath
ECG: Electrocardiogram
EDR signal: ECG-Derived Respiration signal
EEG: Electroencephalogram
EOG: Electrooculogram
FIR filter: Finite impulse response filter
HF: High frequency
HRV: Heart rate variability
KDS: Karolinska Drowsiness Scale
KSS: Karolinska Sleepiness Scale
LF: Low frequency
LS: Learning set
MIT: Massachusetts Institute of Technology
NN interval: Normal sinus to normal sinus inter beat interval
PSD: Power spectral density
PVT: Psychomotor vigilance task
ROC: Receiver operating characteristic curve
RSA: Respiratory sinus arrhythmias
STD: Standard deviation
TS: Test set

Chapter 1

Introduction

Drowsiness is an important cause of major injuries or deaths in car accidents throughout the world. For instance, in the European Union, drowsiness at the wheel is estimated to be responsible for 20 to 25% of road accidents [2]. This proportion is similar to the one of car accidents caused by drunk driving (25% according to European SafetyNet study devoted to alcohol [3]). Different studies have shown that driving after 24 hours of sleep deprivation has the same effect as driving with 0.9 g/l of alcohol in the blood (an illegal rate everywhere in the UE) [2].

According to another study of the AAA Foundation for Traffic Safety published in 2014, in the United States, 13% of the car crashes involved people that then were hospitalized and 21% of the fatal car crashes involved a drowsy driver [4]. However, objective data are difficult to find. Indeed, it is very complicated to estimate the drowsy state of the conductor just before the accident. On one hand, there are of course the physical consequences (invalidity or death) of these accidents. But on the other hand, there are also the economical consequences of such accidents. For instance, for the year 2011, the total amount of costs (direct and indirect) of these accidents was estimated at 160 billion Euros in Europe [2].

To overcome this problem, research and technologies to detect drowsiness events in cars are in progress. However, the reliability of these systems is not always proved. Moreover, lots of these systems try to assess the state of the driver based on his/her behavior (i.e. steering wheel movements, car motion on the road, ...). They are thus not based on a monitoring system recording physiological data from the driver.

Currently, through the scientific community, the method that is considered as the reference to monitor drowsiness from physiological data is the polysomnography consisting of the electroencephalography (EEG) and the electrooculography (EOG) signal. Electroencephalography (EEG) monitors the electrical activity of the brain thanks to electrodes placed on the scalp of the patient. Electrooculography (EOG) monitors the activity of the eye. More precisely, there exists a difference of potential between the cornea and the retina with the cornea being taken as the positive side. This difference of potential generates an electric field in the cells around the eye. When the eye moves, this electric field modifies its direction. Therefore, the movement of the eye can be recorded via electrodes placed near the eye [5].

However, the problem with these 2 techniques is that they are quite intrusive for a driver in a car. The work presented in this thesis is therefore focused on the Electrocardiogram (ECG). In the ECG technique, the electrical activity of the heart

is monitored thanks to electrodes. This technique is considered in this thesis because it is less intrusive to monitor the ECG compared to the use of the EEG and/or EOG. For instance, Jung et al. developed a system based on capacitive active electrodes placed on the seat and that detects the ECG of the driver without direct contact [6]. An other example of a non-intrusive technique to monitor the ECG is the AliveCor's Apple Watch ECG. The society AliveCor is improving the Apple smart watch with electrodes on it to be able to monitor the ECG. Up to now they are in the prototype phase. The release in the US is planned during the 2016 year when the product will have been approved. Therefore, the possibility to capture the ECG in a fully non-intrusive way for the driver is really present.

However, the use of the ECG to detect drowsiness events is still a new field in research, and studies have not yet proved completely the feasibility or not of the ECG to detect drowsiness. This is why this work will also be devoted to see whether the development of an automatic system based on the ECG to detect drowsiness is achievable or not.

More precisely, this thesis is based on a study performed in 2014 at the University of Liège. During this study, several subjects performed three psychomotor vigilance tasks at three different degrees of sleep deprivation. During these tests, several physiological signals were monitored including the EEG, EOG and ECG. The analysis of the EEG and the EOG was performed as part of projects that preceded this thesis work.

As the EEG and EOG signals are considered as the standards among the scientific community to monitor drowsiness, the EEG and EOG are used as the reference to determine if a subject is drowsy or awake. Based on the analysis of the ECG, prediction of the state of the subjects of the experiment will be done. These predictions will be compared to the reference known states of the subjects, determined from EEG and EOG, to assess the reliability or not of the ECG to predict drowsiness.

To perform this work, the thesis is divided in four main chapters that represent chronologically all the work performed during it.

Chapter 2 gives an introduction about drowsiness, the physiology of the heart, and the ECG. The chapter also introduces the heart rate variability concept and the use of the tachogram. It is from this heart rate variability that drowsiness will be predicted based on the ECG.

Chapter 3 explains all the processing chain of the ECG implemented in this thesis. Then, the rest of the chapter presents the review of the standard parameters defined among the scientific community that can be computed to assess the heart rate variability. The different domains in which these parameters can be obtained and the way to compute them are detailed all along the chapter.

Chapter 4 gives an introduction to the machine learning technique that will be used to develop the automatic system. In the chapter, the way the reference states of subjects were obtained by analyzing the EEG and EOG will also be explained. Finally, the chapter will detail the machine learning procedure developed during this thesis and the performance measures that will be used to evaluate the monitoring system of drowsiness.

Chapter 5 presents the results obtained when the parameters computed in Chapter 3 are used with the machine learning models presented in Chapter 4. Based on the

results obtained on the experiment this thesis is using, the feasibility or not of the ECG to predict drowsiness will be discussed.

Finally, the last chapter consists of the conclusion of the thesis and the possible perspectives for future works.

The main personal contributions of this thesis are as follows:

- I made the review of the literature to identify the different parameters indicative of drowsiness that can be computed from the ECG.
- I have implemented all the feature extraction part on *Matlab*. For this part, I first needed to identify the R peaks in the ECG. To do so, I implemented the R peaks detection algorithm explained in the paper of Benitez et al. [1]. Then, I computed all the features in their respective domains. As the respiratory signal can be derived from the ECG (ECG-Derived Respiration signal), I decided to add parameters of the respiratory domain to the learning algorithms in this thesis. To do so, I derived the respiratory signal from the ECG using the variations of amplitudes of the R peaks as explained in [7].
- After defining the procedure to follow in order to build and assess the models for the machine learning part, I implemented them in *Python* using the *Scikit-Learn* library.
- I also tried 3 different techniques in order to try to compensate the variability between the subjects.

Chapter 2

Drowsiness and electrocardiogram (ECG): does there exist a possible link?

In this chapter, drowsiness is defined and generalized in the first part. Then, the ECG, the physiology of the heart and the possible configurations of an ECG will be approached. Next, the state of the art of the detection of drowsiness using the ECG will be addressed. Finally, the flowchart of the work performed during this thesis will be presented.

2.1 Drowsiness

2.1.1 Definition

Drowsiness is generally defined as a state between wakefulness and sleep. During drowsiness, the person irresistibly wants to fall asleep if there is no stimulation [8]. Staying awake require some efforts and sleeping is the only solution that can efficiently responds to this need and to durably restore wakefulness. Generally, drowsiness and fatigue are often mixed in the usual language. However, in medical applications, there are not the same. Fatigue corresponds to a moral or physical weakening felt after a sustained effort. The rest (not necessarily sleeping) is the only solution to restore a sufficient level of performance. For instance, nervous fatigue occurs roughly every two hours. The consequences of this fatigue are a diminution of attention and concentration. This phenomenon explains for example why drivers are encouraged to take a rest of 15-20 min after two hours of driving on a rest area [9]. During this resting time, they do not necessarily have to sleep. They just need to stop their driving activity to be more focused on the road when they get back in their cars.

Sleep and wakefulness cycles are regulated by two physiological mechanisms:

- The first one is the circadian cycle. This cycle is the internal clock of our body that maintains our activity according to a 24-hour rhythm. Typically, with the circadian cycle, the sleepiness peaks are in the middle of the afternoon

(2-4 pm) and in the early morning (2-6 am). Figure . 2.1 shows the influence of this cycle on the sleep propensity.

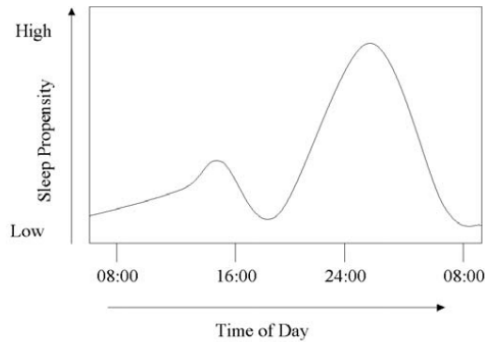


Figure 2.1: Circadian cycle influence on sleep propensity (taken from [10]).

- The second mechanism is the homeostatic regulation of sleep. It is measured through the sleep pressure. The longer the waking period, the higher the sleep pressure. Therefore, this pressure increases all along the day with the mental activity. When we sleep, the sleep pressure decreases accordingly to the sleep time. Through the sleep/wake cycle, there is an homeostatic regulation of the sleep pressure.

Circadian cycle and sleep pressure are thus the two systems regulating the sleep/wake cycle. However, external factors can also influence this cycle like: drugs, alcohol, sleep deprivation, task performed (monotonous or stimulating, unpredictable or not, short or long, ...), the environment (quite or noisy, bright or dark, alone or in crowded environment), etc.

The main consequences of drowsiness are the followings [2]: loss of alertness, difficulty keeping the eyes in focus, loss of concentration and wandering thoughts, reduced awareness of the environment, memory lapses, increase of reactions times.

2.1.2 How to assess drowsiness ?

Drowsiness can be assessed in two different ways. The first one is to assess drowsiness based on subjective tests. In these tests, the subject responds to a series of questions. On the basis of the responses chosen, a drowsiness score can be attributed. However, this technique is far from reliable. The second method to assess drowsiness is the objective methods based on physiological data. In the scientific world, the reference to assess drowsiness is the polysomnography. This method uses the electroencephalogram (EEG) and the electrooculography (EOG). More precisely, specific activity in a given frequency spectrum of the EEG, or slow eye movements of the EOG, are considered to be indicators of drowsiness. Based on these two signals, a drowsiness score can be attributed to the subject. This score is called the KDS (Karolinska Drowsiness Scale) score [11, 12]. The way to obtain this score will be explained in more detail with a specific example in chapter 4 of this thesis. Note that the KDS score, which is an objective method to assess drowsiness, must not be confused with the KSS (Karolinska Sleepiness Scale) score, which has nearly the same name but is a subjective method.

2.2 The electrocardiogram

2.2.1 A few words about the physiology of the heart

The role of the heart is crucial for living beings. Indeed, it is through the pumping movement of the heart that blood circulates through the different vessels of the organism and brings oxygen and nutrients to its different cells.

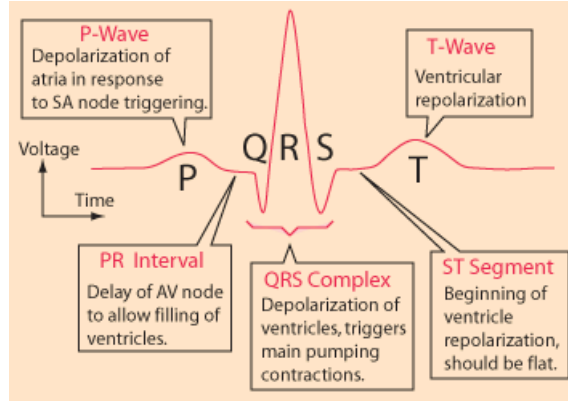
During each beat of the heart, the depolarization and repolarization of the myocardium (layer of the heart made of contractile cells) generate electrical activity [13]. Therefore, thanks to two electrodes placed on the skin of the body, we can detect a variation of potential between these two electrodes during the activity of the heart. This variation of potential through the time is known as the electrocardiogram (ECG). The global activity of the heart is recorded in an ECG.

Each heart beat corresponds to a specific pattern in the ECG. This pattern comprises five waves (P, Q, R, S, and T) as it can be seen in Fig. 2.2. The first wave is the P wave. It corresponds to the atrial depolarization initiated by the sinoatrial node (SA node). It is during this phase that the contraction of the auricles occurs (= atrial systole). Then it is the QRS complex consisting of the ventricular depolarization just before its contraction for the ventricular systole. During this complex, there is also the atrial repolarization (= atrial diastole). Finally, the last wave is the T wave corresponding to the ventricular repolarization leading to the ventricular diastole.

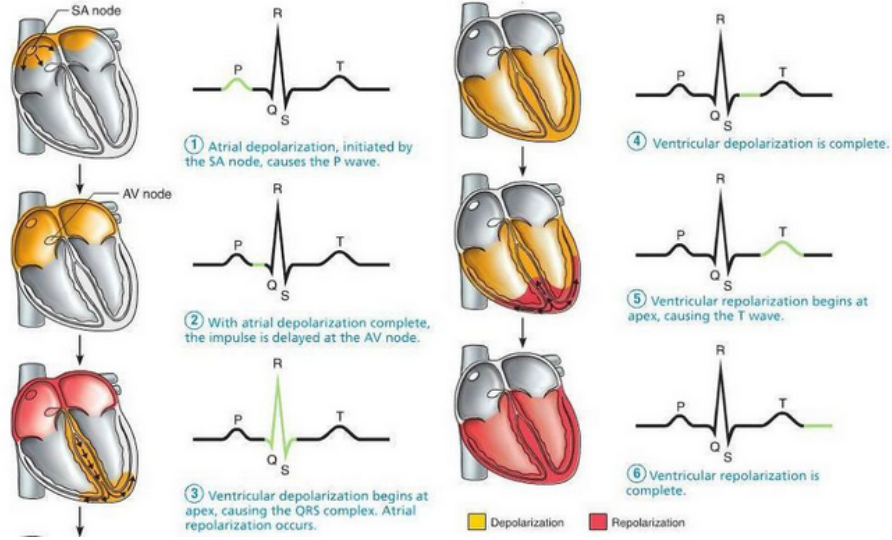
The heart has the ability to autonomously generate depolarization at a specific frequency. This phenomenon can be seen for instance during heart transplant. The heart continues to beat when it is extracted from the body of the donor. This ability comes from the pacemaker cells present in the heart. These cells slowly depolarize themselves from their resting potential. After a certain time, the potential of the membrane will reach the threshold potential. Therefore, an action potential will be generated resulting in a heart beat. There are 3 main parts for the pacemakers cells [16] as it can be seen in Fig. 2.3 :

- Sinotrial node: it is situated at the junction between the superior vena cava and the right atrium (Fig. 2.3). This node has frequency of 60 to 100 beats per minute. Action potential is transferred from this node in the atrium until the atrioventricular node.
- Atrioventricular node: this node is situated near the valve of the right atrium (Fig. 2.3). The propagation signal is delayed during a small time at this node to allow the full contraction of the atria. Then, it is transmitted through the ventricles.
- His bundle: this bundle is a link between the atrium and the ventricle (Fig. 2.3). This bundle is thus splitted in two parts to reach the two ventricles. The propagation goes down through this bundle until reaching the Purkinje fibers. It is in the Purkinje fibers that the myocardial cells are excited leading to the ventricular contraction.

In a normal heart, it is the sinotrial node that has the highest depolarization frequency. It is thus this node that determines the heart beat.



(a) ECG wave (taken from [14])



(b) Depolarization wave trough the heart (taken from [15])

Figure 2.2: Link between the heart activity and the ECG pattern.

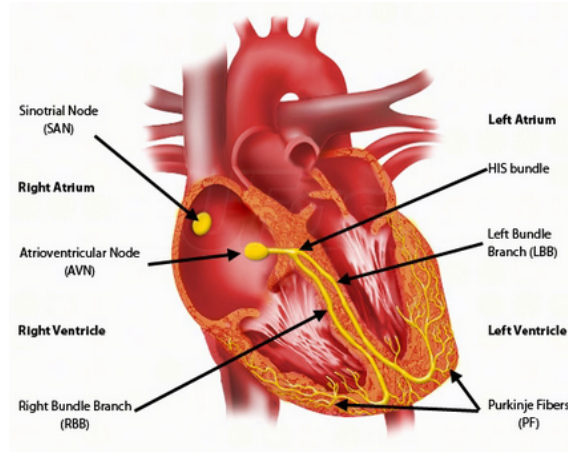


Figure 2.3: Cardiac conduction system (taken from [17]).

2.2.2 Action of the nervous system on the heart

As said before, the heart has the ability to generate heart beat autonomously thanks to pacemaker cells. However, the autonomic nervous system (ANS) is able to interact with the heart to change its activity [18]-[10]. This system is divided in two parts. The first one is called the parasympathetic nervous system and slows down the heart.

This system uses the vagal nerve as the main road. This is why the parasympathetic action is also called the vagal activity. The second system is the sympathetic nervous system. It can speed up the beats frequency of the heart. The sympathetic activity is also able to increase the contractile force of the heart.

The parasympathetic nervous system releases acetylcholine from the terminals of its ganglia whereas it is noradrenalin that is released for the sympathetic system. The acetylcholine released by the vagal activity fix in the M2 muscarinic receptor. This binding triggers the inhibition of the spontaneous depolarization of the sinotrial node (also called the negative chronotropic effect). For the sympathetic system, the noradrenalin attach to beta 1 receptors. The consequence is an increase of the speed of the depolarization of the sinotrial node which increases the heart frequency (positive chronotropic effect). Moreover, the action of the noradrenalin will also lead to a rise of the intracellular calcium concentration thanks to a series of biocehmical reactions. This rise of calcium will increase the contractile force of the heart [18]-[19]. These two antagonist systems trigger thus variations of heart rate. However, the time-scale of their actions is not the same. There is a time delay of 5s after the stimulation for the sympathetic response. This response remains for 20-30s. The parasympathetic reaction is faster (400ms of delay) and is shorter [20].

2.2.3 Lead configuration

The electrodes required in the ECG can be placed at different locations on the body. However, these locations are not anywhere on it. Indeed, if two ECGs of the same patient are compared, any changes detected in the ECGs records must be caused by a heart problem and not simply a change of the positions of the electrodes. Therefore, the location of the electrodes are universal. This leads to the universal 12-lead ECG. The lead term is used to represent the difference of voltage between two electrodes. Each lead measures the electrical activity of the heart along one direction. A 12-lead ECG machine consists of 10 wires [21]:

- 6 chest electrodes placed on the rib cage. There are labeled as V and are numbered from 1 to 6. Fig. 2.4a shows the position of these electrodes.
- 4 limb electrodes placed at the periphery of the body. Two are on the right and left wrists and two are on left and right ankles as it can be observed in Fig. 2.4b. The electrode connected to the right ankle is used as the neutral point. This neutral point allows for instance the so-called *driven right leg configuration* [22]. In this configuration, the common mode is reduced thanks to a negative feedback into the right leg electrode which contributes to a higher reduction of electric interference. Therefore, the driven right leg configuration enhances the ECG signal. The three other electrodes are labeled aV for augmented vector.

We have thus a total of 10 wires , one of which 1 is used as the neutral. This leads to a number of 9 leads called the unipolar leads. The three remaining leads are called the bipolar leads and are placed in a triangle configuration with the heart at the middle (Fig. 2.5).The lead I is between the aVR and the aVL, lead II is between the aVR and the aVF and lead III is between the aVL and the aVF.

At the end we have thus a 12-leads ECG as it can be seen in Fig. 2.6 and each lead conveys information about the heart activity along one direction. However, the

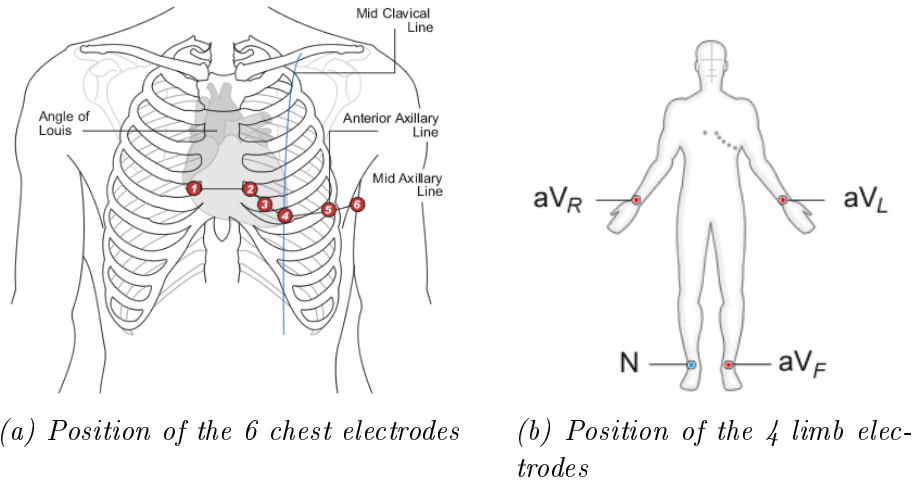


Figure 2.4: Universal 12-lead ECG (taken from [23]).

ECG data collected during the experiment from which this thesis is based only came from a lead I configuration.

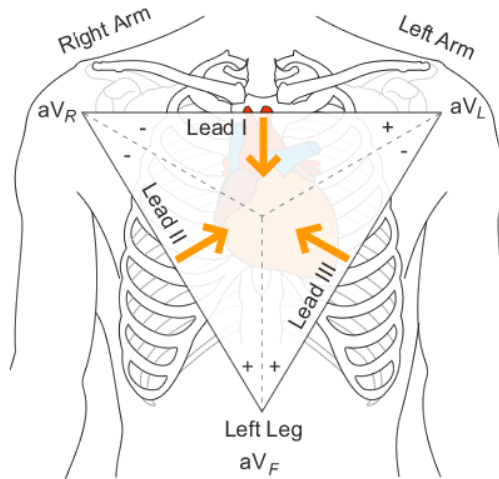


Figure 2.5: Bipolar leads forming the Einthoven's triangle with the heart on the middle (taken from [23]).

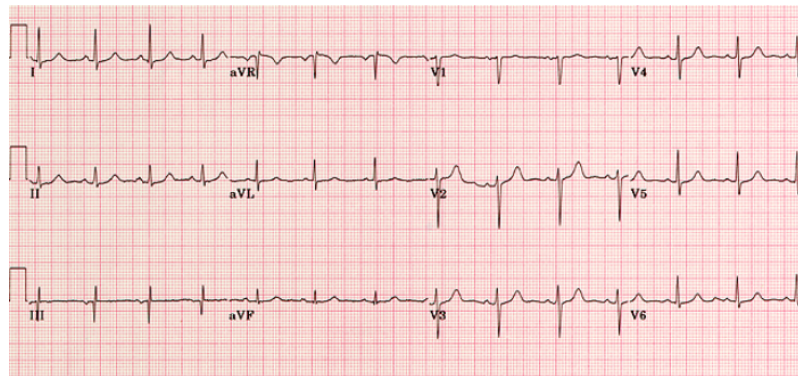


Figure 2.6: Example of a 12 lead ECG recording (taken from [24]).

2.2.4 Heart rate variability (HRV)

The ECG signal is not constant over time. More precisely, there are variations in the time difference between successive R peaks. The heart rate variability refers to this beat-to-beat alterations in the heart rate [25]. In the literature, RR variability is a common synonym used when speaking of HRV [26]. A term that is also often used is the NN interval. This term stands for normal sinus to normal sinus inter beat interval [27].

The first thing to do when dealing with HRV is to compute and draw the RR tachogram. As can be seen in Fig. 2.7, a tachogram is a graph where both axes have time units. The horizontal one corresponds to the beats locations in the time domain. The vertical axis indicates the time elapsed between the current beat and the previous one. It is important to notice that, as the tachogram reflects HRV, it is an unevenly (irregularly) spaced signal along time (on the horizontal axis) [10]. The tachogram of Fig. 2.7 is a tachogram computed during this thesis. We will explain later how this signal was obtained.

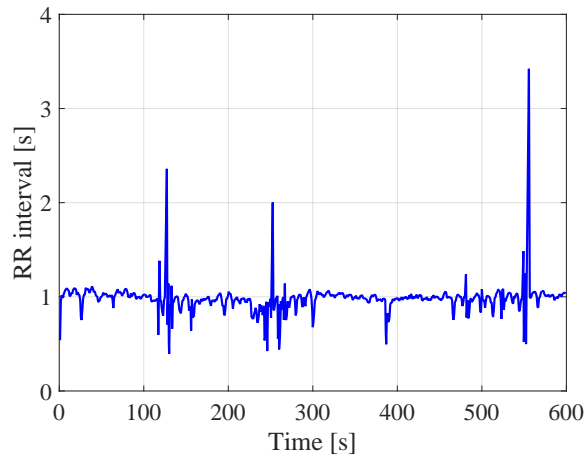


Figure 2.7: Example of a tachogram computed during Chapter 3.

Once this tachogram is available, HRV can be quantified by using some metrics. These metrics can be analyzed in three main different domains [26]:

1. Time domain: in this domain, different statistical parameters are computed based of the time elapsed between R peaks.
2. Geometrical domain: in the geometrical domain, NN time series are displayed according to a specific pattern. For instance, it can be the density distribution of the duration of NN intervals.
3. Frequency domain: variations of the RR intervals can also be observed in the frequency domain. The analysis in the frequency domain is done thanks to the power spectral density (PSD) of the tachogram. For instance, there is the power in the low frequency part (LF: $[0.04; 0.15]$ Hz), the one in the high frequency part (HF: $]0.15; 0.40]$ Hz), ...

The HF interval is mainly driven by the parasympathetic system. This result is generally accepted by scientists [26]. However, for the LF interval, there are lots of controversies. Some scientists think that the vagal and sympathetic activities are present in this interval. Others consider that the LF part only reflects the activity of the sympathetic activity. The LF/HF ratio is thus

estimated as the sympathovagal balance between the vagal and sympathetic systems. Generally, in a PSD, two peaks can be observed [10]. One is in the LF interval around 0.1Hz due to Meyer waves. The other is around 0.25Hz in the HF band and is due to respiratory oscillations (respiratory sinus arrhythmia (RSA)).

2.3 State of the art of the use of ECG to detect drowsiness

Based on the different parameters obtained from the HRV analysis, different research teams have conducted experiments to try to predict the state of drowsiness of people.

Mahachandra et al. carried out a study in which they attempted to predict the sleepiness event thanks to HRV [28]. More precisely, they asked several drivers to drive in a simulator during three and one half hours starting at 8 pm. The domination of the θ activity in frequency band of the EEG during 3 to 14 seconds was used to detect sleepiness events. The authors wanted to find HRV parameters that would allow to detect sleepiness events eight minutes before they are recorded on the EEG. The results of this study showed that the root mean square of squared differences of successive RR intervals (RMSSD) and the SD1 parameters of the Poincaré plot were the two parameters with the best sensitivity/discriminability index (d') to detect drowsiness. In their study, they noticed that the parameters of the frequency domain did not perform well. The authors explain it by the fact that parameters were computed eight minutes before each sleepiness event really occurred in the driver's body and were detected by the EEG.

Another study by Ibanez et al. tried to assess the drowsiness of ten professional driver based on their HRV parameters [29]. During the experiment, subjects drove in their real environment. Videos were recorded when subject was driving. Thanks to these videos, two independent observers classified the drowsy states of subject based on their eyes movements, eyes closed time, etc. The results have shown a reduction of the beat frequency and a higher HRV. Moreover, this study concludes that HRV parameters can be used to detect drowsiness. However, these parameters must be used together.

In the work performed by Chua et al., 24 subjects were forced to remain awake during 40 hours in a constant environment beginning at 8 am. Every 2 hours, they performed a psychomotor vigilance task (PVT) during 10 minutes [30]. In these tests, subjects had to respond the fastest they could to stimuli on a screen. The time between successive stimuli was randomly chosen between 2 and 10 seconds. If the patient had a reaction time higher than 0.5 seconds, it was considered as a PVT lapse. The number of lapses was considered as a reflection of the drowsiness state. During the 16 first hours, the number of lapses found was quite small. Then, during the normal hours of sleep, this number increased reaching a peak after 24 hours of wakefulness. After this moment, the number of lapses decreased a bit but remained much higher than the number obtained during the 16 first hours of the procedure. This recovering was assumed to be the action of the circadian cycle.

To establish their results, Chua et al. first normalized with the z-normalization

the PVT lapses and the different physiological parameters within subjects. Then, thanks to the Pearson coefficient, they found the best HRV parameter that linearly correspond to the PVT profile. This parameter was the RR-interval PSD in the 0.02 to 0.08 Hz frequency range. Next, they compared the ability of this parameter with the EEG PSD and the PERCLOS (=percentage of eyelid closure over the pupil over time) to forecast the increase of PVT lapses above different thresholds (25%, 50%, and 75%) thanks to a binary classification. This increase of PVT lapses is relative to performance of the subject during the baseline (= beginning of the experiment when the subject was considered as fully awake). Based on the ROC curves, the optimal threshold is found for each feature. At its optimal threshold, RR-interval PSD in the 0.02 to 0.08 Hz frequency range provided similar results to the PERCLOS and the EEG. These results tend to conclude that, if the RR-interval PSD in the 0.02 to 0.08 Hz observed at a specific moment is compared to the one when he is well rested, it is possible to predict the increase of PVT lapses above specific threshold.

Vicente et al. also tried to predict the drowsiness based on the HRV [31]. In this study, two databases provided by FICO MIRRORS S.A. were used. The first one consists of 11 subjects deprived of sleep in a driving simulator during 120 minutes. The second contains the ECG of ten professional drivers without sleep deprivation was recorded during six hours. Every two hours, these drivers needed to take a break in their drive. In each of these data banks, an external observer classified, in every minute, the state of the driver as drowsy or awake based on the EEG record. The ECG was analyzed during five minutes. Thanks to a moving window by step of one minute, they were able to predict the subject's condition every minute. The HRV signal was obtained thanks to the *integral pulse frequency modulation model*. Then the HRV parameters were computed from the *Smoothed Pseudo Wigner-Ville Distribution*. The classifier used in this study was based on a linear discriminant analysis. When using the two data banks together to have roughly the same proportions of drowsy (1663) and awake (2162) states and only the best features, a positive predictive value of 93% and a sensitivity of 85% was obtained.

The use of the HRV derived from the ECG to detect drowsiness is a new field of study. Therefore, even if some research has found good results for predicting drowsiness based on the HRV, there exist some variations in the results. For instance the ratio LH/HF has been regarded a lot as being the reflection of the sympathovagal balance. Machandra et al found that the ratio LH/HF increases when drowsiness at wheel occurs [28]. On the other hand, in a study of Sahayadhas et al [32], a decrease of this ratio was observed when subjects were very drowsy. Shinar et al also studied the frequency component of the HRV during the transition between wakefulness and sleep in subjects without sleep deprivation [33]. They noticed a reduction of the LF power in normal subjects (without sleep troubles) around the sleep onset contrary to HF part that did not significantly changed. Therefore the sympathovagal balance shifts towards an increase of the parasympathetic activity.

The main problem in all these studies is that the results do not always come from the same protocol of experiment. In some studies the reference to detect drowsiness is the EEG; in others it is the recording of videos that is used as the reference. Therefore, the definition of drowsiness is not always the same from one study to another. An other problem may be the variability of the state of the driver between different experiments. In some research work, drivers were sleep deprived before doing the test while it was not the case for other research work. Finally, the use of HRV is still a new field of research.

For all of these reasons, the aim of this work is to establish the feasibility, or not, of the ECG to detect drowsiness.

2.4 Flowchart of the work performed in this thesis

The core of this thesis is about the development of an automatic drowsiness monitoring system using the ECG. As previously said, the detection of the drowsiness using the ECG is a recent field of research and development. Therefore, different different studies do not always reach the same conclusions and/or the same protocol of experimentation. For these reasons, the goal of this thesis is not only to develop the drowsiness monitoring system but is also to assess the feasibility of the ECG to detect drowsiness.

The experiment that this thesis is based on is an experiment conducted 2 years ago at the University of Liège. 25 healthy volunteer subjects were chosen to perform three psychomotor vigilance task (PVT) tests at three different degrees of sleep deprivation. The evolution of the experiment with time is the following:

Each subject arrived in the lab for the first PVT test at 8 am after a normal night of 8 hours of sleep (subjects were asked to wake up at 7 am). The first PVT test was done between 8 and 9 am. The following signals were collected during each test: EEG (F_z , C_z , P_z , C_3 , and C_4 canals), ECG, EMG, EOG and eyes images. The drowsy state of each subject is known by using the ECG, EOG. The study of these two signals was analyzed previously to this work. Clémentine François and Baudouin Fortemps de Loneux formed by the Doctor Robert Poirrier analyzed these signals and gave the corresponding KDS score reflecting the drowsiness state of the subject. The second PVT test was done between 2 and 3 am after 19 hours of sleep deprivation. Finally, the last test took place between 11 and 12 am after 28 hours of sleep deprivation. Each PVT test lasts 10 minutes and consists of a light stimulus displayed on a screen during 400 milliseconds. The time elapsed between two light stimuli is chosen randomly between 5 and 15 seconds. At each stimulus, the subject has to press the fastest possible on a button. This task, considered as boring, is used to facilitate the drowsy events.

The work flow of this thesis can thus be summarized in Fig. 2.8. The first block in this Figure is the problem definition. The problem definition has already be explained previously in this chapter such as the data collection. The next steps of the flowchart are the data processing and the features extraction. These two blocks will be explained in details in chapter 3. Then, the next block is the machine learning part and the presentation of the results. Chapters 4 and 5 will deal of these two blocks. Finally, a discussion of the results will also be performed at the end of this thesis.

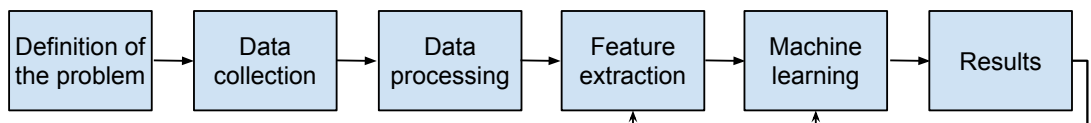


Figure 2.8: Flowchart of the work performed in this thesis.

Chapter 3

Feature extraction

In order to develop a system that automatically predicts the level of drowsiness of a person, a machine learning model needs to be developed. This model will be explained in detail in the next chapter. For the moment, a machine learning model can be seen as a black box. This black box receives parameters (also called features) as input. Then, based on the algorithm insides this black box, the model predicts the output by using the parameters given as input. The input parameters are thus important for the model to work. In our problem, the different features computed should normally be a function of the drowsiness. This chapter is devoted to the computation of these features. In the first part, all the signal filtering steps to detect the R peaks in the ECG is explained. Once this filtering step is done and the R peaks are identified, the features from the ECG can be computed. Therefore, the second part of this chapter describes the features of the heart rate variability (HRV) in the different domains. The definitions and the formulas of these features are also explained in this chapter. Finally, the last part of this chapter deals with the respiratory features. More precisely, it explains how it is possible to derive the respiratory signal from the ECG and how features in the respiratory domain can be derived from this last signal.

3.1 R peaks detection algorithm

The ECGs recorded during the different PVT tests can not be used directly in order to extract the different features indicating the level of drowsiness of the subjects. Indeed, a processing step must be performed in order to find the R peaks of the ECG and then extract these parameters.

The R peaks detector algorithm developed in this thesis is strongly based on the article of Benitez et al. [1]. In the literature, lots of algorithms developed use the MIT-BIH arrhythmia database to evaluate their accuracy. This database is an open-source database [27]. Each ECG record of the database has been annotated independently by two cardiologists. Thanks to these annotations, it becomes easy to assess the different algorithms. The one designed by Benitez et al. detects the R peaks in the QRS complex of the ECG with a sensitivity of 99.94%, a positive prediction rate of 99.93% and an average detection rate of 99.87%. As it provides good results on this database, we decided to choose this algorithm for the

This algorithm is thus reimplemented based on the description of Benitez et al. [1]. The main components of this algorithm are summarized in Fig. 3.1. The rest of this section will be devoted to the explanation of the role of each component. In Fig. 3.1, the first part is the filtering part. As its name says, it filters the raw ECG signal. Then, the derivative of the signal is computed. After this step, the Hilbert transform of the derivative is taken. Based on this last signal, a first peak detector is applied to locate the positions of the R peaks candidates. Finally, based on the positions of these candidates, the true peaks in the original signal are found.

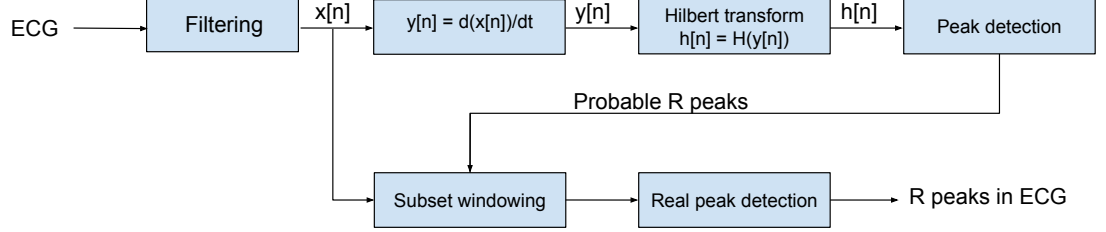


Figure 3.1: Flowchart of the algorithm developed for the R peaks detection based on [1]

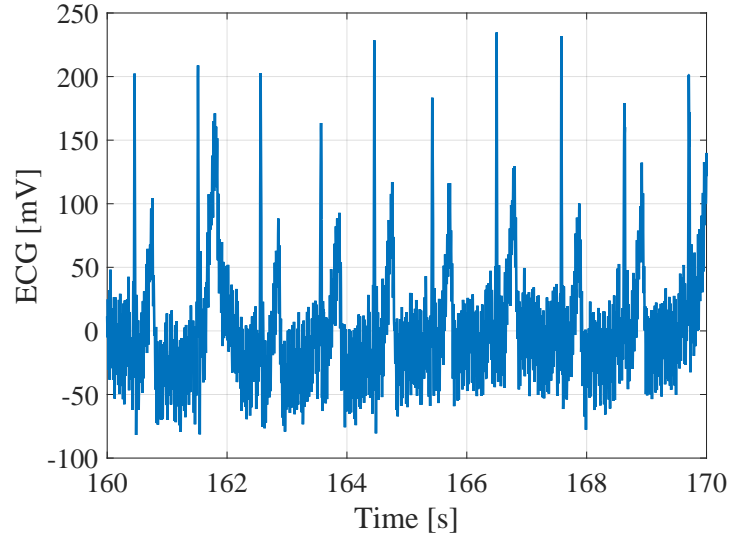
3.1.1 Filtering

If the plot of a raw ECG as a function of time is observed, this signal contains lots of noise. In fact, there exist several sources of noises for the ECG [34, 35]. The frequency components of these noises are different for each one. As the useful bandwidth of an ECG is comprised between 0.05 and 100 Hz, these noises act on different parts of the spectrum of the ECG.

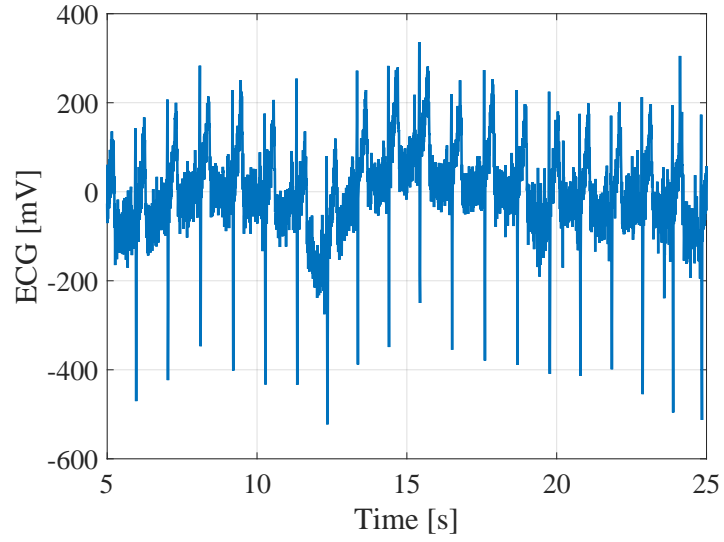
The first kind of noise is the power line interference at 50 Hz (as the study was conducted in Europe). Figure 3.2a shows the consequence of this noise. The second kind of noise is mainly due to respiration. This noise is a low frequency noise (below 0.1-0.5 Hz) [34]. The consequence of this noise is that the horizontal reference line of the ECG is moving with time. This phenomenon is called the baseline drift. An example of baseline drift is displayed in Fig. 3.2b. Another noise is the motion artifacts. This noise occurs when the patient is moving during the recording and is shown in Fig. 3.2c. It is characterized by big spikes in the ECG. Finally the last source of noise is the muscular noise (also called the EMG noise). This kind of noise generally interacts in the high frequency part of the useful band of the ECG. The consequences of this noise are generally spikes at a high frequency. This noise was not clearly identified in the dataset of the experiment conducted at the University of Liège. The probable explanation is that this noise was present but is hidden by the important contribution of the power line interference in the signal.

In order to get rid of these noises, a filtering process has been used. The filter implemented is a band pass finite impulse response filter (FIR filter) using a Kaiser-Bessel window.

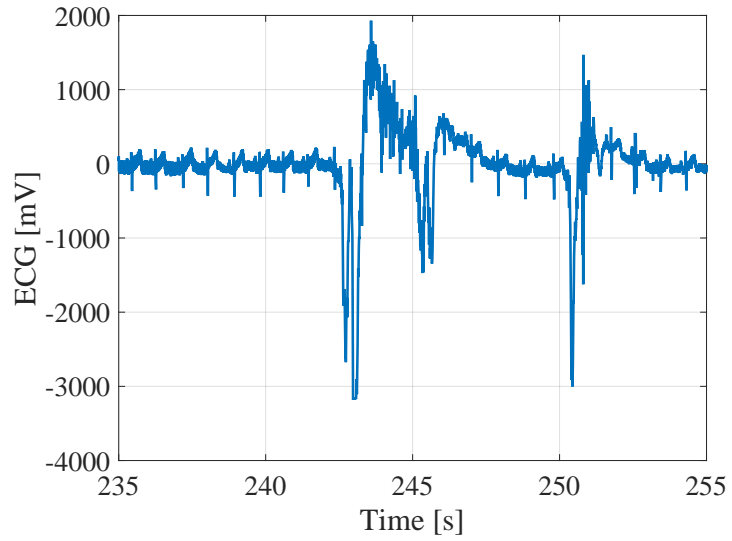
A FIR filter can be represented by the block diagram in Fig. 3.3. In this figure, the order of the FIR filter is equal to N . As the filter has an order N , there are also N delays. Therefore, a FIR filter is also called a tapped delay line. Based on Fig. 3.3,



(a) Power line interference at 50 Hz in an ECG.



(b) Baseline drift effect in an ECG.



(c) Motion artifact occurring in an ECG.

Figure 3.2: Noise effects occurring in different ECGs used in this project.

the impulse response of this filter can simply be written as:

$$h[n] = \begin{cases} 0 & \text{if } n < 0 \\ b_n & \text{if } 0 \leq n \leq N \\ 0 & \text{if } n > N \end{cases} \quad (3.1)$$

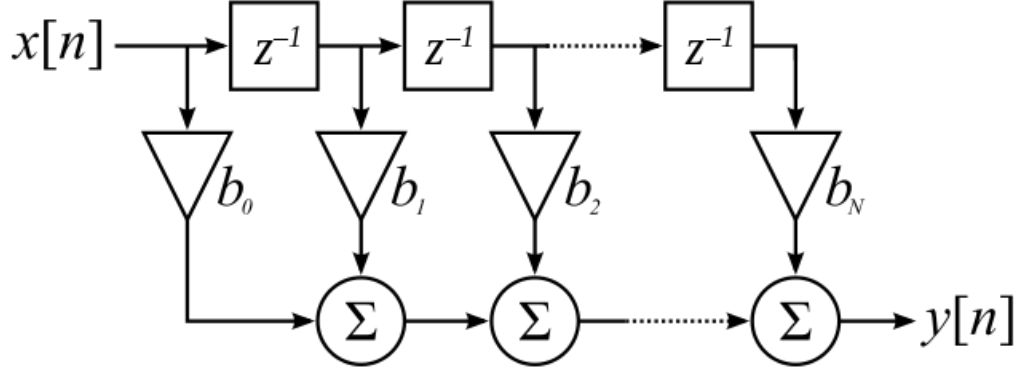


Figure 3.3: Block diagram of an FIR filter (taken from [36]).

Based on Eq. 3.1, the number of elements b_n different from zero (M) is equal to the order of the filter (N) plus one ($M = N + 1$). This number M is clearly finite. Therefore, impulse response is also finite which gives the name to the finite impulse response. The output of the filter can be written as:

$$\begin{aligned} y[n] &= \sum_{l=-\infty}^{l=+\infty} h[l] \cdot x[n-l] \\ &= \sum_{l=0}^{l=N} b_n \cdot x[n-l] \\ &= b_0 \cdot x[n] + b_1 \cdot x[n-1] + \dots + b_N \cdot x[n-N]. \end{aligned} \quad (3.2)$$

Another interesting property of the FIR filter is that it can have a linear phase. This condition is met when the coefficients b_i are symmetric with respect to the coefficient at the center $b_{\frac{N}{2}}$. If this condition is true, the phase of the frequency response is linear. Therefore, the delay is the same for all the samples (half of the order of the filter) and there is no delay distortion. This delay distortion can also called phase distortion.

The FIR filter implemented in this work is a band pass filter between 8 and 20 Hz. The Kaiser-Bessel window was used as the window technique. In fact, all types of filter (band pass, band stop, high pass) can be understood and obtained from the low pass filter type. An ideal low pass filter is a filter that let pass all the frequencies below its cut-off frequency and it is the inverse for all the frequencies higher than this cut-off frequency. Figure . 3.4 shows the modulus of the frequency response of the ideal low pass filter. The frequency axis of Fig. 3.4 is expressed in rad/s. Note that as the signal is in the digital world, the maximal achievable frequency is half of the sampling frequency according to the Nyquist criterion.

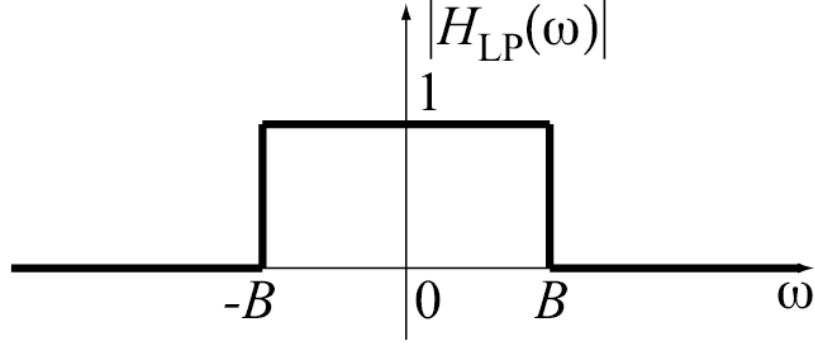


Figure 3.4: Modulus of the frequency response of the ideal low pass filter (taken from [37]). The frequency is expressed in rad/s and the cut-off frequency is named B .

The impulse response of this ideal filter is equal to:

$$\begin{aligned}
 h[n] &= \text{DTFT}^{-1}(H(\omega)) \\
 &= \frac{1}{2\pi} \int_{-\omega_c}^{+\omega_c} e^{j\omega n} d\omega \\
 &= \frac{1}{2\pi j n} [e^{j\omega_c n} - e^{-j\omega_c n}] \\
 &= \frac{\sin[\omega_c n]}{\pi n} \\
 &= 2f_c \text{sinc}[2f_c n],
 \end{aligned} \tag{3.3}$$

where ω_c and f_c are respectively the cut-off frequency in rad/s and in Hz (with the relation $\omega = 2\pi f$).

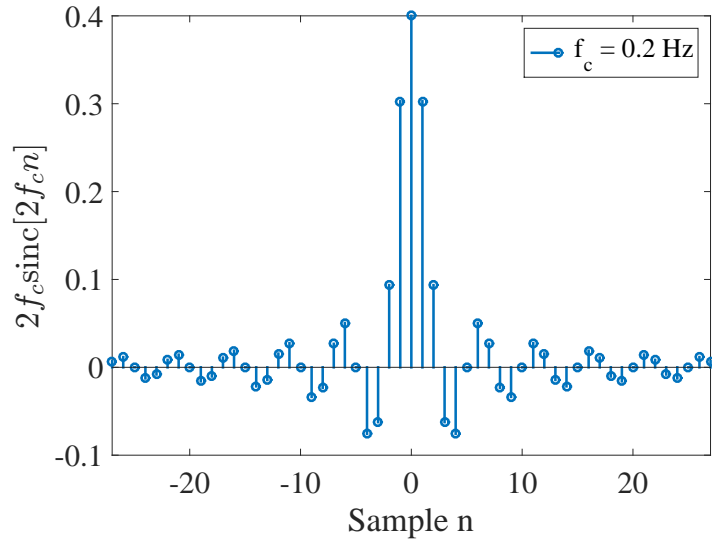


Figure 3.5: Impulse response of an ideal low pass filter (the cut-off frequency = f_c).

The problem of the impulse response of Fig. 3.5 is that this impulse response is infinite and non-causal. However, the one of the FIR filter (Eq. 3.1) has a finite length and is causal. To tackle this problem, first the impulse response is shifted to the left. This step introduces the delay between the signal at the output of the filter and the signal at its input. Then, a window technique is used to obtain a

finite number of samples. This window technique can be of different shape: Kaiser-Bessel, Hamming, Rectangular, Bartlett, ... This window has a finite number of coefficients different from 0. As each element of the impulse response is multiplied by the corresponding element of the window, the result becomes finite. Indeed, only the samples of the impulse response multiplied by the non-zeros coefficient of the window will be different from 0. Having a finite number of samples in the impulse response has as consequence that ripples appear in the frequency response of the filter. Moreover, the transition band is not a straight line as in the ideal low pass filter.

Taking into account the delay and the windowing, the impulse response of the FIR low pass filter can be finally written as:

$$h_{FIR}[n] = \begin{cases} W[n] \cdot 2f_c \text{sinc}[2f_c(n - \frac{N}{2})] & \text{if } n \neq \frac{N}{2} \\ W[n] \cdot 2f_c & \text{if } n = \frac{N}{2} \end{cases} \quad (3.4)$$

where $W[n]$ is the value of the window at sample n , f_c is cut-off frequency, and N is the order of the filter.

Here a band pass filter is required for filtering the ECG. If h_1 is a low pass filter with a cut-off frequency f_{c1} and h_2 is another low pass filter with a cut-off frequency f_{c2} ($f_{c2} > f_{c1}$), a band pass filter between f_{c1} and f_{c2} can be obtained just by subtracting h_2 to h_1 . Therefore, the impulse response of the band pass FIR filter can directly be written as:

$$h_{BPFIR}[n] = \begin{cases} W[n] \cdot \left(2f_{c2} \text{sinc}[2f_{c2}(n - \frac{N}{2})] - 2f_{c1} \text{sinc}[2f_{c1}(n - \frac{N}{2})] \right) & \text{if } n \neq \frac{N}{2} \\ W[n] \cdot (2(f_{c2} - f_{c1})) & \text{if } n = \frac{N}{2} \end{cases} \quad (3.5)$$

In this work, $f_{c1} = 8$ Hz and $f_{c2} = 20$ Hz (in absolute frequency). The window used is a Kaiser-Bessel window. Therefore, $W[n]$ can be developed as:

$$W[n] = \frac{I_0 \cdot \left(\beta \sqrt{1 - \left(\frac{n - \frac{N}{2}}{\frac{N}{2}} \right)^2} \right)}{I_0(\beta)}$$

$$\beta = \begin{cases} 0.1102 \cdot (\alpha - 8.7) & \text{if } \alpha > 50 \\ 0.5842 \cdot (\alpha - 21)^{0.4} + 0.07886 \cdot (\alpha - 21) & \text{if } 21 \leq \alpha \leq 50 \\ 0 & \text{if } \alpha < 21, \end{cases}$$

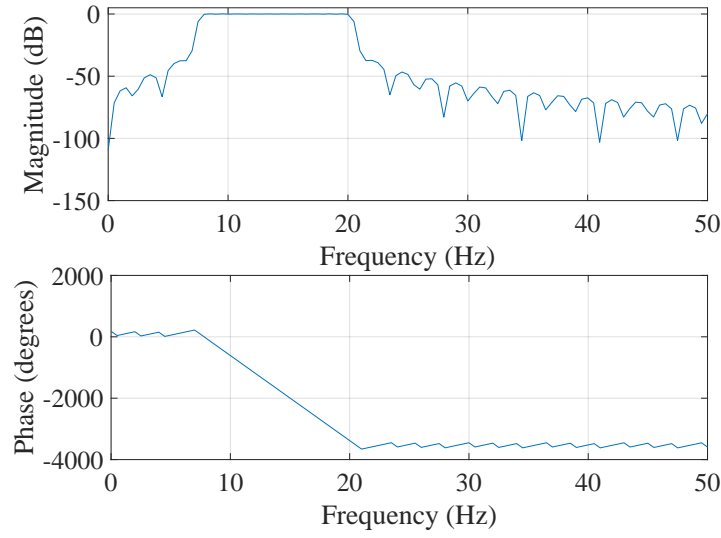
where I_0 is the modified Bessel function of order 0 and α is the sidelobe attenuation in decibels (dB).

The optimal order of the filter and the window coefficients $W[n]$ were found using the method *kaiserord* of Matlab. The specifications were to have an attenuation of 30dB between 7Hz ($= f_{c1} - 1$) and 8Hz ($= f_{c1}$) and between 20Hz ($= f_{c2}$) and 21Hz ($= f_{c2} + 1$). The maximal pass band ripple was chosen to be 10%. Given all these specifications, the function *kaiserord* designed a filter of order $N = 788$. The frequency response and the impulse response of this filter are shown respectively in Fig. 3.6a and Fig. 3.6b. From the impulse response shown in Fig. 3.6b, it is clearly visible that the condition of symmetry around the center point ($\frac{N}{2}$) is verified.

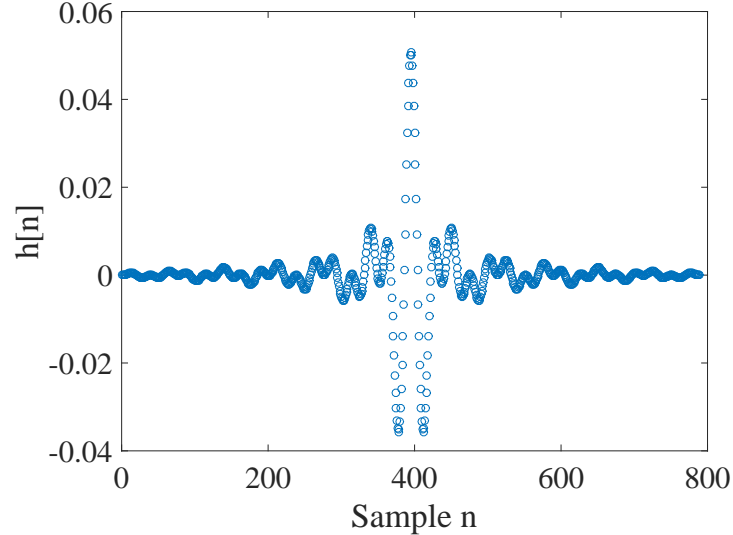
Therefore, the filter has a linear phase as it can be also seen in Fig. 3.6a. There is thus no phase distortion and the delay introduced by the filter is equal to:

$$\text{delay} = \frac{N}{2} = \frac{788}{2} = 394 \text{ samples}_{f_s=512\text{Hz}} = 7.69 \cdot 10^{-1} \text{ second}$$

There is thus a delay of $7.69 \cdot 10^{-1}$ second between the raw ECG and the filtered one. As in the flowchart of Fig. 3.1 the original signal is used at the end to find back the true position of the R peaks. There must not be any delay between the raw ECG and the filtered one. To do so, the filtered signal must be shifted to the left of a value equal to the delay of the filter to compensate this delay. This delay compensation has a cost. Indeed, the 394 last samples will be lost with this operation. However, as it only represents $7.69 \cdot 10^{-1}$ second, a negligible part of the ECG is lost in this operation.



(a) Frequency response of the band pass FIR filter.



(b) Impulse response $h[n]$ of the band pass FIR filter.

Figure 3.6: Frequency response and impulse response of the band pass FIR filter implemented in this work. The passband is between 8Hz and 20Hz.

Figure 3.7 shows the three noisy cases of Fig. 3.2 after the filtering step. In fact, by removing all the frequencies that do not belong in the pass band frequency of the

filter, all the noise that has a frequency outside this pass band (for instance 50Hz noise) is removed. However, the shape of the ECG waveform is a bit modified as all the frequencies of the useful spectrum of the ECG are no longer present. But the remaining frequencies highlight the QRS complex of the ECG [1].

3.1.2 Derivative

The second step in the processing chain is to take the derivative of the signal. The ECG signal here is in the digital domain. Therefore, the numerical derivative must be taken. For all the samples except the 2 at the limit, the central difference approximation is thus used [38]:

$$ECG'[n] \approx \frac{ECG[n+1] - ECG[n-1]}{2T}, \quad (3.6)$$

where $ECG'[n]$ is the derivative of the ECG at sample n , T is the inverse of the sampling frequency f_s ($= 512\text{Hz}$).

For the first and last point of the ECG, this method can not be used. The backward and forward derivative are respectively used for these 2 points [38].

3.1.3 Hilbert transform

For the third step of the flowchart, the Hilbert transform of the derivative of the ECG obtained at the last step is computed. If $x(t)$ is a real signal, its Hilbert transform (noted $\hat{x}(t)$) is defined as:

$$\hat{x}(t) = x(t) \otimes \frac{1}{\pi t},$$

where the symbol \otimes is used to represent the convolution.

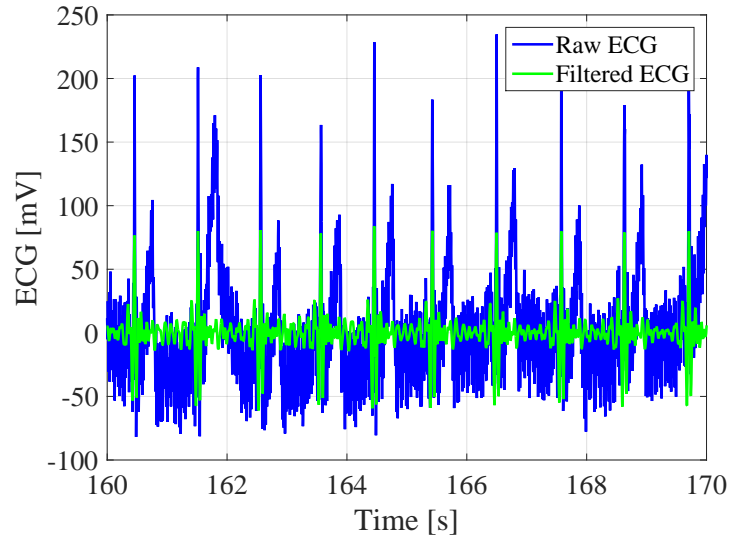
If $X(f)$ represents the Fourier transform of $x(t)$, the Hilbert transform can also be defined as:

$$\begin{aligned} \hat{X}(f) &= -j \cdot \text{sgn}(f) \cdot X(f) \\ &= \begin{cases} -j \cdot X(f) & \text{if } f > 0 \\ 0 & \text{if } f = 0 \\ +j \cdot X(f) & \text{if } f < 0 \end{cases} \end{aligned}$$

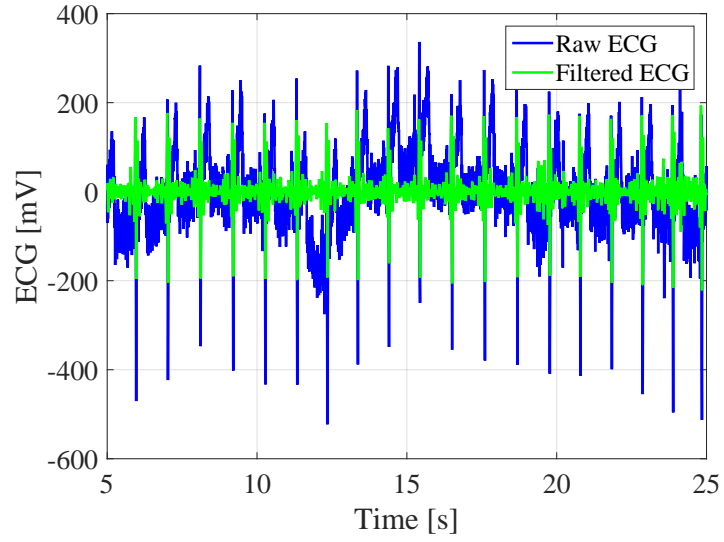
where $\text{sgn}(f)$ is the sign of the frequency (i.e. 1 if $f > 0$ and -1 if $f < 0$).

Therefore, in the frequency domain, the Hilbert transform can be easily understood. For the frequencies higher than 0, the signal $X(f)$ is shifted by $-\frac{\pi}{2}$. And a shift of $+\frac{\pi}{2}$ is applied for the frequencies below 0.

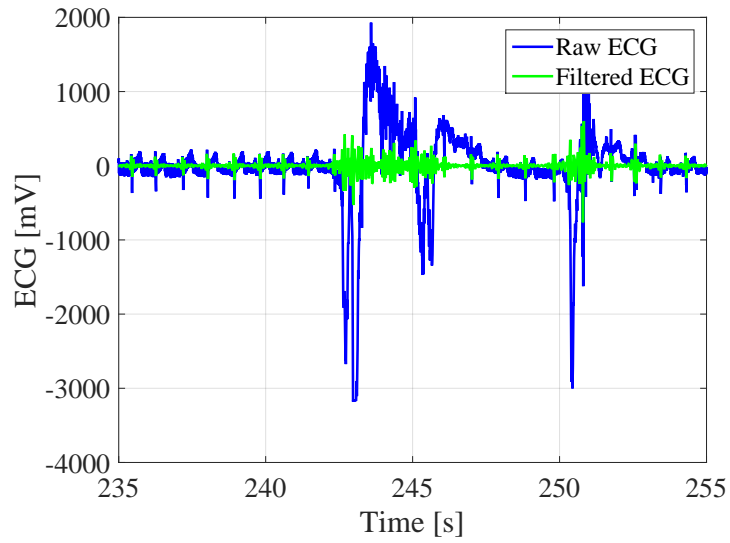
The Hilbert transform is an odd filter. It means that the Hilbert transform will be equal to zero every time there is an inflection point in the original signal. Therefore,



(a) Power line interference at 50Hz in an ECG.



(b) Baseline drift effect in an ECG.



(c) Motion artifact occurring in an ECG.

Figure 3.7: Impact of the FIR filter on the different noise effects of Fig. 3.2. In blue, the original noisy signal. In green, the FIR filtered signal.

a crossing of the horizontal axis in the original signal located between a positive and a negative inflection point will be translated as a peak in the Hilbert transform signal [1]. As the derivative of the ECG crosses 0 between a positive and a negative inflection point, locating the peaks in the Hilbert transform of the derivative of the ECG can be used to detect the positions of the R peaks in the ECG.

The Hilbert transform is computed using the function *hilbert* of Matlab.

3.1.4 First R peak detector and subset windowing

Once the Hilbert transform is computed, an adaptive threshold (τ) is applied to it to find the positions of the R peaks candidates. In fact, a moving window of 1456 points is taken. The length of this window is obtained by adapting the length taken by Benitez et al. ([1]) given the sampling frequency used in this work. At each time, the beginning of the next window corresponds to the last R peak found in the previous window. First, the level of noise in the current window is computed. This level of noise is obtained by using the root mean square (RMS) value of the signal, and it is used later to find the threshold to locate the peaks. The maximum value of the Hilbert transform signal is also computed in the window and is named $M(H[n])$. Another parameter γ equal to 1.4 is also introduced. Then:

- if $\text{RMS noise} \geq 0.18 \cdot M(H[n])$, where \cdot means multiplication, a high level of noise is present in the current window. Therefore, the threshold is set to $\tau = 0.39 \cdot \gamma \cdot M(H[n])$.
- if $\text{RMS noise} \geq 0.18 \cdot M(H[n])$ and if $M(H[n]) > 2 \cdot M_{-1}(H[n])$, where $M_{-1}(H[n])$ is the maximum value of the Hilbert transform in the previous window, a high level of noise is present in the current window but the threshold is adjusted to $\tau = 0.39 \cdot \gamma \cdot M_{-1}(H[n])$.
- if $\text{RMS noise} < 0.18 \cdot M(H[n])$, the level of noise is low in the current window. The threshold level (τ) becomes equal to $\tau = 1.6 \cdot \gamma \cdot \text{RMS noise}$.

The parameter γ has been added compared to the work of Benitez et al. [1]. The main reason is that the ECG signal in our work seems to be more noisy than the one of the MIT-BIH arrhythmia database. That is why the threshold τ is incorporated in this work. However, we recognize that it will be more robust to this optimal threshold automatically.

Once the threshold is obtained for the current window, the R peaks candidates can be simply found by searching the peaks above this threshold. This step is done by using the function `findpeaks` of Matlab. However, the peaks in the Hilbert transform cannot be too close. In the work of Benitez et al., the minimum time duration between successive peaks is equal to 0.2 s. However, here the criterion of Goldberger et al. is used [27]. Therefore the minimum duration between successive peaks is equal to 0.4 s. It means that only one peak can be kept when 2 successive peaks are separated by less than 0.4 s. The one kept is found by using an adaptive criterion. More precisely, we keep the one that is closer compared to the previous located true peak to the mean interval duration between successive peaks and which is also closer to the mean amplitude of the previous true peaks found.

3.1.5 Second peak detector

Finally, the last step of the filtering (see Fig. 3.1) consists in finding the true location of each peak in the initial ECG. This is done by searching the true maximum in ± 15 samples around the peak candidate.

The R peaks in the ECG are therefore found. This was the first step in the feature extraction problem.

3.1.6 Result of the R peaks detection algorithm

Once the algorithm to detect the R peaks is programmed, it can be tested on the ECGs obtained during the PVT tests. To illustrate the accuracy of the algorithm developed, the noisy ECGs of Fig. 3.2 can be used for test in order to determine the performance of the algorithm on noisy ECGs. The result is shown in Fig. 3.8. In this figure, the ECG is shown in blue and the identified R peaks are indicated by green crosses. Figure 3.9 represents the results obtained after the Hilbert transform of the three noisy signals.

From Fig. 3.8a and Fig. 3.8b, one can conclude that the algorithm correctly finds the R peaks in presence of power line interference and baseline drift. In Fig. 3.8c, the problem of the motion artifacts effects can be seen. In fact, the motion artifacts generate lots of close peaks in the Hilbert transform. These peaks are nearly all above the threshold of the current window as it can be seen in Fig. 3.9c. The chosen peaks are based on the minimum time duration between successive peaks and the location with the previous true peaks as explained above. Even if this situation can lead to detection of false R peaks, in the next section, we will propose an additional criterion in order to filter these false R peaks found.

3.2 Computation of the features

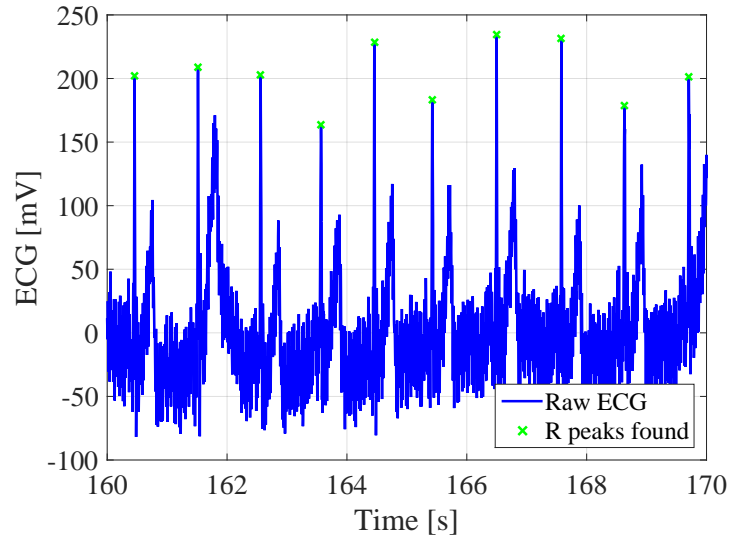
3.2.1 Creation of the tachogram

Once the R peaks are found, the tachogram can be generated. Just as a reminder, a tachogram is a graph where the horizontal axis corresponds to the beats locations in the time domain. The vertical axis indicates the time elapsed between the current beat and the previous one.

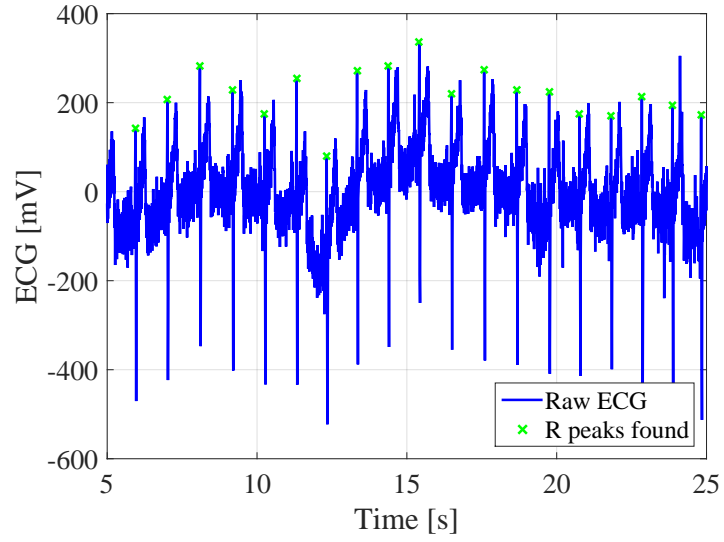
To compute all the features that the model will need, only the normal to normal (also called NN) intervals must be kept. There are intervals that are located between 2 consecutive QRS complexes and triggered by the depolarization of the sinotrial node [26]. Therefore, all the non NN intervals are removed. The criterion to decide if an interval must be kept or not is based on the work of Gari D. Clifford [10]. If RR_n is the time interval between the R peak occurring at time n and the one at time $n-1$, therefore this interval is kept if and only if:

$$0.755 \cdot RR_{n-1} < RR_n < 1.325 \cdot RR_{n-1}$$

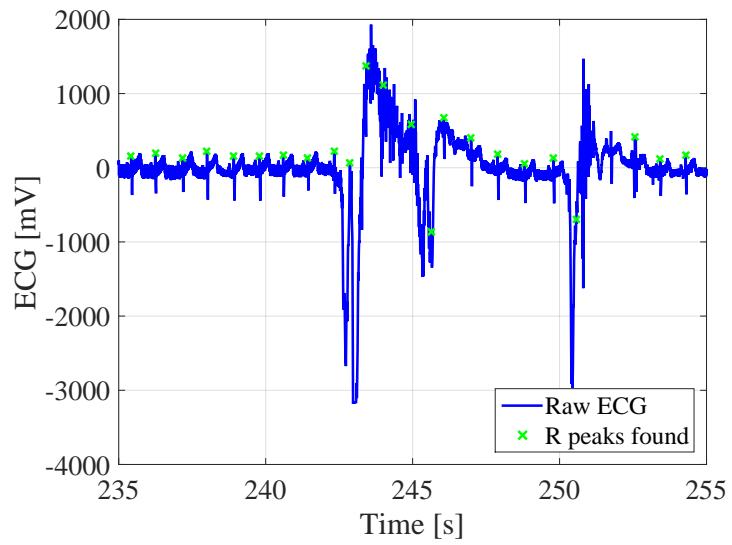
where RR_{n-1} is the time interval between the R peak occurring at time $n-1$ and the one at time $n-2$.



(a) Power line interference at 50Hz in an ECG.

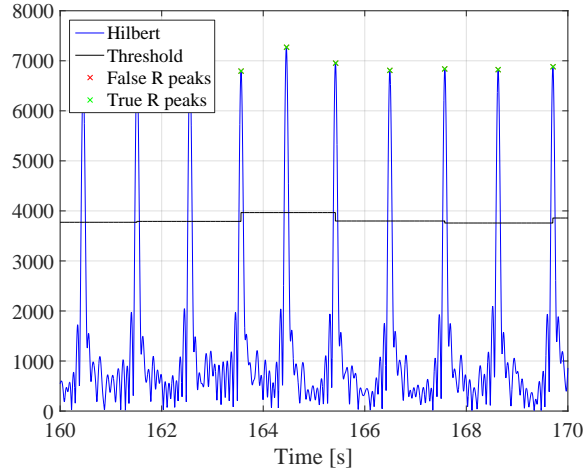


(b) Baseline drift effect in an ECG.

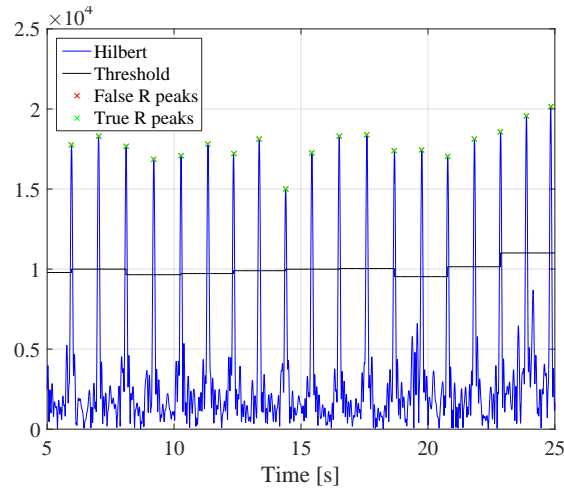


(c) Motion artifact occurring in an ECG.

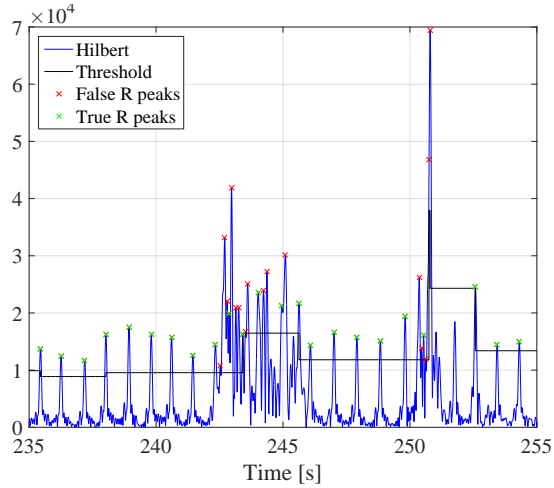
Figure 3.8: Results of the R peaks detector algorithm on the noisy ECGs of Fig. 3.2. The R peaks are shown by green crosses.



(a) Hilbert transform the power line interference at 50Hz in an ECG of Fig. 3.2a.



(b) Hilbert transform of the baseline drift effect in an ECG of Fig. 3.2b.



(c) Hilbert transform of the baseline drift effect in an ECG of Fig. 3.2c.

Figure 3.9: Hilbert transform of the three ECGs with noise shown in Fig. 3.2. The Hilbert transform is shown in blue. The threshold computed according to the level of noise ECG in that window is shown in black. The red crosses are peaks of the Hilbert transform above the threshold that are false peaks. The green crosses are the peaks found that are correct true peaks.

Then, it can still remain some problems in the remaining NN intervals. Indeed, sometimes, the R detection algorithm can detect a R peak that is not correct or it can miss a peak. The consequence is that some outliers can be present in the tachogram and, therefore, the features computed from this tachogram could be affected by these outliers. It is thus also important to remove them. Just to give an example of the importance of removing the outliers, according to Goldberger et al., some features in the frequency domain that will be explained later can be incorrect by more than 1000% just due to the presence of outliers [27].

Therefore, we must remove these outliers. The rule of Goldberger et al. is used here. First, if a NN interval is greater than two seconds, this interval is removed. Then, we use a moving window. The length of this window is equal to 41 samples. This window is centered on the current NN interval that is evaluated. Therefore, there are 20 NN intervals in each side of the current NN interval and we take the mean of these 40 NN intervals (named $\Delta_{t_{40}}$). We then compute the range of the possible time duration for the current interval as $[0.8 \cdot \Delta_{t_{40}}; 1.2 \cdot \Delta_{t_{40}}]$. If the current evaluated interval is not in this range, we remove it and it is considered as an outlier.

The impact of removing these outliers on the tachogram is shown in Fig. 3.10. This figure shows, at the top, the raw tachogram of the R peaks coming from the R peaks detection algorithm. This tachogram is the one obtained from the subject with motion artifacts of Fig. 3.2c. The tachogram lasts during all the PVT session here. At the middle of Fig. 3.10, we only keep the NN intervals. Finally, the bottom of Fig. 3.10 shows the NN intervals with the outliers removed. In Fig. 3.8c, we observed that the detection of R peaks results sometimes to detection of false R peaks for that subject during motion artifacts (around 250s on the horizontal axis). However, we can see in Fig. 3.10 that these false peaks are removed and, at the end, the tachogram without non NN intervals and outliers is obtained. However, it is important to remark that for some subjects, the number of non NN intervals and the number of outliers is smaller.

3.2.2 Time duration from which the features are computed

The features can be computed on time intervals of different lengths. The greater the time interval, the higher information in the ECG there is. However, a prediction of the model must be available as frequently as possible. Indeed, imagine a system that predicts the drowsiness state of the driver every five minutes. It means that every five minutes, the model predicts if the driver has been drowsy or not during the last five minutes. This time interval is too large to be useful. Indeed, in five minutes the driver has lots of time to fall asleep and to cause an accident. Therefore, the analysis of the features should be done on short time intervals of the ECG.

Another important constraint for this work is that the three PVT tests done during the experiment had a duration of 10 minutes each. This means that for each test, the ECG signal lasts 10 minutes. Therefore, to reach a sufficiently big dataset, features can not be extracted on an ECG of long duration. Moreover, during this experiment, the drowsy state of each subject was assessed using polysomnography signal (EEG and EOG). This assessment was based on windows of 20 seconds of EEG and EOG. Therefore, the predictions of the state of the subjects are only available every 20 seconds.

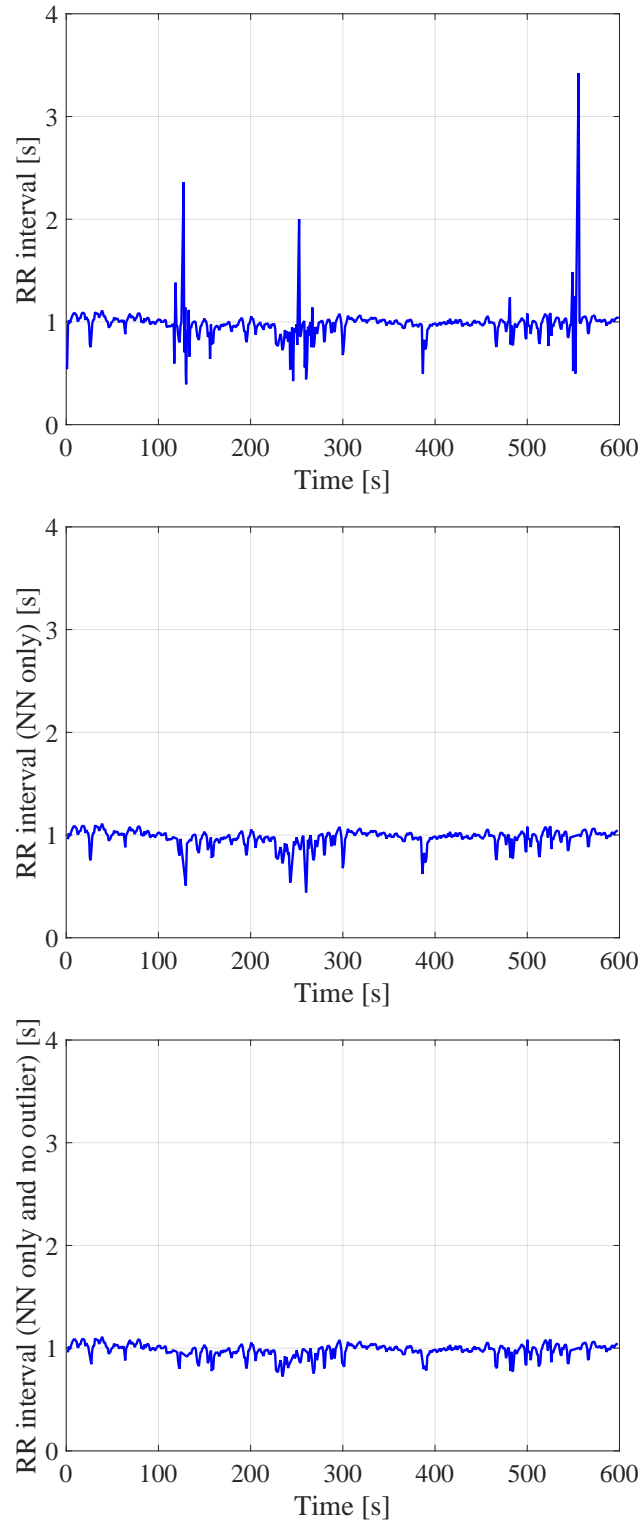


Figure 3.10: Successive steps in the generation of the final tachogram. Top: raw tachogram just after the R peak detection algorithm. Middle: the result when only the NN intervals are kept. Bottom: the outliers are also removed.

Finally, the *Task Force of the European Society of Cardiology and Others* have defined a minimum time duration for computing parameters in a given domain [26]. For instance, in the frequency domain, short term recording of the ECG are comprised between two and five minutes.

For all these reasons, we choose a moving window principle to compute the features on the ECG. Figure 3.11 shows this principle that can be summarized as follows: For each subject i during the the PVT test j , the tachogram on all the duration of the test (600 seconds) is computed using the technique explained previously. Then, we take only a window of 120 seconds of the complete tachogram. The features are computed on this window. Each window is then shifted by 20 seconds and we repeat the process until reaching the end of the tachogram. The model will therefore predict the state of the driver every 20 seconds based on an record of 120 seconds.

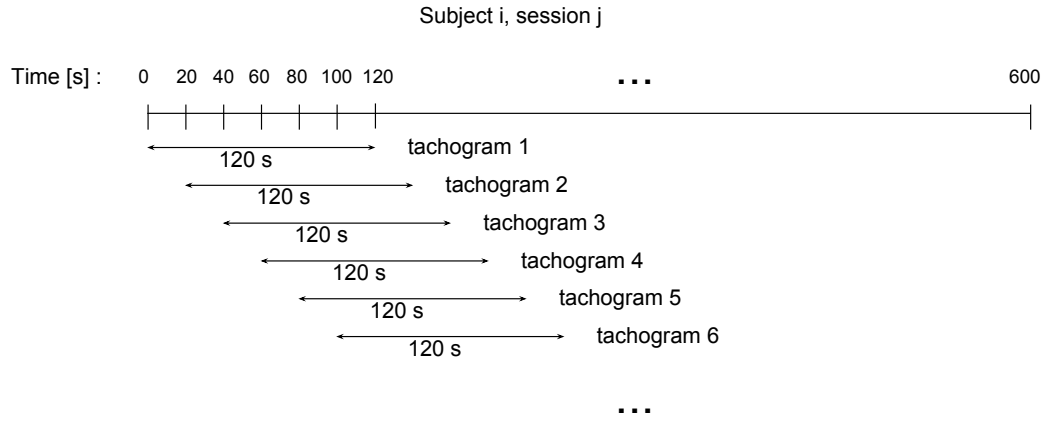


Figure 3.11: Illustration of the principle of the moving window used to compute the features for the machine learning model.

3.2.3 The HRV features in the different domains

The *Task Force of the European Society of Cardiology and Others* defined in its work the different parameters that can be computed from an ECG in order to assess the heart rate variability (HRV) [26]. It is these parameters that will be used as the features for the model.

These parameters can be computed in different domains. Here, three domains are used: time and the statistical domains, non linear domain, and. the frequency domain.

Note that as we only keep the NN intervals in the tachogram, in the next of this thesis, a RR interval will always be a NN interval.

We know present the three domains of interest

Time and statistical domains

From all the features that can be computed, the ones in the time and statistical domain are the easiest to compute and to understand. The parameters computed

in this domain are summarized in Table 3.1. Some parameters are the mean of a specific value like the mean of selected RR interval series (mean RR), and the mean of the hear rate (mean HR). The square root of the mean of the square of differences of successive RR intervals (RMSSD RR) is also based on the computation of a kind of mean. An other parameter is equal to the standard deviation of the NN intervals (SDNN). Finally, we also compute the number of consecutive RR intervals that differ more than 50 ms (NN50) and the percentage of this number in the whole intervals (pNN50).

Type	Units	Definition	Formula
Mean RR	s	Mean of selected RR interval series	$\frac{\sum_{i=1}^{nb_RR} RR_i}{nb_RR}$
SDNN	s	Standard deviation of selected RR interval series	$\sqrt{var(RR_intervals)}$
Mean HR	$\frac{1}{min}$	Mean heart rate of selected RR interval series	$\frac{60}{MeanRR}$
RMSSD RR	s	Root mean square of differences of successive RR intervals	$\sqrt{\frac{\sum_{i=1}^{nb_RR-1} ((RR)_{i+1} - (RR)_i)^2}{nb_RR-1}}$
NN50	[-]	Number of consecutive RR intervals that differ by more than 50 ms	$\sum_{i=2}^{nb_RR} 1 \text{ if } RR_i - RR_{i-1} > 50\text{ms}$
pNN50	[-]	Percentage of consecutive RR intervals that differ by more than 50 ms	$\frac{NN50}{nb_RR}$

Table 3.1: Table summarizing the names, units, definitions, and the formulas of the features in the time and statistical domains.

Non linear domain

Two parameters can be computed in the non linear domain, sometimes also called the fractal analysis [28]. These parameters are derived from the Poincaré plot. A Poincaré plot is a graph in which the horizontal axis is the duration of the current RR interval (RR_n). The vertical axis corresponds to the duration of the next RR interval (RR_{n+1}). Figure 3.12 represents an example of a Poincaré plot. The two non linear parameters that can be extracted from the Poincaré plot are called SD1 and SD2 [39]. SD2 represents the standard deviation (std) according to the identity line (i.e. the diagonal line with an angle of 45 degrees). SD1 is the std in the perpendicular direction to this identity line. SD1 and SD2 are in fact respectively the semi-minor axis and the semi-major axis of an ellipse. This ellipse is also shown in Fig. 3.12. More information about how to compute SD1 and SD2 can be found in Table 3.2.

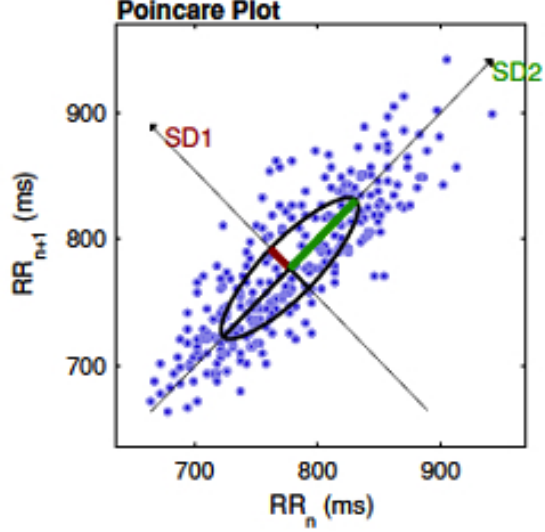


Figure 3.12: Example of a Poincaré plot. The figure shows the two parameters $SD1$ and $SD2$ extracted from this plot. (Taken from [40]).

Type	Units	Definition	Formula
SD1	s	Dispersion of the points (std) in the perpendicular direction of the identity line	$SD1^2 = 0.5 * variance(RR_n - RR_{n+1})$
SD2	s	Dispersion of the point (std) in the direction of the identity line	$SD2^2 = 2SDNN^2 - SD1^2$

Table 3.2: Table summarizing the names, units, definitions, and the formulas of the features in the non linear domain.

Frequency domain

The variations of the RR intervals can also be observed in the frequency domain. The analysis in the frequency domain is done thanks to the power spectral density (PSD) of the tachogram. The explanation of how to compute this PSD will be given later in this section.

By definition, the PSD of the tachogram, is divided in the frequency domain in specific intervals: the ultra low frequency (ULV: ≤ 0.03 Hz), the very low frequency (VLF: $]0.03; 0.04]$ Hz, the low frequency (LF: $]0.04; 0.15]$ Hz) and the high frequency part (HF: $]0.15; 0.40]$ Hz) [26]. However, in recordings from two to five minutes like the ones used here (two minutes), the Task Force recommends to only compute the VLF, LF, and HF components. For the VLF component, the Task Force even admits that it is a parameter that must be avoided for short recordings. Therefore, only the LF and the HF components are computed. Generally, it is advised to express these 2 parameters in absolute and normalized units. The normalization process enables to get a perception of the symptovagal balance. All the parameters in the frequency domain computed in this work as well as their units, and the way to compute them can be found in Table 3.3.

To compute the features in the frequency domain, the PSD of the tachogram must first be computed. Generally, the Fourier transform or an auto-regressive method

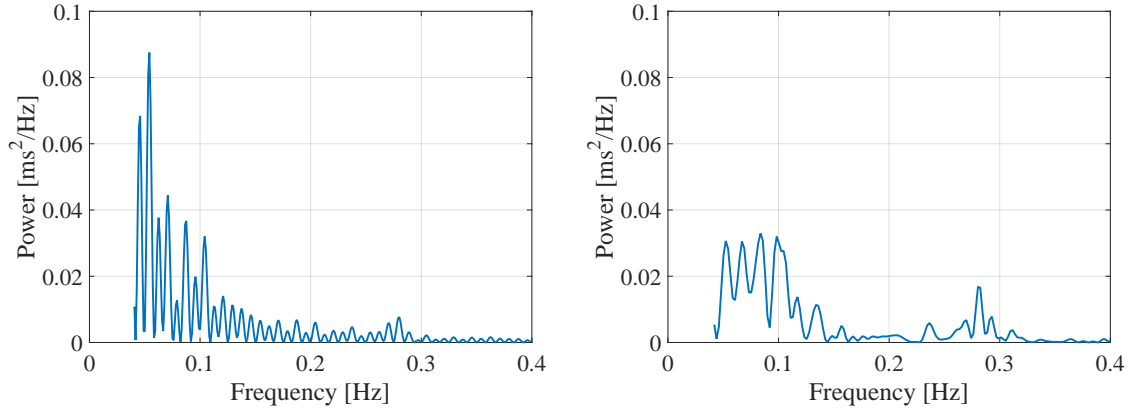
Type	Units	Definition	Formula
LF_f_peak	Hz	Frequency of the max in the LF band	$\max(\text{PSD}([0.04; 0.15] \text{ Hz}))$
HF_f_peak	Hz	Frequency of the max in the HF band	$\max(\text{PSD}([0.15; 0.40] \text{ Hz}))$
LF_power	s ²	Power in the low-frequency band	$\int_{f=0.04}^{f=0.15} \text{PSD}(f) \cdot df$
HF_power	s ²	Power in the high-frequency band	$\int_{f=0.15}^{f=0.4} \text{PSD}(f) \cdot df$
LF_power_norm	[-]	Power in the low-frequency band divided by the sum of the power in the low and high frequency bands	$\frac{\text{LF_power}}{\text{LF_power} + \text{HF_power}}$
HF_power_norm	[-]	Power in the high-frequency band divided by the sum of the power in the low and high frequency bands	$\frac{\text{HF_power}}{\text{LF_power} + \text{HF_power}}$
$\frac{\text{LF}}{\text{HF}}$	[-]	Ratio of the power in the low-frequency band divided by the power in the high frequency bands	$\frac{\text{LF_power}}{\text{HF_power}}$

Table 3.3: Table summarizing the names, units, definitions, and the formulas of the features in the frequency domain.

are the two most common ways of computing the PSD. However, these two methods operate on an evenly sampled signal (a signal in which a data is available every T second where T is the inverse of the sampling frequency f_s). But here, the tachogram is by definition an unevenly sampled signal. Indeed, there are variability in the heart rate: if it was not the case, it would not be possible to assess the HRV and the tachogram would be a constant, i.e. a perfect horizontal line. Therefore, to use the Fourier transform technique or the auto-regressive method, the tachogram needs to be re-sampled uniformly. This can be done for instance by using the linear or cubic-spline re-sampling techniques [10]. However, these techniques are not perfect. Indeed, they make the assumption that the tachogram signal evolves according to a specific shape between the different heart beats (RR intervals). Fortunately, there exists a technique that can compute the PSD on an unevenly sampled signal. This method is called the Lomb periodogram. It was not very popular at the beginning of the HRV analysis because of its high computational complexity. However, today, there are fast algorithms that have been implemented. Moody conducted a study in which he compared the Fourier transform, the auto-regressive method, and the Lomb technique [41]. His conclusion was the following: "*The Lomb method avoids all of the complications and pitfalls of re-sampling and replacement of outliers, and introduces no drawbacks of its own; in consequence, it is the method of choice for PSD estimation*". Clifford also concluded in his work that the Lomb method was better than the Fourier transform technique applied to a signal re-sampled uniformly

using cubic spline interpolation and where ectopics beats had been removed [10].

For all these reasons, in this thesis, the Lomb method is chosen as the technique to compute the PSD estimation in the frequency domain. Figure 3.13 shows the result of the PSD estimation using the Fourier transform and the Lomb method. As can be seen in this figure, it also seems that the Lomb method is more precise than the Fourier transform one. Indeed, it seems to have periodic pic in the Fourier transform. Moreover, the 2 peaks of the PSD can be more clearly with the Lomb method. The first peak is due to the Meyer wave (0.1Hz) and the second peak is due to respiratory oscillations (0.25Hz) (respiratory sinus arrhythmia (RSA)) [10]. Moreover, the Meyer waves (0.1Hz) peaks . The second one is around 0.25Hz in the HF band and is due to respiratory oscillations (respiratory sinus arrhythmia (RSA))



(a) PSD using the Fourier transform and cubic spline interpolation.

(b) PSD using the Lomb method.

Figure 3.13: Comparison of the PSD estimation using the Fourier transform and the Lomb method. Only the frequency in the LF and HF parts of the spectrum are represented as the PSD is based on a short term record of 120 seconds. (a): result of using the Fourier transform. Before applying this transform, the signal was re-sampled by using cubic spline interpolation at 8 Hz. (b): result of the Lomb method applied to the unevenly sampled signal.

The Lomb method implemented in this work is the Lomb-Scargle periodogram. Matlab already implements a method called *plomb* to estimate this periodogram. Therefore, this method was used in this project.

A periodogram is a power spectral density estimate of a signal. If $x(t_k)$ ($k = \{1, \dots, N\}$) are the N equally spaced measurements of the signal at times t_k , the periodogram is defined as

$$\begin{aligned} P_x(\omega) &= \frac{1}{N} \sum_{k=0}^{N-1} |x(t_k) e^{-j\omega t_k}|^2 \\ &= \frac{1}{N} \sum_{k=0}^{N-1} [(x(t_k) \cos(\omega t_k))^2 + (x(t_k) \sin(\omega t_k))^2]. \end{aligned} \quad (3.7)$$

Scargle modified the formula of the periodogram for the case of an unevenly signal as follows: If $x(t_i)$ ($i = \{1, \dots, N\}$) are the N unevenly spaced measurements of the

signal at times t_i , the periodogram is defined as

$$P_x(\omega) = \frac{\left[\sum_{i=1}^N x(t_i) \cos(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \cos(w(t_i - \tau))^2} + \frac{\left[\sum_{i=1}^N x(t_i) \sin(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \sin(w(t_i - \tau))^2}, \quad (3.8)$$

with τ such as

$$\sum_{i=1}^N \cos(w(t_i - \tau)) \sin(w(t_i - \tau)) = 0.$$

Scargle shows in the appendix C of his paper that this definition of the periodogram was identical to the equation of the least squares fitting of the unevenly sampled signal by sine waves that Lomb previously established [42, 43].

Indeed, in his paper, Lomb fits the unevenly signal by sine waves to estimate the frequency component. More precisely, if $x(t_i)$ ($i = \{1, \dots, N\}$) are the N measurements of the signal at times t_i , unevenly spaced, each sample $x(t_i)$ is fitted by a function x_f

$$x_f(t_i) = A \cos(wt_i) + B \sin(wt_i).$$

The mean square differences between the model and the samples of the signal is given by

$$E(\omega) = \sum_{i=1}^N [x(t_i) - x_f(t_i)]^2.$$

The best model is the one with the A and B values that will reduce $E(\omega)$ to its minimum value, called $E_{min}(\omega)$. This last value, $E_{min}(\omega)$, can then be minimized as a function of ω . Scargle rewrites then differently the problem as

$$\Delta E(\omega) = \sum_{i=1}^N [x(t_i)]^2 - \underbrace{\sum_{i=1}^N [x(t_i) - x_f(t_i)]^2}_{=E_{min}(\omega)}, \quad (3.9)$$

and searching the minimum of $E_{min}(\omega)$ is the same as finding the maximum of $\Delta E(\omega)$. Lomb has shown in its paper ([43]) that if the model to fit the sample is modified a little bit with

$$x_f(t_i) = A \cdot \cos(w(t_i - \tau)) + B \cdot \sin(w(t_i - \tau)),$$

with τ such as

$$\sum_{i=1}^N \cos(w(t_i - \tau)) \sin(w(t_i - \tau)) = 0,$$

therefore Eq. 3.9 can be simplified to

$$\Delta E(\omega) = \frac{\left[\sum_{i=1}^N x(t_i) \cos(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \cos(w(t_i - \tau))^2} + \frac{\left[\sum_{i=1}^N x(t_i) \sin(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \sin(w(t_i - \tau))^2}. \quad (3.10)$$

Scargle concluded in his paper that this last equation (3.10) corresponds to the definition of the periodogram of Eq. 3.8. Fitting the unevenly signal with sine waves can thus be used to estimate the periodogram. We also see from Eq. 3.10 that the frequency that maximizes the periodogram will be the same as the one that maximizes $\Delta E(\omega)$.

Equation 3.8 correspond to the case where the unevenly signal has a zero mean and a unit variance. In the general case, the Lomb-Scargle periodogram can be written as

$$P_{LS}(\omega) = \frac{1}{2\sigma^2} \left\{ \frac{\left[\sum_{i=1}^N (x(t_i) - \bar{x}) \cos(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \cos(w(t_i - \tau))^2} + \frac{\left[\sum_{i=1}^N (x(t_i) - \bar{x}) \sin(w(t_i - \tau)) \right]^2}{\sum_{i=1}^N \sin(w(t_i - \tau))^2} \right\}, \quad (3.11)$$

where \bar{x} and σ^2 are respectively the mean and the variance of the unevenly signal $x(t_i)$. Equation 3.11 corresponds to the formula used by the *plomb* function of Matlab.

3.2.4 The respiratory features

In addition to the HRV features, we decided to compute respiratory variability (RV) features. Generally, the respiratory signal is used to detect apnea during sleep. However, some studies also established the link between the respiratory variability and the psychological state of people [44, 45, 46]. For instance, during mental stress, it was shown that the RV increased [44]. The mental stress in [44] was triggered by a competition. More precisely, during the mental stress task, the subjects had to respond the fastest they could to arithmetic tests. The subject who responded the fastest and with the best accuracy received a specific reward. To the contrary, during sustained attention states, the RV decreased. In the same study ([44]), the task requiring sustained attention was also an arithmetic test but with no pressure of time and reward.

In the work presented in this report, subjects had to respond the fastest they could to stimuli occurring at random times during the PVT tests. These tests can be defined as sustained attention tasks as no pressure was put on the subjects (there was neither reward nor penalty if they were too slow). And when subjects became drowsy, it can be assumed that their attention for the task was reduced, which can be interpreted by a variation in the RV. This is why the RV is used in this thesis. Moreover, in the literature, there are only few scientific papers using the respiratory signal directly as a feature to detect the drowsiness.

In the experiment of this thesis conducted at the University of Liège, the respiratory signal was not directly available. Indeed, the subjects did not wear any devices during their PVT tests that would directly record the respiratory signals. Therefore, if we wish to have the respiratory signal, we must derive it from the ECG. This can be done by using the technique of the ECG-Derived Respiration signal (or EDR signal).

The method developed to compute the EDR signal is based on the work performed by Widjaja et al. in [7]. The EDR signal is based on the following principle: when the subject is breathing during the experiment, his/her rib cage inflates and deflates

at the whim of breathing. It means that the distance between the electrodes and the heart is varying according to time. There is thus a modification of the impedance, and the ECG is altered by this modification. Therefore, there are small variations in the amplitude of the ECG and the amplitudes of the R peaks will be also modified in time due to the respiration. By interpolating these variations, the respiratory signal is obtained. In [7], the variations of the R peaks must be compared to the baseline. Therefore, the baseline must be removed from the ECG. We thus take the ECG after the FIR filter of Fig. 3.1 as the baseline is already removed on this signal in order to get the real amplitudes of the R peaks. The locations of the R peaks are taken back from the R peaks detection algorithm previously described in this thesis. Once the amplitudes of the R peaks are found, the signal is interpolated using cubic spline interpolation at 10 Hz. The interpolated signal is then filtered by a low-pass FIR filter with a cut-off frequency of 0.4 Hz. The order of this filter is equal to $N = 48$ and a Hamming window is used, the equation of which is given by

$$W[n] = 0.54 - 0.46 \cdot \cos\left(2\pi \frac{n}{N}\right),$$

where N is the order of the filter.

Then, we find the local extrema in the signal. These extrema corresponds to the inspirations and exhalations. A breath to breath (BB) interval is defined as the time elapsed between two maxima (resp. between two minima). Finally, two additional criteria are used to validate or not the BB intervals [7]:

- in time: a BB interval must at least have a duration of 1500ms to be accepted.
- in amplitude: the amplitude between a pair of successive maximum and minimum must be greater than 15 % of the mean of the previous and the next differences of amplitudes.

Figure 3.14 shows the result of the plot of the BB intervals as a function of time for a subject during a PVT test. From this figure, we can compute the mean duration of the BB intervals which is equal to 4.3 seconds and which corresponds to a respiratory rate of 15 breaths per minute. This result is within the usual range for a human adult which is between 12 and 18 breaths per minute [47].

Once the correct BB intervals are found, the features in the respiratory domain can be computed. The names, units, definitions, and the formulas of these features used in this work are summarized in Table 3.4.

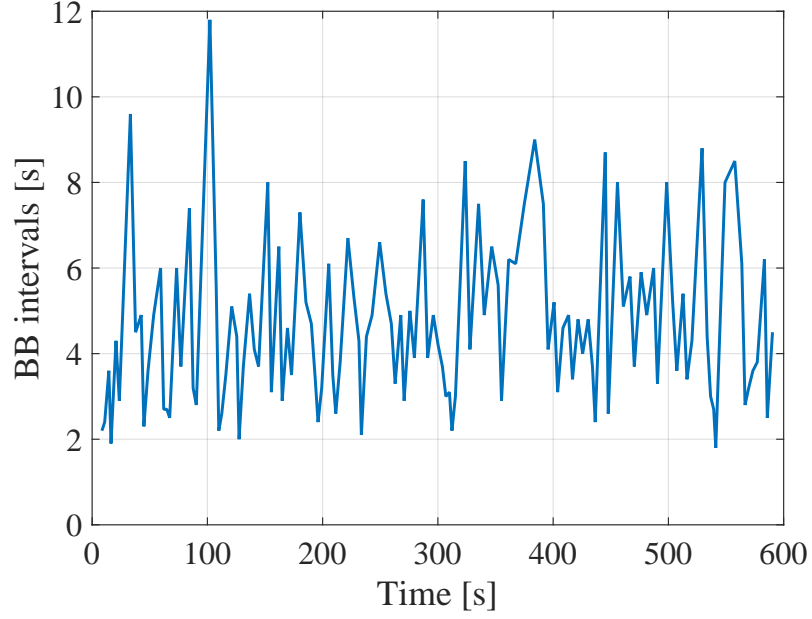


Figure 3.14: BB intervals computed from the ECG-Derived Respiration signal for a subject during a PVT test.

Type	Units	Definition	Formula
mean BB	s	Mean of the BB intervals	$\frac{\sum_{i=1}^{nb_BB} (BB_i)}{nb_BB}$
SDBB	s	Standard deviation (std) of the BB intervals	$\sqrt{var(BB_intervals)}$
diff BB	s	Difference between the largest and the smallest BB interval	$max(BB) - min(BB)$
RMSSD BB	s	Square root of the mean of the squared difference of successive BB intervals	$\sqrt{\frac{\sum_{i=1}^{nb_BB-1} ((BB)_{i+1} - (BB)_i)^2}{nb_BB-1}}$
SDSD	s	Standard deviation (std) of the difference of successive BB intervals (signal notated BB^*)	$\sqrt{var(BB^*)}$

Table 3.4: Table summarizing the names, units, definitions, and the formulas of the features in the respiratory domain.

Chapter 4

Machine learning

The previous chapter was devoted to the signal processing of the ECG and the computation of the features. Once these features are computed, the algorithm for the machine learning part of this thesis can be implemented in order to predict the drowsiness state of people. First, this chapter explains the logic and the functioning of a machine learning model. Second, the reference method used to determine if someone is awake or drowsy will be detailed. Third, we explain the method that is implemented in this thesis. We also describe in this third part the performance measures that can be used in order to assess the reliability of the drowsiness monitoring system developed.

4.1 Introduction and theory about machine learning

The machine learning field can be seen as a sector that develops algorithms in order to enable machines or computers to learn. In a machine learning problem, there are always observations on the input side of the model. These observations are also called samples. For each observation, there is a number k of features. The features can be of really different natures. For instance, in a medical application, a feature can be the weight of a person represented as a real number. Another feature could be the sex of the person represented as a class. This last feature can only take two possibilities, male or female, and can be represented in different ways. For instance the number 1 could be used for a male and 0 for a female. Another possibility is the use of the letters M or F.

The goal of the machine learning model is to learn properties from the features of the observations that the model has at its disposition. Once these properties are learned, the algorithm receives unknown observations at its input and tries to predict as accurately as possible the properties of these unknown observations based on what it has previously learned [48].

More generally, the machine learning problems can be separated in two main domains:

- Supervised learning: each observation at the input of the training is associated with one or several outputs. The algorithm will therefore try to find the

function predicting the output. If this function is called f , the output y can be simply written as

$$y = f(x_1, x_2, \dots, x_k)$$

where x_1, x_2, \dots, x_k are the k features of the current unknown observation.

- **Unsupervised learning:** the model tries to find specific patterns in the data without looking at input-output relations. Specific fields exist like clustering, where the algorithm tries to find families of data. There is also the density estimation for which the distribution of the observations in the input space is estimated. Finally, there is the dimensionality reduction. This process consists of projecting data from a high dimensional space to a lower one (generally 2D or 3D) in order to be able to visualize the data more easily.

In this thesis, the supervised learning technique is used. Indeed, for each observation, there is a known output value which is the state of the subject in our case. Once the model has been trained, it will try to predict these states from the features of unknown data. The supervised learning technique can be separated in 2 parts:

1. **Regression:** the output is one or several numerical value(s).
2. **Classification:** the output is a class. The class of the output can only belong to a specific set of possible classes. If the output can only belong to 2 classes, this problem is called binary classification.

The main difference between classification and regression is that for classification the output is a discrete value in a given set of possibilities whereas, for regression, the output is continuous. In this thesis, a choice had to be made between these 2 possibilities. The choice made and the reasons that justify this choice will be explained later in this chapter.

4.2 The output of the model

4.2.1 The KDS score

As previously explained in Chapter 2, each subject of the experiment performed three PVT tests at three different moments of the day (Fig. 2.8). Each PVT test lasts 10 minutes. During each test, different physiological signals were recorded: EEG, EOG and ECG. In the literature, the reference to assess the drowsiness state of people is the EEG and EOG. The analysis of these two physiological signals were done by three people at the University of Liège before this thesis: the Doctor Robert Poirrier, François Clémentine, a PhD candidate at the University of Liège in the drowsiness field and Baudouin Fortemps de Loneux, during his master thesis [49]. Based on this analysis, the level of drowsiness for each subject at different moments during the PVT tests is known and we can give a score from 0 to 10 to on the *Karolinska drowsiness scale* of the drowsy state of the subject [11, 12]. The lower the score, the more awake and alert the subject is.

Table 4.1 shows an example of how the assessment of the level of drowsiness of a subject during the PVT test is done. In this table the subject is the subject number 8 during his third PVT test (called session 7 in Table 4.1). As already mentioned,

each PVT lasts 10 minutes. These 10 minutes are split in 30 windows of 20 seconds. Each window of 20 seconds is then separated again in 10 intervals of 2 seconds. For each interval of 2 seconds, the activity of the EEG and the EOG is analyzed.

More precisely, if, during this interval, there is an activity α or θ in the EEG or if there is specific slow eye movements in the EOG, it is indicated (by a letter) in the corresponding box of the Table 4.1, meaning that there is a drowsiness sign for this interval. Having a letter, like α or θ , during an interval can be seen as putting a value equal to 1 for this interval. When two activities are present during the same interval (like for instance for the 7th interval of the 5th window of Table 4.1 where there is at the same time an activity α and θ of the EEG), this case also corresponds to a value equal to 1 for this interval. To the contrary, if there is no activity of the EEG or EOG in the interval, a value of 0 is simply put for that interval. The global score for the window of 20 seconds is then simply obtained by summing the different values of the 10 intervals of 2 seconds. By doing so, the level of drowsiness of a person is a discrete number comprised between 0 and 10. This discrete number is called the KDS score [11, 12]. The acronym KDS stands for *Karolinska drowsiness scale*. The lower the score, the more awake and alert the subject is. When the KDS score increases, the subject becomes more drowsy. For instance, in the example of Table 4.1, the subject has big KDS scores, which seems logical as the subject was performing this test after roughly 28 hours of sleep deprivation.

As explained in Chapter 3, the duration of the time window from which the ECG features are computed is equal to 120 seconds. The moving window technique of Fig. 3.11 was also explained in Chapter 3. As the KDS score is available every 20 seconds, a technique must be used to combine the different scores of the six intervals of 20 seconds in order to have the global score of the window of 120 seconds. We decided to deal with the maximum score of the different windows. The main problem of the mean of several intervals is that it smooths the KDS score of the window of 120 seconds. Therefore, the resulting score could be low if a large score is surrounded by lower ones. For this reason, we keep the maximum score of the 6 intervals.

4.2.2 Choice between regression and classification

We decided to take the supervised learning method for this thesis. Indeed, the output values of the different observations are the KDS scores and the learning model will use these outputs. However, as said in the introduction section of the present chapter, a supervised learning method can be implemented either using regression or classification. A supplementary decision has to be taken to decide if we use the regression or the classification in this thesis.

The final choice was to deal with classification. Indeed, as a first reason, in the literature, this approach is generally the one used. The machine learning model will try to predict if someone is drowsy or not at a given time based on its ECG recorded during the past 120 seconds. In this case, the classification problem seems to be simpler. Indeed, compared to the regression, the binary classification can be done more easily. A given threshold is put on the KDS score before training the model. In the literature, the threshold is normally fixed at 5 on a scale from 0 to 10. However, in Anund et al. [12]; big decrements of performances were observed for lower values than 5. Moreover, if the reaction time to the stimuli is observed, it can be seen that there is already an important rise of the reaction time for KDS

Protocol	PVT	KDS Scoring Sheet	Legend									
Subject	8	Scorer	alpha	α	spindle	s						
Session	7	Scoring date	theta	θ	vertex	v						
Starting time	11:12:38		eye	e	K-complex	k						
Epoch	Time	Score	1	2	3	4	5	6	7	8	9	10
1	11:12:38											
2	11:12:58	5	e	e	e			θ		e		
3	11:13:18	4	e						e	e		e
4	11:13:38	6	e		e	e	e	e	e			
5	11:13:58	4				e			$e\theta$	e		e
6	11:14:18	7		e	e	$e\theta$	e		e		e	e
7	11:14:38	5				$e\theta$		$e\theta$	e	e	e	
8	11:14:58	2					e			e		
9	11:15:18	8	e	e		e		e	e	e	e	e
10	11:15:38	4			e		θ		e	e		
11	11:15:58	5		e	θ	e	e	e				
12	11:16:18	6	e	e	e	e	e	e				
13	11:16:38	1								e		
14	11:16:58	8		e	θ	e		e	e	$e\theta$	e	e
15	11:17:18	5	e	e				e		e	e	
16	11:17:38	4			e	e	e	e				
17	11:17:58	3				e	e				e	
18	11:18:18	3					e	e	e			
19	11:18:38	3				e			e		e	
20	11:18:58	0										
21	11:19:18	0										
22	11:19:38	2			e	e						
23	11:19:58	3								e	e	e
24	11:20:18	4			e	e	e					e
25	11:20:38	1	e									
26	11:20:58	3						e	e	e		
27	11:21:18	4						e	e	e	e	
28	11:21:38	2								e	$e\theta$	
29	11:21:58	3	e	e							θ	
30	11:22:18	3				e	e					e

Table 4.1: Example of the scoring sheet used to assess the level of drowsiness of a subject during a PVT test on the Karolinska drowsiness scale.

scores lower than 5 compared to the windows for which the subjects have a KDS score of 1 or 2. Therefore, it was decided to choose a threshold of 4. It means that if someone has a KDS score below 4, this person is considered as awake. If the person has a score equal or larger than 4, this person is considered as drowsy. Therefore, the output of the model is a class that can only take two values: 1 if the person is drowsy or 0 if not.

As previously said in this thesis, the feasibility or not of the ECG to detect drowsiness is still not proven. Therefore, another reason for which the binary classification is that it is not sure that this system will be precise enough to discriminate the different states on the KDS scale. That is why we decided to see whether the ECG can be used to more generally discriminate between the 2 classes: drowsy or awake.

Finally, another reason for which classification is used instead regression is that the regression approach would lead to a continuous KDS score at the output of the machine learning model. A threshold on this output should then be used to obtain at the end the state of the person (awake or drowsy). However, this threshold must be adapted with the training set. Indeed, if there are more awake states in the training set, the model will probably tend to predict lower KDS scores. Therefore, the threshold to find if someone is drowsy or not will not be the same as the one in the literature. It also means that this optimal threshold should be found. Therefore, as regression requires to find the optimal continuous threshold to find if someone is drowsy or not, the classification method was chosen by simplicity.

4.3 Machine learning model developed

4.3.1 The final dataset

The approach to develop the supervised learning method in binary classification must be developed. All the subject for which the KDS score was available were taken for the development of the model (four subjects did not get the KDS score and were therefore not considered). Therefore, the data set comprises 21 subjects. For each subject, there are 25 observations for each PVT test, consisting of 20 features on the input side and the binary class on the output side. However, if we observe the dataset with more details, we can notice that some subjects have strange behaviors. Some subjects resulted with no drowsy events during the three tests. On the other hand, others were considered as drowsy for all the three PVT tests. These subjects were thus removed from the dataset. The reason is that when the algorithm is tested, evaluations parameters are used. These parameters, which will be explained later in the present chapter, require to have unknown observations from both classes. Moreover, a problem with the acquisition of the ECG signal was also noticed for a male subject of the experiment. Figure 4.1 shows the ECG recorded during this acquisition problem. One can see in this figure that the ECG signal is full of noise and no useful information can be derived from this signal. The reason is that the electrodes were not correctly placed on the skin of the subject when the subject had lots of hairs on his chest. At the end, after removing these "curious subjects", there remains 16 subjects inside the dataset.

One must divide the remaining data set in two parts. One is the learning set (LS). They are the data used by the model to learn. The other part is the test set

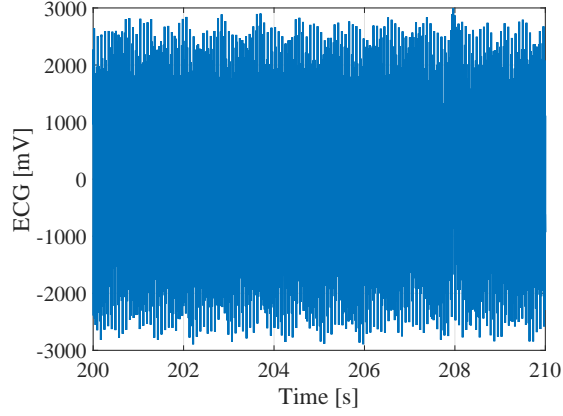


Figure 4.1: Example of a male subject that was removed from the dataset due to an acquisition problem. The ECG is totally corrupted and full of noise. No useful information can be derived from a signal with so much noise.

(TS). They are the unknown data for the model. Once it has been trained, the model receives only the features of each observation in the TS. Based on what it learned with the LS, it predicts the output of the unknown data. By comparing the estimated outputs to the true ones in the TS, it is possible to assess the reliability of the model. However, the main problem of this method is that the model could be specialized to correctly fit the given TS. If a totally new TS is given to the model, the results could be worse than expected. This phenomenon is called overfitting. The next section will explain a method that allows to avoid this problem.

4.3.2 The cross-validation

In the previous section, the importance of avoiding a specialization on a single LS-TS case was highlighted. Moreover, we also need to pay attention to avoid putting observations of the same subjects in the TS and the LS. Indeed, if the LS contains already have several observations of a given subject in its LS, when others observations on that subjects will be tested, it will have less difficulty to predict the level of drowsiness. Therefore, the performance of the model will be overestimated. When the model will be really used on a totally new subject, the results will probably be less as expected.

Therefore, we use the *k-fold* cross-validation to assess the model. This technique consists of splitting the dataset in k parts that are called the folds [50]. Then, the following procedure is repeated k times: the model is learned on the whole dataset except the k^{th} part. This k^{th} part is used for the test. The error on each fold is then averaged to have a global assessment of the model. Here, the *k-fold* cross-validation is a little bit different to take into account the requirement that different subjects must be used for the learning and the testing parts. Therefore, the implemented procedure is the following: all the subjects except 2 are used for the TS. The 2 remaining subjects are placed in the TS. The model is trained on the LS and assessed with the TS. This process is repeated k times until the cross-validation is finished. The number of folds is therefore equal to 8 and there are 2 subjects by fold.

4.3.3 Different estimators used to build the model

Different types of machine learning estimators exist in order to build a model. For this work, two main estimators were chosen. The first one is the support vector machine (SVM). This method offers several advantages [48]. First of all, SVM is memory effective. It means that it only keeps specific points from the LS, called support vectors, for the decision function. Moreover, this technique is generally efficient in high dimensional spaces by using a specific method called the kernel trick [50]. Finally, it is possible to use easily different kernel functions (linear, gaussian, polynomial, ...). In this thesis the linear kernel and the Gaussian Radial Basis Function (rbf) kernel were used. [48].

The second estimator tried is the random forest (RF). Its name really indicates what it is. This estimator consists of a series of trees. A tree is one of the most representative machine learning technique. It consists of a series of nodes. At each node, a criterion separates the data in two parts as a function of a specific criterion. At the end, it remains leaves of the tree. Each leaf corresponds to a specific value of the output. If there are k trees in the forest, k different learning samples will be drawn from the original one by using the bagging technique (= sampling with replacement). Each tree of the forest thus predicts the class corresponding to the unknown observations. The final class chosen is the one that gets the majority of votes among the forest. The random forest belongs to the ensemble of methods family [50, 48]. In fact, each tree is built on a bootstrap samples of the LS (by using the sampling with replacement). By building the trees in different bootstrap LS, the variance as a function of the LS is reduced. However, the precision of the model built with a RF estimator is a bit reduced due to the fact that information in the different LS is smaller than in the original LS due to the sampling with replacement.

Others general estimators like the nearest neighbor or the neural network also exist in the literature. However, these estimators were not used for this thesis. The reason depends on the technique. For the nearest neighbor, it is very sensitive to noise. Therefore this estimator was avoided here as it is suspected that there is a consequent number of noise in the dataset. For the neural network method, the network is quite complicated to build, and this method was thus no used in this thesis.

4.3.4 Performance measures

To assess the precision of a machine learning model in binary classification, the confusion matrix (also called the contingency matrix) is a nice tool to summarize the results. This matrix takes the form of Table 4.2. The positive class corresponds to a drowsy event in this thesis. It is the opposite for the negative class. Based on this matrix, we can defined four important properties:

1. Sensitivity (Se): $\frac{TP}{TP+FN}$. It corresponds to the percentage of correct true predictions (drowsy estimations) in all the true classes of the TS. The sensitivity can also be named the recall (R).
2. Specificity (Spe): $\frac{TN}{TN+FP}$. It corresponds to the percentage of correct false predictions (awake estimations) in all the false classes of the TS.

3. Positive predicted value (PPV): $\frac{TP}{TP+FP}$. It represents the proportion of true positive predictions among all the positive predictions of the model. This parameter is often named the precision (Pr).
4. Negative predicted value (NPV): $\frac{TN}{TN+FN}$. It represents the proportion of true negative predictions among all the negative predictions of the model.

Ideally, the number of FP and FN is equal to zero. Therefore, all the four parameters are equal to 1 for the ideal case. If the number of FN and FP increases, these parameters decrease and the model is less precise. Note that the number of FP corresponds to the number of false alarms. The number of FN corresponds to the number of times we predict a subject as awake whereas his true state is drowsy.

One additional parameter can be added to the four ones previously described. This additional parameter is the accuracy and is equal to:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

If we know the accuracy, the error rate can be easily derived as

$$\text{error rate} = 1 - \text{accuracy}.$$

		True (known) values	
		Positive	Negative
Predicted values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Table 4.2: Example of a confusion matrix used in binary classification.

Generally, in binary classification, the output of the model is a class probability. Then a specific threshold is set. If the class probability is above this threshold, the observation belongs to this class. If it is not the case, in binary classification, it belongs to the other class. Let's take an example (in binary classification). If the class probability to belong to the first class is 0.3 and the threshold is set to 0.5, this observation will be predicted as belonging to the second class as its class probability for the first class is below the threshold.

For a specific threshold, the confusion matrix as in Table 4.2 can be built. However, this threshold can also be modified to impact the sensitivity and specificity. This principle is used to build a ROC curve (receiver operating characteristic curve). A ROC curve is a plot representing the sensitivity as a function of one minus the specificity (1-specificity) for different thresholds values. Figure 4.2 shows an example of a ROC curve. In a ROC curve, the optimal point is the one in the upper left corner. Indeed, this point reaches a sensitivity of 1 and a specificity of 1. Based on the ROC curve, the optimal threshold value can be found in order to maximize the sensitivity and specificity. Another performance measure that can be derived from the ROC curve is the area under the curve. The higher this area is, the better it is. The diagonal dashed line in Fig. 4.2 corresponds to the result of a random classification (also named random result in this thesis). If the ROC curve is very close to this diagonal, it means that the model has nearly the same results as the random predictor and did not learn useful information during the training phase to

be able to correctly predict the new data in the TS. However, if the model is far away below this diagonal line, it does not mean that the model is not good. It is just that the predicted output must be inverted in order to have the real class. Note that in Fig. 4.2 the ROC curve is not a continuous curved line. It comes from the fact that only a discrete number of threshold values were used.

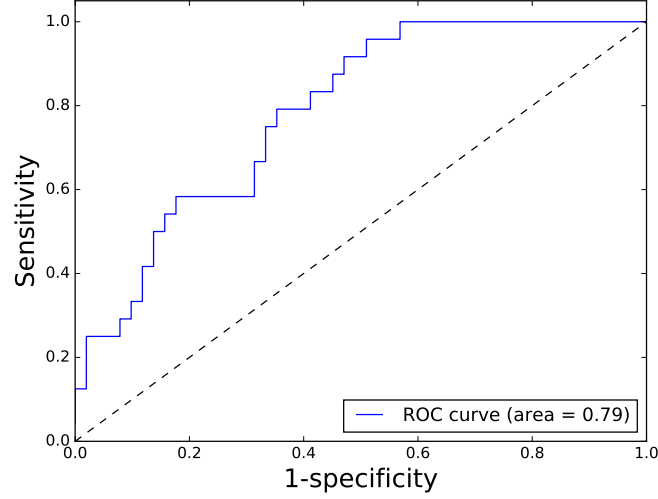


Figure 4.2: Example of a ROC curve. The diagonal dashed line corresponds to the random result.

4.4 Conclusion of the chapter

In this chapter, the procedure of the cross-validation without mixing the different subjects in the folds has been chosen and detailed. Moreover, we also explained how the two possible classes at the output of the model (drowsy or awake) were obtained. Then the different performances measures that can be used to evaluate the model were explained. Based on the features computed in Chapter 3 and all the procedure and performance measures detailed in the present chapter, the automatic model in order to detect the drowsiness can be developed in order to see if it is possible or not to detect drowsiness based on the ECG. The results obtained will be explained in the next chapter.

Chapter 5

Results obtained and discussion

In the previous chapter, we detailed the the procedure of the machine learning approach developed in this thesis. We also explained the performance measures that enable the evaluation of the model developed. In the present chapter, this procedure will be tested on the data from the subjects. The results obtained will be analyzed. Some tests performed in order to enhance the quality of the model will also be explained in this chapter. Moreover, a discussion of the results obtained will also be conducted through this chapter. Finally, some perspectives in order to improve these results will be given.

5.1 Balanced and normalized dataset

For the first test, all the features computed after the processing of the ECG and explained during the Chapter 3 are used. The dataset used is the one with the 16 subjects as explained in Chapter 4. The three models with the three estimators used in this thesis (the random forest and the SVM using the linear and Gaussian kernels) are tested on this dataset with the default parameters of the estimators. The performance measures are then obtained during a cross-validation. This cross-validation consists of 8 folds. In each fold, there are 2 different subjects. The model is therefore trained on 7 folds (14 subjects) and tested on the remaining fold (2 subjects). This operation is repeated 8 times in order to have each fold used once as TS. Therefore the same subjects are never used for the training and testing phase.

The first block of Table 5.1, called *No balancing*, shows the results obtained during a cross-validation for the 3 models. The same cross-validation is used for the 3 different models. As can be observed, the results of this solution are not good. In fact, the model nearly always predicts an unknown observation as awake. The sensitivity is very low as the model does not recognize drowsy events in the TS.

However, if the data collected during the experiment are analyzed in more detail, we can observe that the number of observations labeled as drowsy events is much smaller than the number of awake events. Indeed, on average during the cross-validation, only 30% of the LS is composed of drowsy events. To tackle this issue, the LS is adapted to be balanced. More precisely, in each LS, the number of observations is equal to the number of moving windows of 120 seconds in the three PVT tests. As each moving window is separated by 20 seconds (Fig. 3.11), there are thus 75

windows during the three PVT tests per subject. As there are 14 subjects in each LS, this gives 1050 input-output relations in the train. The number of outputs equal to 1 (drowsy events) is then computed. Then, the same number of awake events (outputs equal to 0) are randomly picked from the LS. By doing so, the LS is balanced between the two classes and the model will be less prone to overestimate awake events. The main drawback of this solution is that the size of the dataset is reduced.

	Estimator	Se	Spe	PPV	NPV
No balancing	SVM: linear	0.126	0.902	0.318	0.705
	SVM: Gaussian	0.139	0.918	0.272	0.700
	Random forest	0.366	0.844	0.458	0.753
Balancing	SVM: linear	0.509	0.571	0.359	0.718
	SVM: Gaussian	0.278	0.718	0.249	0.688
	Random forest	0.456	0.752	0.452	0.762
Balancing + Normalization	SVM: linear	0.488	0.605	0.369	0.743
	SVM: Gaussian	0.447	0.650	0.418	0.760
	Random forest	0.423	0.791	0.471	0.765

Table 5.1: Results of the sensitivity (*Se*), specificity (*Spe*), positive predicted value (*PPV*), and the negative predicted value (*NPV*) for the different machine learning models built with the 3 estimators used with their default parameters. The first block, called *No balancing*, consists of the raw dataset. The second block, called *Balancing*, consists of the balanced dataset. The third block, called *Balancing + Normalization*, consists of the balanced and normalized dataset.

This balanced dataset was then tested. The sensitivity (*Se*), specificity (*Spe*), positive predicted value (*PPV*) and the negative predicted value (*NPV*) were assessed for the 3 models. The second block of Table 5.1, called *Balancing*, shows the improvement when the balancing is done.

However, even with the balancing, the model built on the SVM estimator using the Gaussian kernel provided still low results. This comes from the fact that this estimator is not a scale invariant method [48]. Moreover, in this work, the order of magnitudes of the different features is really different from one feature to another (for instance the *meanHR* (60 to 80 beats per minute) compared to the *LF_f_peak* ([0.04; 0.15] Hz). Therefore, the features were normalized. The mean and the variance of each feature of the LS are computed. Then each feature of the LS and the TS are normalized by using the z-normalization

$$X_z = \frac{X - \bar{X}}{\sigma_X}$$

where X_z , X , \bar{X} , and σ_X are respectively the z-normalized feature, the original value of the feature, the mean, and the standard deviation of that feature for all the subjects in the LS.

Once the dataset is balanced and the features are normalized, the *Se*, *Spe*, *PPV*, and the *NPV* can be re-computed. The results are shown in the third block of Table. 5.1 named *Balancing + Normalization*. From these results, it can be seen

that the features normalization is only important for the model using the Gaussian kernel.

From Table. 5.1, it can also be observed that the results obtained depend on the estimator used. On the other hand, even for the same estimator, there exist parameters that can be modified (for instance the number of trees in the forest) and that can influence the results. Therefore, each model was optimized by searching for these optimal parameters. To do so, parameters were tuned by hand and the ones providing the best results were kept. More precisely, another cross-validation was performed. The performance measures were obtained for each fold. Then, they were averaged over the different folds.

For the random forest, the best parameters were: a number of trees of 150, in which the number of features looked at each node to find the one leading to the best split is equal to the square root of the number of the features, and finally the minimum number of samples to split a node is equal to 9.

For the SVM estimator using a Gaussian kernel, two parameters are important to tune. The first one is the penalty parameter C . This parameter is a trade-off between the complexity of the model and the miss classification weight of the LS. If this parameter is increased, the model tends to correctly classify more and more LS cases and increases its number of support vectors [48]. The other important parameter is the γ parameter that reflects the importance of a single observation of the LS. The best parameters for the Gaussian kernel are the following: a penalty parameter C of 25 and a kernel coefficient γ of 0.01.

Finally, for the SVM estimator using the linear kernel, the best parameter found was a penalty parameter of 5.

Table 5.2 shows the four performance measures of the models with these parameters. The results are obtained by taking the mean of the performance measures of a cross-validation. In this table, the class probability to say if a TS observation belongs to the first class or not is the default one ($= 0.5$). From Table 5.2, it can be seen that the specificity is the parameter that increases the most compared to the random case. However, the increase of the sensitivity is lower. All the three models have therefore more difficulty to find the drowsy events compared to the awakes ones.

Also, it can be noticed from Table 5.2 that the best model among the three seems to be the one using the SVM estimator with the linear kernel.

	SVM: Gaus- sian kernel	SVM: linear kernel	Random forest	Random case = luck
Se	0.534	0.593	0.490	0.490
Spe	0.640	0.709	0.687	0.508
PPV	0.391	0.522	0.389	0.300
NPV	0.759	0.801	0.767	0.695

Table 5.2: Results of the sensitivity (Se), specificity (Spe), positive predicted value (PPV), and the negative predicted value (NPV) between the different machine learning models for the balanced LS. The estimators of the models are used with their best parameters found.

The results from Table 5.2 give some general information of the performance of the models. As previously said, the ROC curve is another tool that can be used.

Therefore, the ROC curve can also be drawn during the cross-validation. Figure 5.1 illustrates this ROC curve. In this figure, each color corresponds to a fold. The black dotted line is the mean of the ROC curve of the 8 different folds. From Fig. 5.1, the first thing that can be noticed is that for the 3 models, there are subjects that provide good results; and others for which the model is less good and makes wrong predictions more often. From this figure, it also seems that the model using the SVM estimator with the linear kernel seems better as it shifts nearly all the folds up.

However, results of Fig. 5.1 are for a specific cross-validation. Several different cross-validation have thus been tested in order to see if there were or not lots of variations in the ROC curves in function of the different subjects placed in the folds. Figure 5.2 represents thus the 3 ROC curves for the 3 models for another cross-validation. From this Figure, it can still be observed that the different models provide good results for some subjects and poor results for others. Moreover, if Fig. 5.2 is compared to Fig. 5.1, it can be observed that there are more pairs of subjects that provide low results in Fig. 5.2.

Figure 5.1 and Fig. 5.2 show the ROC curves during 2 different cross-validations. However, it was shown that, even after a cross-validation, there exist variations from one mean ROC curve to another one due to the fact that it is not the same pair of subjects that are used for the different folds. Indeed, the subjects must be randomly shuffled before creating the different folds for the cross-validation. To avoid this variability, all the combinations of 2 subjects among the dataset ($C_{16}^2 = 120$ possible combinations) must be taken. As this number is not so important in this case, all these combinations have been computed.

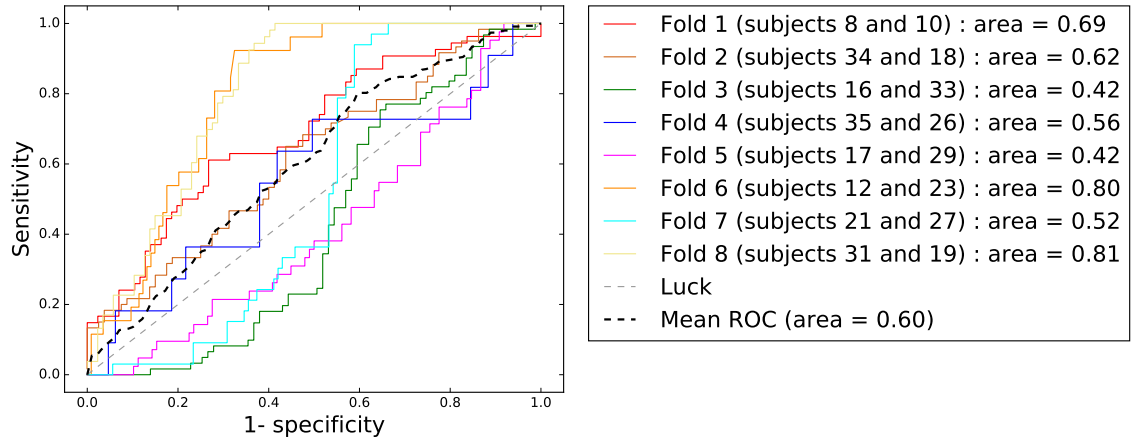
The mean accuracy on all these combinations for the three models is around 60% for each model. The detail of the sensitivity, specificity, positive and negative predicted values is given in Table 5.3.

Estimator	Se	Spe	PPV	NPV
SVM: linear	0.536	0.657	0.414	0.769
SVM: Gaussian	0.526	0.643	0.390	0.760
Random forest	0.466	0.671	0.382	0.757

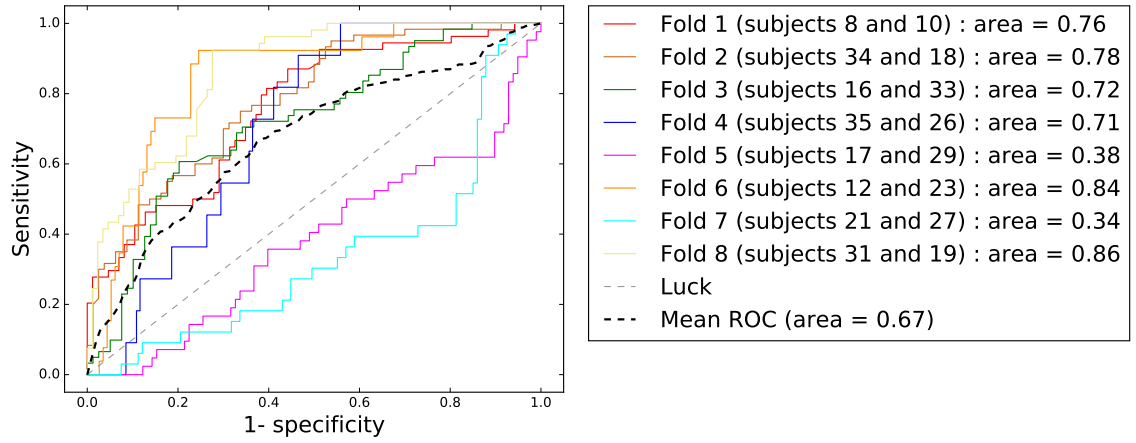
Table 5.3: Results of the sensitivity (Se), specificity (Spe), positive predicted value (PPV), and the negative predicted value (NPV) between the different machine learning models on all the possible combinations. The estimators of the models are used with their best parameters found.

The mean ROC curve of each model can also be drawn from all these combinations as one can see in Fig. 5.3. From this figure, it can be noticed that the 3 models have more or less the same results on all the combinations of the subjects of the experiment. However, the model built on the SVM estimator with a linear kernel is a little bit better than the 2 other models. The best points of each curve are respectively:

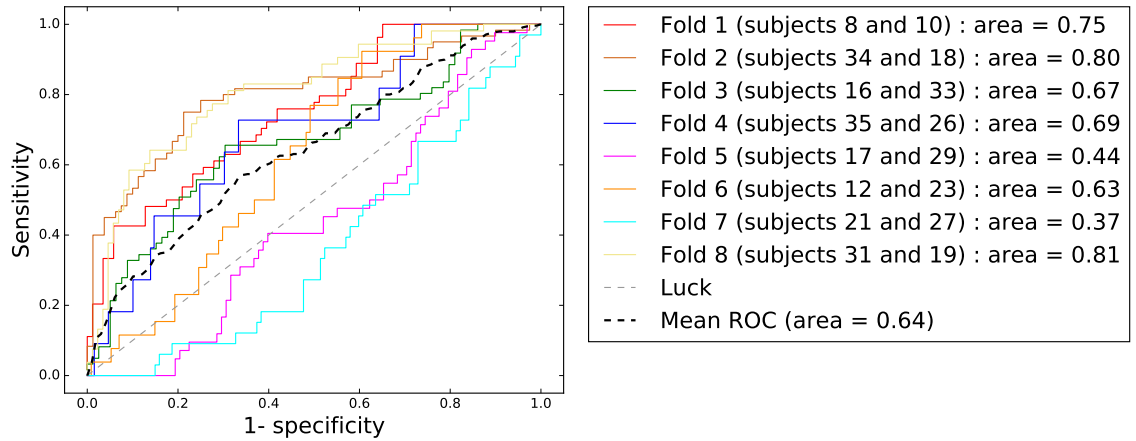
- SVM with Gaussian kernel: Se=0.61, Spe=0.63, for a class probability of 0.46
- Random forest: Se=0.65, Spe=0.63, for a class probability of 0.43
- SVM with linear kernel: Se=0.64, Spe=0.65, for a class probability of 0.49



(a) SVM: Gaussian kernel

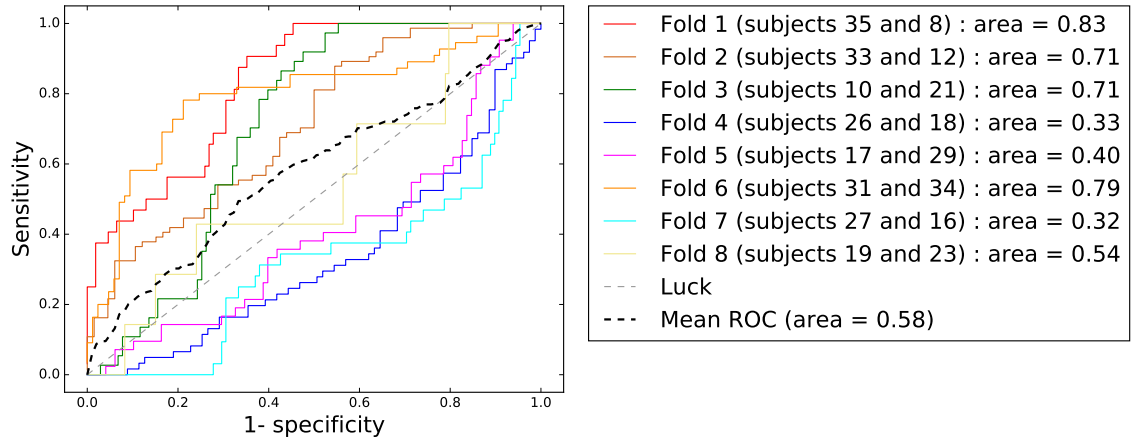


(b) SVM: linear kernel

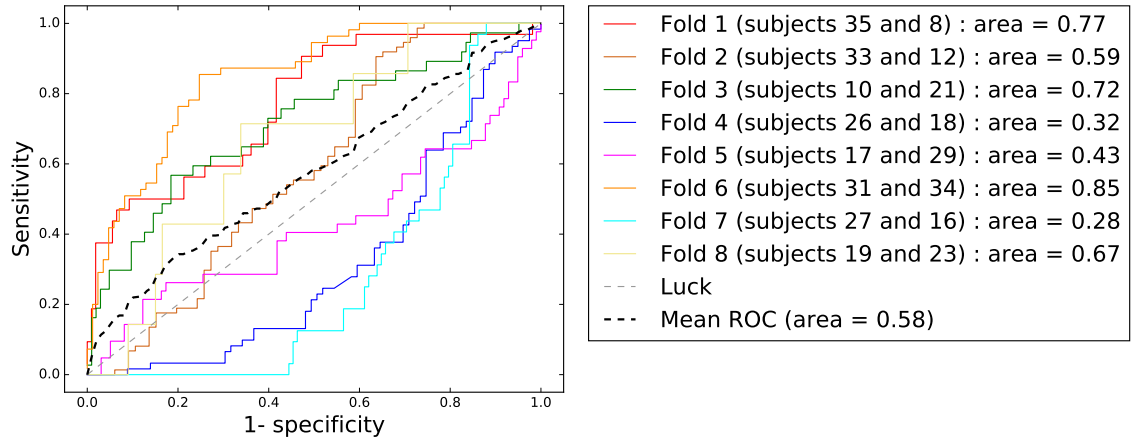


(c) Random forest

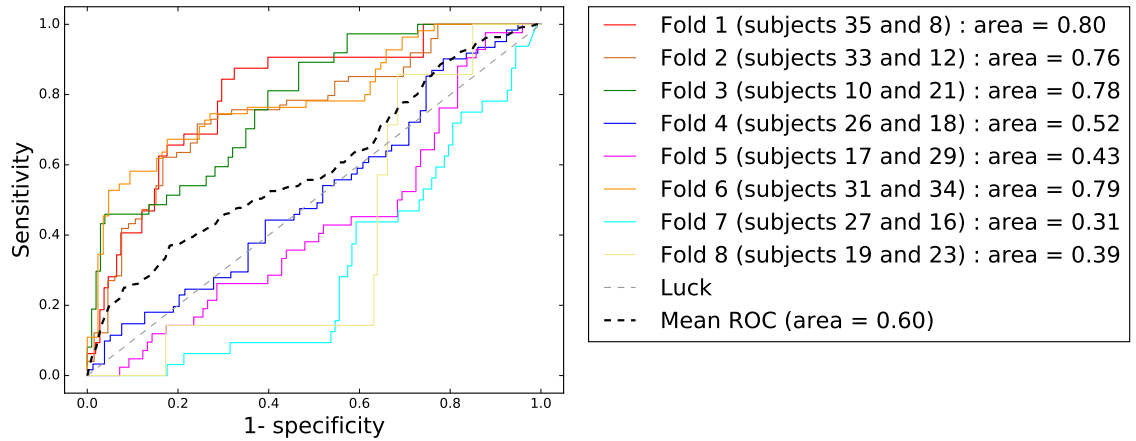
Figure 5.1: The three ROC curves obtained during a cross-validation. Each model is used with all the features and the optimal parameters of the estimators are used. The same cross-validation is used for the 3 models. The area under the ROC curve is also indicated. The gray diagonal dotted line corresponds to random predictions.



(a) SVM: Gaussian kernel



(b) SVM: linear kernel



(c) Random forest

Figure 5.2: The three ROC curves as in Fig. 5.1 during a cross-validation for another set of folds. Each model is used with all the features and the optimal parameters of the estimators are used. The same cross-validation is used for the 3 models. The area under the ROC curve is also indicated. The gray diagonal dotted line corresponds to random predictions.

Note that these values are a bit different from the ones that would be obtained by looking at the optimal point of each curve of Fig. 5.3. Indeed, these values are obtained by computing the mean of the optimum point of each ROC curve obtained for a possible combination which is different as taking the optimum point of the mean of the ROC curves of all the combinations, as shown in Fig. 5.3.

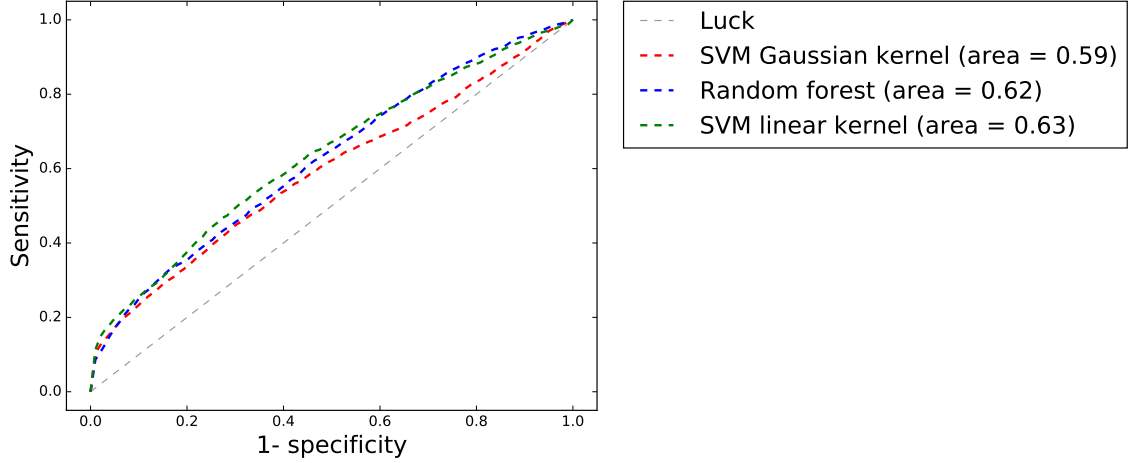


Figure 5.3: Mean ROC curves for the 3 models obtained on all the combinations of the subjects of the experiment. Each color corresponds to a model using a specific estimator. The area under each ROC curve is also indicated. The gray diagonal dotted line corresponds to the random predictions called luck in the legend.

5.2 Discussion of the results obtained for the normalized dataset

From the results of Fig. 5.3, one can conclude that for each of the 3 models, the results obtained are better than for the random case. However, even if these results are better than the random case, the mean accuracy obtained on all the combinations of the subjects was equal to roughly 60%. This result is not really large. Therefore, the use of the ECG to predict if someone is drowsy or not seems to be not precise enough.

In fact, the main problem is that there exists important variations between the subjects. This phenomenon is called subject variability. To illustrate this problem, an inspection of all the features of all the subjects for different KDS scores has been done. More precisely, each feature of the whole dataset (the 16 subjects for the 3 PVT tests) was taken one by one. Each of the features was then separated in function of the different KDS scores. Then, for each feature at a given KDS score, the mean and the standard deviation were computed and plotted for the 16 subjects grouped together.

Fig. A.1 of the Appendix A shows these 20 plots for all the features as a function of the KDS score. In this figure, the mean is represented by the blue dotted line. The standard deviation is represented by the vertical bar. A vertical bar has a length equal to two times the standard deviation (one std from each side of the mean). This vertical bar shows the dispersion of the values of the feature for the same KDS score between the different subjects. They are therefore also called the error bar. From

the 20 plots in Fig. A.1, one can directly observe that the error bars are important for all the features. Therefore, there is an important variability between the subjects for the same feature at the same KDS score. This variability between subjects is one of the main issue and difficulty to predict the drowsiness from the ECG.

We tried different techniques in order to compensate this variability between the subjects:

1. The first technique is also based on the z-normalization. During the first 300 seconds of the first PVT test of each subject, the mean and the standard deviation of each feature is computed. Then, for each time interval, the feature is z-normalized using the corresponding mean and standard deviation. This technique is called *z-baseline* in this thesis.
2. The second technique removes the mean of each feature. More precisely, the mean of each feature is computed during the first 300 seconds of the first test for each subject. Then, the corresponding mean is simply subtracted from all the features of the corresponding subject computed for all the others moments. This technique is called *sub-baseline* in this thesis.
3. The third technique divides each feature by its variance. To do so, the variance of each feature is computed during the first 300 seconds of the first test for each subject. Then, for each subject, each feature is divided by this variance. This last technique is called *div-var-baseline* in this thesis.

These 3 techniques were then used to recompute the different features. To do so, the 3 models were tested with the same optimal parameters as before on these new LS. The performance measures of each model were therefore computed on all the combinations of 2 subjects used in the TS and 14 in the LS. Table 5.4 presents the results. In this table, the default class probability threshold is used (0.5). The last block of Table 5.4, called *no-compensation*, corresponds to the previous case where no variability compensation method between the subjects was used. From this Table, it can be seen that none of the variability compensation method between subjects improves the results. Therefore, the first case with no compensation method was kept as the best one even if it was not able to remove this variability between the subjects.

5.3 Feature selection

In Chapter 3, the 20 features used in this thesis were explained. However, it is possible that some features are less significant than others to determine if a person is awake or drowsy. For instance, it has been decided to incorporate features from the respiratory domain by using the EDR method. However, few studies have used this signal to detect drowsiness. Therefore, some features computed in Chapter 3 may be useless. Some estimators like the random forest are quite robust to the presence of irrelevant features. However, it is better to try to find the best subset of features that can be used for our applications for three reasons:

- The first reason is that if some features are not relevant, there is no need to compute them, so this saves computing resources during the computation of the features.

	Estimator	Se	Spe	PPV	NPV
z-baseline	SVM: linear	0.408	0.585	0.282	0.690
	SVM: Gaussian	0.412	0.623	0.312	0.706
	Random Forest	0.303	0.658	0.260	0.682
sub-baseline	SVM: linear	0.497	0.693	0.391	0.746
	SVM: Gaussian	0.453	0.676	0.385	0.734
	Random Forest	0.461	0.651	0.377	0.731
div-var-baseline	SVM: linear	0.297	0.693	0.225	0.685
	SVM: Gaussian	0.419	0.623	0.244	0.705
	Random Forest	0.249	0.840	0.301	0.736
no compensation	SVM: linear	0.536	0.657	0.414	0.769
	SVM: Gaussian	0.526	0.643	0.390	0.760
	Random Forest	0.466	0.671	0.382	0.757

Table 5.4: Comparison of the Sensitivity (Se), Specificity (Spe), Positive Predicted Value (PPV) and the Negative Predicted Value (NPV) for four cases. The three first cases (z-baseline, sub-baseline, and div-var-baseline) attempts to remove the variability between the subjects. The fourth case (no compensation) is the case with no variability compensation method.

- The second reason is that by incorporating irrelevant features, the number of features in the model increases as its complexity and computing time.
- Finally, the third reason is that these irrelevant features can be seen as noise incorporated in the model.

The model on which this feature selection is done is the one using SVM with the linear kernel which is the best one of Fig. 5.3.

5.3.1 Use of the coefficient of correlation to select the features

A first simple approach to find a good subset of features is to use of the coefficient of correlation. The same procedure as the one used to obtain Fig. A.1 of Appendix A was performed. When each feature for all the subjects was obtained, its linear coefficient of correlation as a function of the KDS score was computed. Then, only the features with an absolute coefficient of correlation above 0.25 were kept. Five features were found out of 20. These features are: SDNN, SD1, SD2, LF_power, mean_BB. Then, a comparison of the results obtained by using the SVM estimator with the linear kernel on all the features or only the 5 best ones was performed. These results, obtained on a cross-validation, are presented in Table 5.5. Note that only the SVM model with the linear kernel was used as it was the one with the best results among the 3 tested models (Fig. 5.3). Table 5.5 shows that the linear SVM model increases its sensitivity to the price of a reduction of the specificity when only the 5 features with a good coefficient of correlation are used compared to the case where all the features are used.

	Se	Spe	PPV	NPV	Accuracy
All features	0.471	0.681	0.407	0.758	0.652
5 best features	0.571	0.618	0.385	0.772	0.642

Table 5.5: Comparison of the performance measures obtained for the model using the SVM estimator with the linear kernel. In the first line all the features are used. In the second line only the 5 features found with an absolute coefficient of correlation above 0.25 are used. The performance measures are: the sensitivity (Se), specificity (Spe), positive predicted value (PPV) and the negative predicted value (NPV).

5.3.2 The use of the random forest to rank the features

To select the features, a second approach can be to use the random forest. Indeed, this forest is composed of several trees. Each tree consists of a series of nodes at which the data set is separated in 2 parts. The tree searches at each node the feature that maximizes the class separation. Therefore, features used at the top of the tree are more important than features used at the bottom as they maximize the class separation in a more general case. Based on that principle, a ranking of the features can be obtained. In the random forest, the ranking of the different features among the trees of the forest will be averaged in order to have this global ranking according to the forest. It therefore decreases the variance compared with the use of a single tree [48]. Here, the exact model used for the feature selection is the extra trees. This model is based on the same principle as the forest. The main difference is that instead of searching the best threshold value for each feature (of the subset used at the node to make the split), the extra trees estimator randomly chooses a threshold for each feature. Then, the feature with the best randomly chosen threshold is kept. This property has the consequence of reducing the variance from one set to another one. But it increases the bias a little bit (the precision of the model compared to the optimal one) [48].

To perform this ranking, again a eight folds cross-validation is used. The result over the eight folds is averaged in order to have the global ranking on the cross-validation. Figure 5.4 shows the ranking obtained. From this figure, one can notice that the standard deviation of the variability of the features in the tree is important.

Based on this ranking, different subset of features have been tried on the same cross-validation. The improvement on the ROC curve is shown in Fig. 5.5. In this figure, one can observe that the best subset of features found is the one in green. The features in this subset are: SDNN, SD2, SD1, mean_BB, RMSSD, LF_power, and diff_BB. This subset consists in fact of the 9 best features of Fig. 5.4 except the mean_HR and the mean_RR. The difference when these 2 last features are added to the 7 best ones is shown in red in Fig. 5.4. It can be seen that these 2 features tend to decrease the performance of the model, whereas they are respectively at the first and at the fourth position of the ranking of Fig. 5.4. The real reason of this consequence has not been found. However, it is suspected that these 2 features differ a lot between the subjects. So, even if they can be useful for the LS, in the TS the subjects have not the same values for these features. Therefore, it triggers miss classification at the end.

With the optimal subset found, all the combinations of 2 subjects among the 16 have been tested. The optimal point obtained has a sensitivity of 0.67, a specificity

of 0.69 for a class probability of 0.51.

Another important remark is that the 5 features with a coefficient of correlation higher than 0.25 are also found in the final subset of the best features.

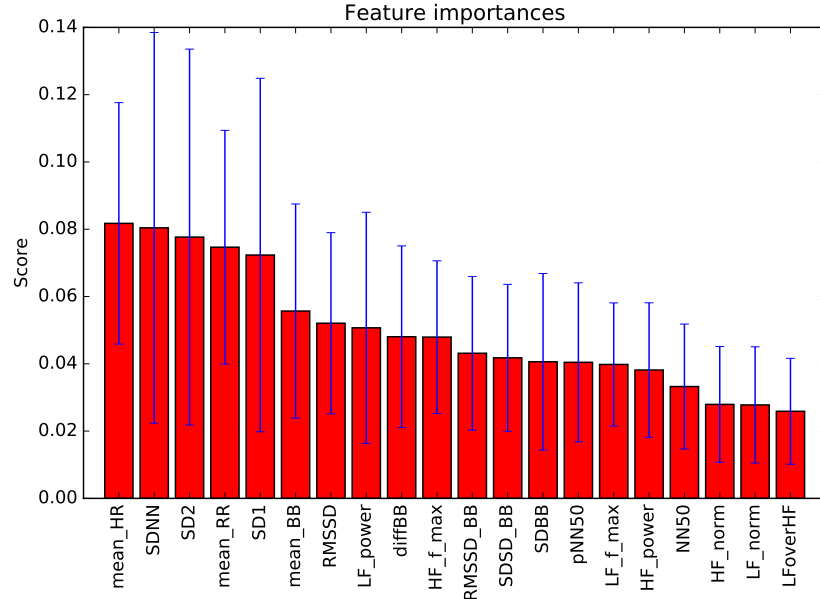


Figure 5.4: Ranking of the 20 features obtained by using extra trees on a eight folds cross-validation.

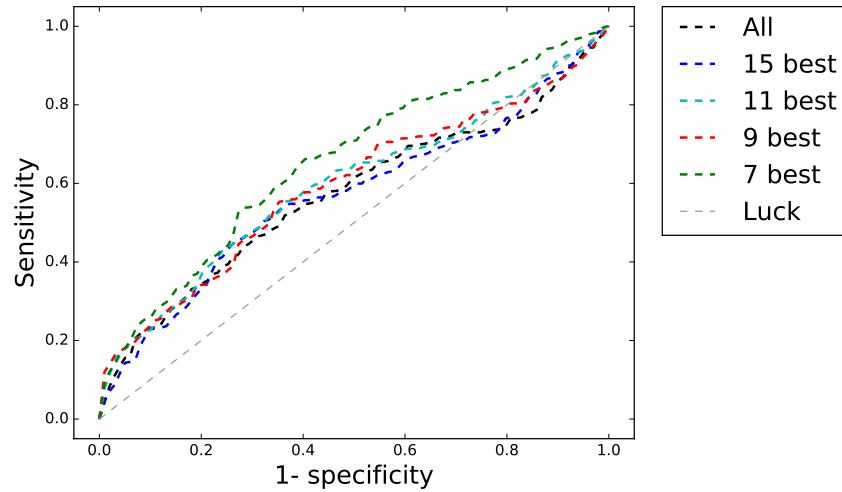


Figure 5.5: Evolution of the ROC curve as a function of the number features selected from the ranking of Fig. 5.4.

5.4 Discussion of the features computed

In Chapter 2, the influence of the parasympathetic and sympathetic nervous system on the heart was explained. These two systems act respectively in the high and low frequency parts of the spectrum of the ECG signal. The ratio of the power in the low frequency spectrum over the one in the high frequency part is thus a

reflection of the symptovagal balance. However, due to the important variability of the features between the subjects (Fig. A.1), it is difficult to draw a general tendency. Moreover, as previously said in chapter 2, different studies do not find the same results. Mahachandra et al. found that this ratio increases when drowsiness occurs [28]. On the other hand, in a study of Sahayadhas et al [32], a decrease of this ratio was observed when subjects were very drowsy. Here, if one just look at the mean of this ratio (Fig. A.1), it can be observed that it tends to increase as drowsiness occurs in concordance to Mahachandra et al.

We also made the choice of adding features computed from the respiratory domain in this thesis. However, in the literature, there are few studies that try to use the features from this respiratory domain to assess the drowsiness. It was previously shown than 2 features in this domain are present in the 7 subset of the best features.

However, there exist others methods that enables the selection of the features. A famous one is the recursive feature elimination with cross validation [48]. This technique automatically finds the set of optimal features by using a cross-validation. More precisely, if the estimator gives weights to its features, it will be trained on the LS with the whole set of features. Then, features with the smallest weights are removed. This procedure is stopped when the accuracy of the model built becomes to decrease (meaning that we reach the important features). Unfortunately, this technique was not implemented in this thesis and is let as a possible improvement.

5.5 Discussion about the sleep deprivation of the subjects

The variability of the features between the different subjects was already highlighted before. However, there is also an important variability of the drowsy profiles. Fig. B.1 of Appendix. B shows the number of windows of 120 seconds in which a given subject gets a given KDS score. From this Figure, one can observe that there are important variations in the way the subjects resisted to sleep deprivation. For instance, there are subjects like subject 19 that only gets a KDS score of 4 (the reference threshold to say that the subject is drowsy) only five times. To the contrary, there are subjects like subject number 18 that gets very often drowsy episodes.

5.6 Conclusion of the results

From this work, the results obtained tend to conclude that the ECG is not precise enough to detect the drowsiness. Three machine learning models have been tested. From the results presented in the present chapter, it is visible that the models developed learn some information from the features computed. Moreover, when all the different combinations are taken, the results are encouraging. However, it was shown that there were subjects for which the model correctly matches, and others for which the model was not able to predict correctly drowsy events. Therefore, the model developed is not robust enough to work on all subjects. Finally, the results obtained in this work are lower than the ones presented in the state of the art of this thesis. There are several possible reasons that explain this:

- First of all, the protocol of the experiment. Indeed, during the experiment, the subjects only performed 3 PVT tests of 10 minutes each. It gives 30 min of ECG signal per subject in this study whereas the studies in the literature generally use ECG recordings of several hours.
- Also, the acquisition of the data was not optimized for the ECG during the experiment. Indeed, the ECG were recorded only by two electrodes on the chest of the subjects. Moreover, experimenters admitted that, for some male subjects, it was difficult to correctly place the electrodes due to hairs on the chest.
- Another constraint of this work is the small size of the dataset. Indeed, only 16 subjects were finally available for the creation and the assessment of the model. Moreover, some subjects were considered only at the limit of drowsiness during few events during the protocol (Fig. B.1 of Appendix B). This issue was highlighted by the need of balancing the dataset in order to have the same number of drowsy and awake events in it (at the price of a reduction of 30% of its size). Therefore, it would be better to restart a new experiment with more subjects and in which the sleep deprivation lasts longer in order to get more subjects with events of larger KDS scores. Moreover, if more subjects are present and the dataset increases its size a lot, it would be possible to cut the dataset in 3 parts (50% for the LS, 25% for the TS and 25% for the validation set (VS)). The different models tested would be trained on the LS and assessed on the VS. Then, the best one would be tuned on the LS+VS and the performance would be estimated based on the TS. As the TS would not be used for the choice of the model or its parameters, the estimated performance would be closer to the one obtained when totally new subjects are tested.
- In this work, features of the respiratory domain were extracted. The respiratory signal was not directly available from the experiment. Therefore, this signal was derived from the ECG. However, if we really want to validate with more accuracy the use of the the features derived from the respiratory domain, it would be better to use a specific device that monitors this signal, like the pneumograph.
- Finally, the main problem encountered during the model evaluation was the important variability between the subjects. Different techniques were tried in order to compensate these variations but they did not improve the results.

Therefore, the conclusion from the present chapter is that the ECG is not suited to detect the drowsiness on its own. This technique should instead be added to others techniques (like the EOG or the EEG) in order to improve their accuracy. However, this conclusion should be validated with another experiment with more subjects and with higher sleep deprivations times.

Chapter 6

Conclusion

The aim of this work was to develop a system that automatically predicts drowsiness based on the ECG signal. Moreover, as the detection of drowsiness is a recent topic, its feasibility to detect drowsiness has not been totally proved among the scientific community. Therefore, another goal of this study was to test the system developed on a study performed 2 years ago at the University of Liège.

In this thesis, we performed a review of the literature in order to determine the different features indicative of drowsiness that can be computed from the ECG. This review highlighted the apparition of different domains in which these features can be computed.

Rapidly, we noticed that the detection of the R peaks in the ECG was the first step to do in order to obtain the different features. Therefore, the first part of this thesis consisted in the filtering of the ECG and in the development of an algorithm to automatically detect these peaks based on [1]. Once these peaks were found, the different features in the time and statistical domains, the non-linear domain, and finally in the frequency domain were computed. As one can derive the respiratory signal from the ECG (ECG-Derived Respiration signal), this thesis also incorporates parameters from the respiratory domain in order to see if this domain can be use to detect drowsiness.

Once these features were computed, three different machine learning models in binary classification were tested. The binary classes were awake or drowsy. The predictions of the model were compared to the true states of the subjects obtained from the KDS (*Karolinska Drowsiness scale*) score. Based on this score, a threshold was put to decide if the true state of the subject at a given time was drowsy or not [12]. The best model among the three tested was the one using the SVM classifier with a linear kernel.

The results of this thesis show that the ECG is not precise enough to predict drowsiness. In fact, the main problem found during this thesis was the important variability between the different subjects. Three techniques were tested in order to compensate this variability. However, none of these techniques manages to improve the results. Due to this important variability between the subjects, the features computed differ a lot from one subject to another, as the precision at the output of the model. This is why the developed model is not sufficient to only detect the drowsiness on its own based on the ECG.

However, this conclusion is based on the results obtained from one experiment. Other work on this very interesting domain must be continued in order to keep developing an automatic drowsiness monitoring system and assess its reliability. There are lots of perspectives for future work on this subject. The main ones are the followings:

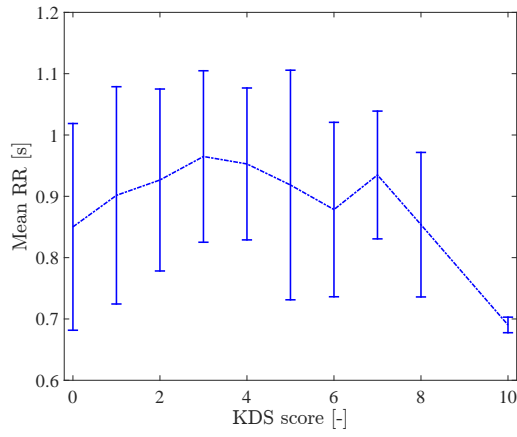
- In the frequency domain, the Lomb method has been used to derive the PSD of the tachogram. This technique was shown to be more beneficial with respect to the Fourier transform. However, the auto-regressive method could be investigated in order to compute the PSD.
- One of the main obstacle of the reliability of the ECG to predict drowsiness was the important variability between the subjects. The 3 techniques used in order to attenuate this variability did not perform in this thesis. Therefore, a research of possible other variability compensation techniques or other features (for instance male/female) could be continued in order to try to compensate this variability.
- Also, another protocol of experiment could be developed. This protocol would have a higher number of subjects. Moreover, the ECG signal would be collected continuously during the experiment instead of at specific time intervals. Finally, subjects with higher sleep deprivation could be taken.
- During this thesis, the binary classification was chosen as the approach for the machine learning model. The regression approach could also be tested instead of the binary classification.
- One could also test new kind of estimators (linear discriminant analysis, boosting methods,...) when building the machine learning model.
- Also, the ECG could be combined to the EEG or EOG to see if it is possible to enhance the reliability of the predictions of these 2 signals to detect drowsiness when the ECG is added.

There are thus lots of possible perspectives for future work.

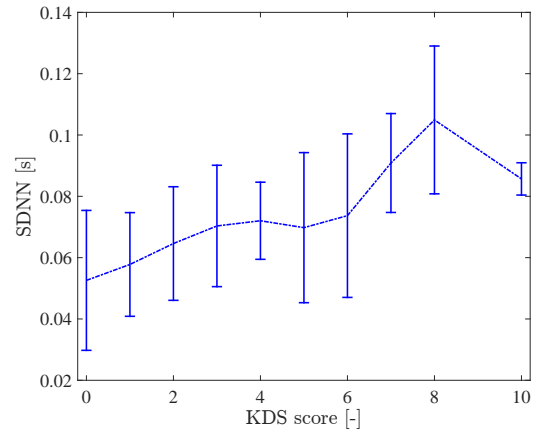
Appendices

Appendix A

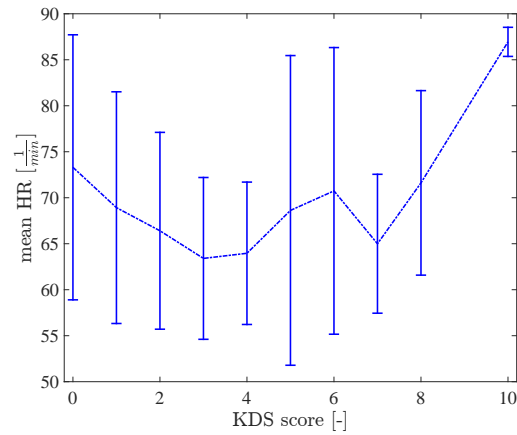
Features as function of the KDS score



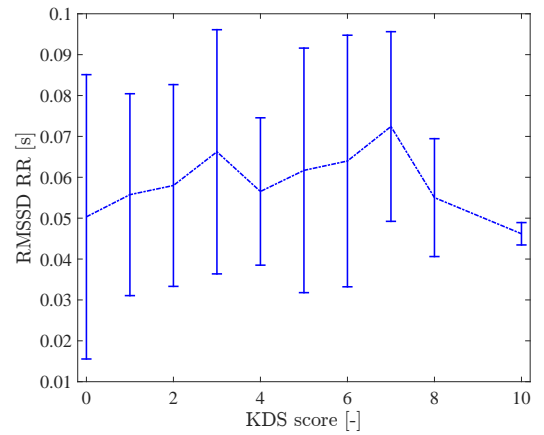
(a) Mean RR vs KDS score



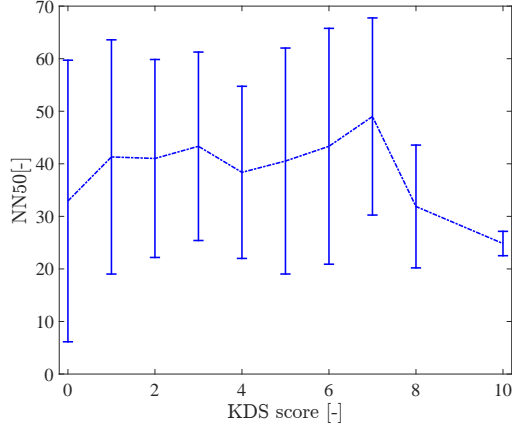
(b) SDNN vs KDS score



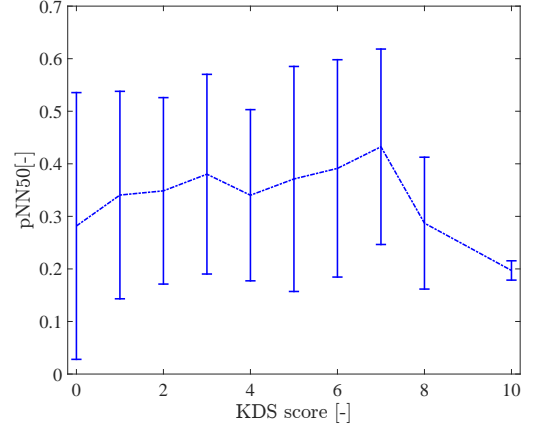
(c) Mean HR vs KDS score



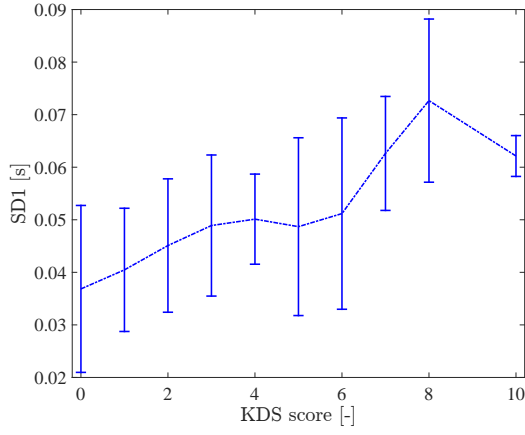
(d) RMSSD RR vs KDS score



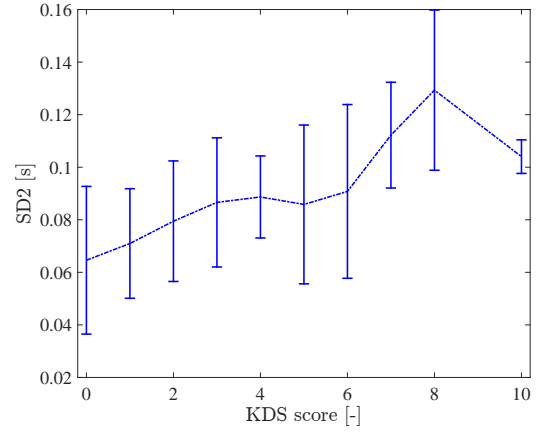
(e) *NN50 vs KDS score*



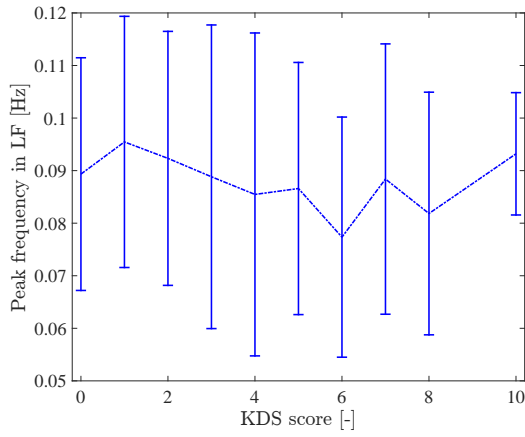
(f) *pNN50 vs KDS score*



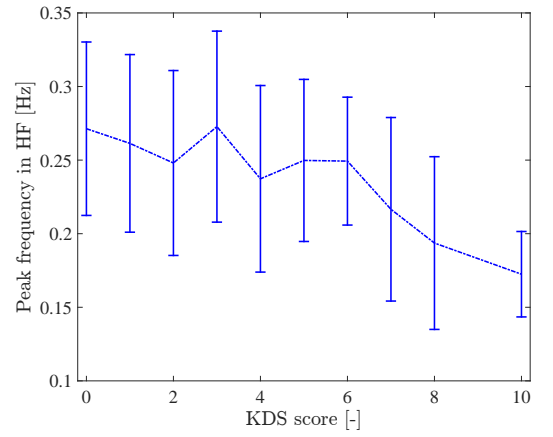
(g) *SD1 vs KDS score*



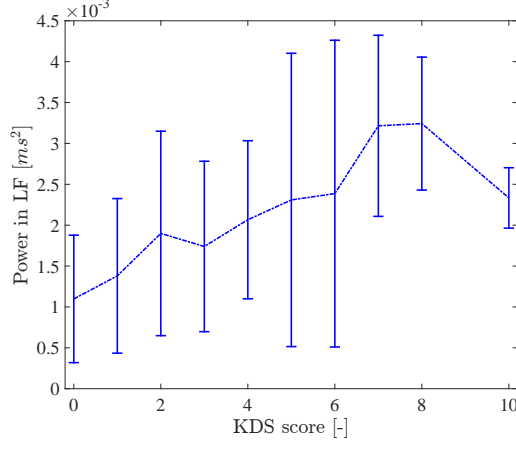
(h) *SD2 vs KDS score*



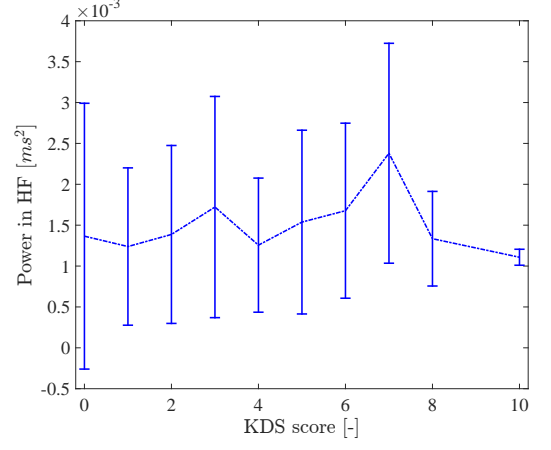
(i) *LF_f_peak vs KDS score*



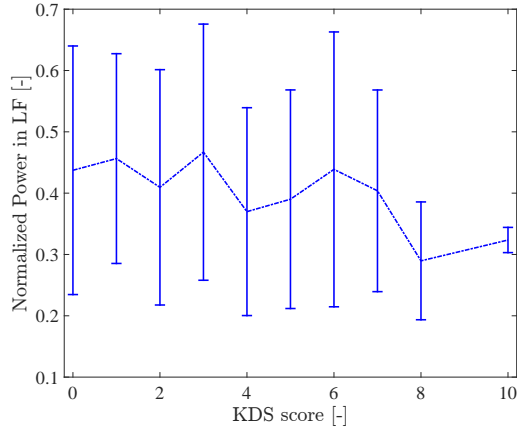
(j) *HF_f_peak vs KDS score*



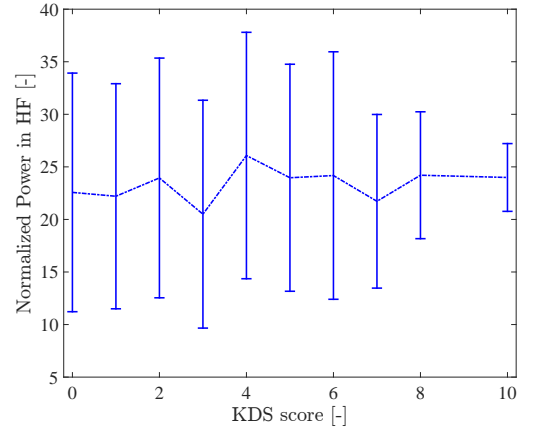
(k) *LF_power vs KDS score*



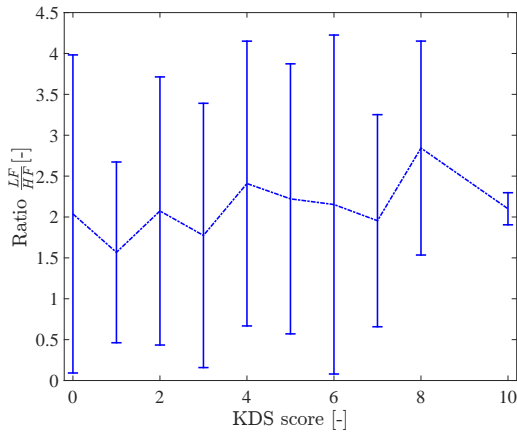
(l) *HF_power vs KDS score*



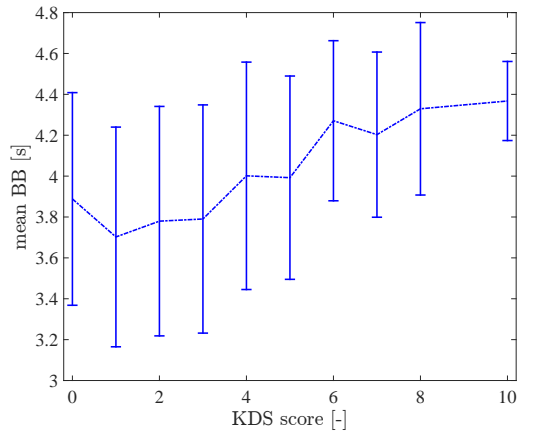
(m) *LF_power_norm vs KDS score*



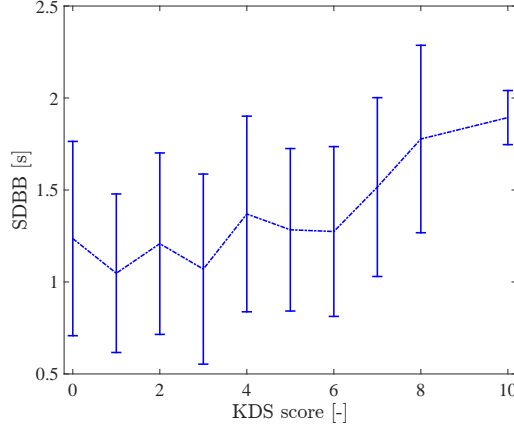
(n) *HF_power_norm vs KDS score*



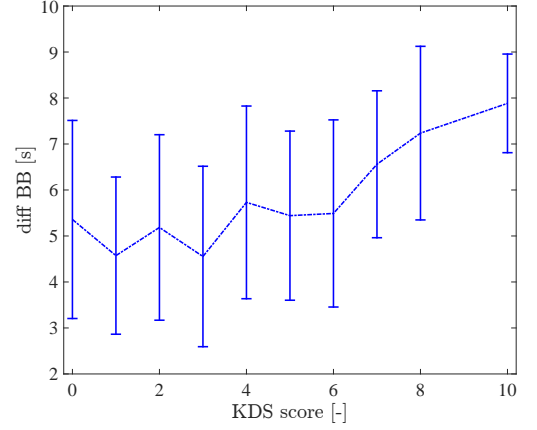
(o) *$\frac{LF}{HF}$ vs KDS score*



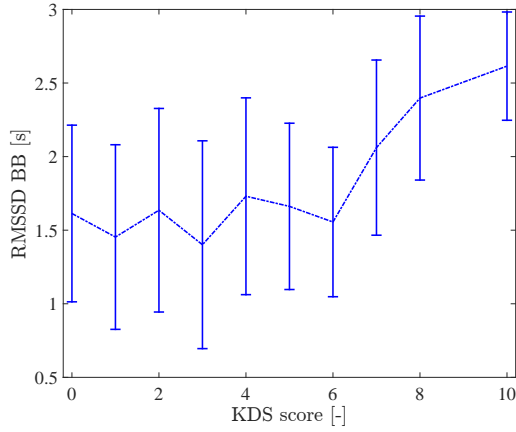
(p) *mean BB vs KDS score*



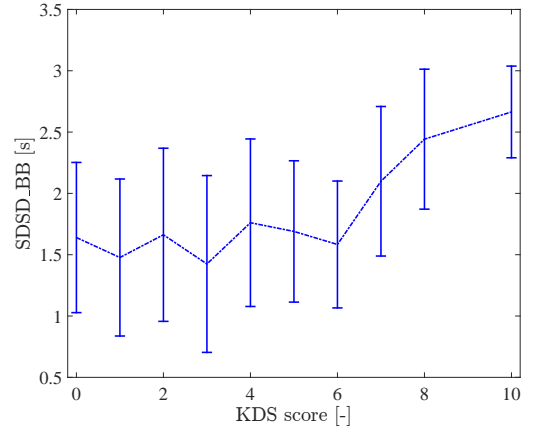
(q) SDBB vs KDS score



(r) diff BB vs KDS score



(s) RMSSD BB vs KDS score



(t) SDSD vs KDS score

Figure A.1: Plots of the mean and the standard deviation of each feature of interest for all the subjects of the experiment as a function of the KDS score. The mean is the blue dotted curve. For each value, the standard deviation (std) is represented by the vertical bar. Each vertical bar has a length of two std. This vertical bar gives information about the error on that value of the feature. A long vertical bar means that there are big variations of that feature for the same KDS score between the subjects. One can find the full name of the acronyms used in Tables 3.1, 3.2, 3.3, and 3.4 of Chapter 3.

Appendix B

Distribution of the KDS score per subject

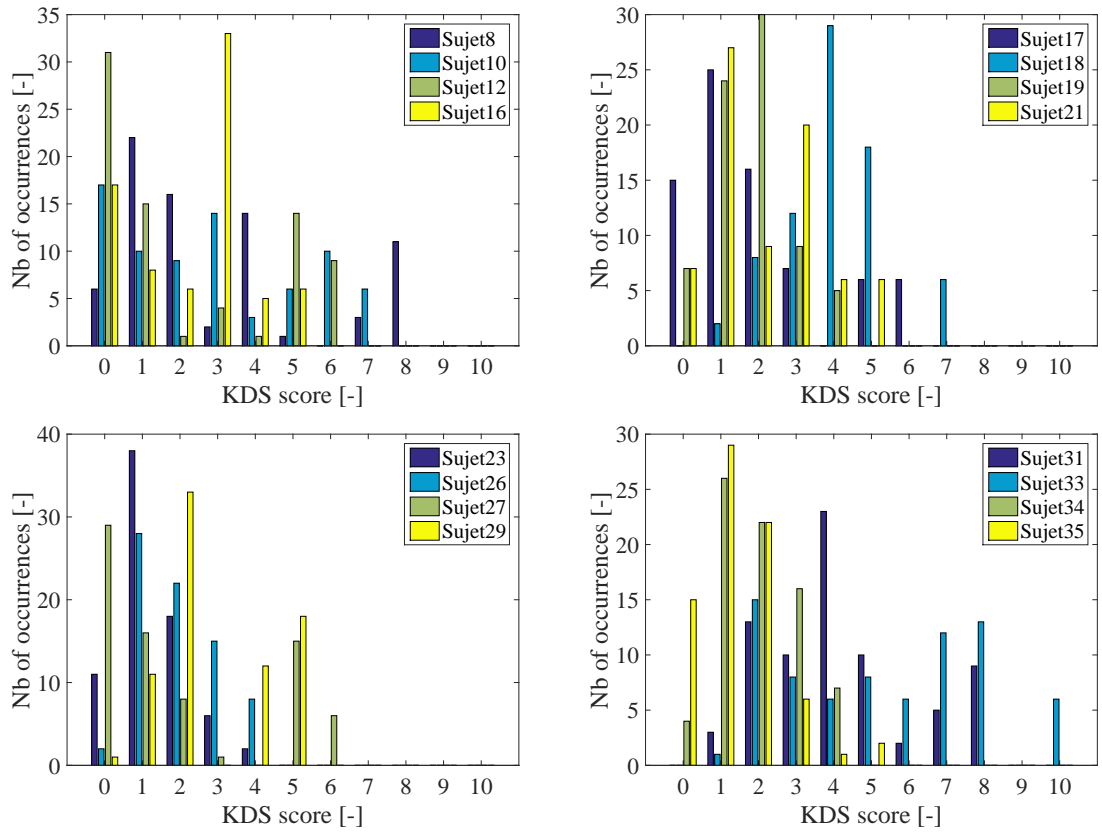


Figure B.1: Distribution of the KDS scores as a function of the subjects of the experiment.

Bibliography

- [1] D Benitez, PA Gaydecki, A Zaidi, and AP Fitzpatrick. The use of the Hilbert transform in ECG signal analysis. *Computers in biology and medicine*, 31(5):399–406, 2001.
- [2] ASFA and ISNV. La somnolence au volant - le livre blanc. <http://www.autoroutes.fr/fr/publications/lasomnolenceauvolantlivreblanc.htm>, Septembre 2013. [Online; accessed 28-April-2016].
- [3] SafetyNet (2009). Alcohol. http://ec.europa.eu/transport/road_safety/specialist/knowledge/pdf/alcohol.pdf. [Online; accessed 28-April-2016].
- [4] B. C. Tefft and AAA Foundation for Traffic Safety. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009-2013. November 2014.
- [5] D. Marghescu. The McGill Physiology Virtual Laboratory. <http://www.medicine.mcgill.ca/physio/vlab/default.htm>, 2016. [Online; accessed 28-April-2016].
- [6] Sang-Joong Jung, Heung-Sub Shin, and Wan-Young Chung. Highly sensitive driver health condition monitoring system using nonintrusive active electrodes. *Sensors and Actuators B: Chemical*, 171:691–698, 2012.
- [7] Devy Widjaja, Joachim Taelman, Steven Vandeput, Marijke AKA Braeken, Renée A Otte, Bea RH Van den Bergh, and Sabine Van Huffel. ECG-derived respiration: Comparison and new measures for respiratory variability. In *Computing in Cardiology, 2010*, pages 149–152. IEEE, 2010.
- [8] Louise Bérubé and Institut Philippe Pinel de Montréal. *Terminologie de neuropsychologie et de neurologie du comportement*. Montréal: Éditions de la Chenelière, 1991.
- [9] Agence wallonne pour la Sécurité routière. Campagne "fatigue au volant". <http://www.awsr.be/campagnes/campagne-fatigue-au-volant-juillet-2014>, 2014. [Online; accessed 28-April-2016].
- [10] Gari D Clifford. *Signal processing methods for heart rate variability*. PhD thesis, Department of Engineering Science, University of Oxford, 2002.
- [11] M Gillberg, G Kecklund, and T Åkerstedt. Sleepiness and performance of professional drivers in a truck simulator—comparisons between day and night driving. *Journal of Sleep Research*, 5(1):12–15, 1996.
- [12] Anna Anund, Göran Kecklund, Björn Peters, Åsa Forsman, Arne Lowden, and Torbjörn Åkerstedt. Driver impairment at night and its relation to physiological

- sleepiness. *Scandinavian journal of work, environment & health*, pages 142–150, 2008.
- [13] Juan Sztajzel. Introduction à l'électrocardiogramme (ECG). http://www.medecine.unige.ch/enseignement/apprentissage/module2/circ/apprentissage/intranet/cours/intro_ecg_dias.pdf. [Online; accessed 28-April-2016].
 - [14] Department of Physics and Astronomy. Hyperphysics. <http://hyperphysics.phy-astr.gsu.edu/hbase/hframe.html>. [Online; accessed 28-April-2016].
 - [15] E Marieb and K Hoehn. *Human anatomy and physiology.(8th)*. San, 2010.
 - [16] *Soins infirmiers en medecine et chirurgie 2*. Fonction cardiaque et système de conduction, de Boeck, 3rd edition edition, 2011.
 - [17] Ceufast.com. Congestive heart failure: The essence of heart failure course - nursing continuing education (ce) - nurse ce. http://www.ceufast.com/CHF_The_Essence_of_Failure__nurse_ceu_course.aspx. [Online; accessed 28-April-2016].
 - [18] P. Kolh. Physiologie des systèmes. Course at the University of Liège, 2013-2014.
 - [19] Philippe Lechat. Pharmacologie. Course at the University Pierre and Marie Curie, 2006-2007.
 - [20] R Hainsworth. The control and physiological importance of heart rate. *Heart rate variability*, pages 3–19, 1995.
 - [21] University of Nottingham School of Health Sciences. Cardiology teaching package. Online ressources.
 - [22] Nitish V Thakor. Biopotentials and electrophysiology measurement. *The Measurement, Instrumentation, and Sensors Handbook*, 74, 1999.
 - [23] University of Nottingham. Practice learning resources : Cardiology teaching package. <http://www.nottingham.ac.uk/nursing/practice/resources/cardiology/index.php>. [Online; accessed 28-April-2016].
 - [24] Queen's University. Analysis and interpretation of the electrocardiogram. <https://meds.queensu.ca/central/assets/modules/ECG/>. [Online; accessed 28-April-2016].
 - [25] Ichiro Kawachi. Heart rate variability. Online ressources, 1997.
 - [26] Task Force of the European Society of Cardiology and Others. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, 17:354–381, 1996.
 - [27] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: <http://circ.ahajournals.org/cgi/content/full/101/23/e215> PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

- [28] Manik Mahachandra, Iftikar Z Satalaksana, Kadarsah Suryadi, et al. Sensitivity of heart rate variability as indicator of driver sleepiness. In *Network of Ergonomics Societies Conference (SEANES), 2012 Southeast Asian*, pages 1–6. IEEE, 2012.
- [29] Noelia Rodriguez-Ibañez, Miguel A García-Gonzalez, Maria Aurora Filigrana de la Cruz, Mireya Fernández-Chimeno, and Juan Ramos-Castro. Changes in heart rate variability indexes due to drowsiness in professional drivers measured in a real environment. In *Computing in Cardiology (CinC), 2012*, pages 913–916. IEEE, 2012.
- [30] Eric Chern-Pin Chua, Wen-Qi Tan, Sing-Chen Yeo, Pauline Lau, Ivan Lee, Ivan Ho Mien, Kathiravelu Puvanendran, and Joshua J Gooley. Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*, 35(3):325–334, 2012.
- [31] José Vicente, Pablo Laguna, Ariadna Bartra, and Raquel Bailón. Detection of driver’s drowsiness by means of HRV analysis. In *Computing in Cardiology, 2011*, pages 89–92. IEEE, 2011.
- [32] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. Drowsiness detection during different times of day using multiple features. *Australasian Physical & Engineering Sciences in Medicine*, 36(2):243–250, 2013.
- [33] Zvi Shinar, Solange Akselrod, Yaron Dagan, and Armanda Baharav. Autonomic changes during wake–sleep transition: A heart rate variability based approach. *Autonomic Neuroscience*, 130(1):17–27, 2006.
- [34] Aihua Zhang and Fenghua Liu. Drowsiness detection based on wavelet analysis of ecg and pulse signals. In *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference*, pages 491–495. IEEE, 2012.
- [35] Joseph D Bronzino. *Biomedical engineering handbook*, volume 2. CRC press, 1995.
- [36] The free encyclopedia Wikipedia. Finite impulse response. https://en.wikipedia.org/wiki/Finite_impulse_response. [Online; accessed 1-May-2016].
- [37] Maxim Raginsky. Signals and systems, course at the the duke university. <http://maxim.ece.illinois.edu/teaching/fall08/lec12.pdf>. [Online; accessed 1-May-2016].
- [38] R. Dahyot. Numerical methods and advanced mathematical modelling. Course at the the Trinity College Dublin, 2007-2008.
- [39] AC Yang. Poincare plots: A minireview. *PhysioNet Heart Rate Variability-2006 Techniques. Applications and Futute Directions*, 2006.
- [40] London Scientific neurotherapy. Biopsychosocial medicine and neuromodulation. http://londonscientificneurotherapy.com/what_we_do.html. [Online; accessed 3-May-2016].
- [41] George B Moody. Spectral analysis of heart rate without resampling. In *Computers in Cardiology 1993, Proceedings.*, pages 715–718. IEEE, 1993.

- [42] Jeffrey D Scargle. Studies in astronomical time series analysis. II-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.
- [43] Nicholas R Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.
- [44] Elke Vlemincx, Ilse Van Diest, and Omer Van den Bergh. A sigh following sustained attention and mental stress: effects on respiratory variability. *Physiology & behavior*, 107(1):1–6, 2012.
- [45] Elke Vlemincx, Joachim Taelman, Steven De Peuter, Ilse Van Diest, and Omer Van Den Bergh. Sigh rate and respiratory variability during mental load and sustained attention. *Psychophysiology*, 48(1):117–120, 2011.
- [46] Mariel Grassmann, Elke Vlemincx, Andreas von Leupoldt, and Omer Van den Bergh. The role of respiratory measures to assess mental load in pilot selection. *Ergonomics*, pages 1–9, 2015.
- [47] William F Ganong and Kim E Barrett. *Review of medical physiology*. Appleton & Lange Norwalk, CT, 1995.
- [48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [49] B. Fortemps de Loneux. Validation définitive d’appareils de mesure de la vigilance par application d’une échelle d’évaluation KDS. Master’s thesis, University of Liège, Belgium, 2015.
- [50] L. Wehenkel P. Geurts. Applied inductive learning. Course at the University of Liège, 2015-2016.