DOCTORAL THESIS

Development and validation of algorithms for automatic and real-time characterization of drowsiness

Author: Clémentine François
Supervisor: Professor Jacques G. Verly

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering Sciences

October 2018
Department of Electrical Engineering and Computer Science
Faculty of Applied Sciences
University of Liège

Abstract

Drowsiness is characterized by impairments of performance which can lead to disastrous accidents, in particular in all types of transportation and high-risk industrial plants. Therefore, it is crucial to be able to measure the ability of an individual to perform correctly and safely a task based on his/her level of drowsiness (LoD).

We thus developed a new, objective, automatic, and real-time drowsiness characterization system based on the analysis of ocular parameters extracted from images of the eye. The use of images of the eye is called, in this thesis, photooculography (POG), and this technique is recognized as one of the most significant and pratical technique to characterize drowsiness. In order to validate our POG-based drowsiness characterization system, we compared, for a number of participants, the LoD determined by our POG-based system to several references: (1) the LoD obtained by visually analyzing polysomnographic (PSG) signals, (2) the performance in the accomplishment of a Psychomotor Vigilance Test (PVT), (3) the performance in the accomplishment of a Driving Session in a professional driving simulator, and (4) the self-assessed LoD using the Karolinska Sleepiness Scale (KSS).

The visual analysis of PSG signals is a very time-consuming task but is essential for the validation of drowsiness characterization systems. PSG is indeed considered by many experts as the gold standard for sleep. Therefore, we also developed, in this thesis, a new, automatic PSG-based system for characterizing drowsiness. This system is primarily intended to be used as a reference for the validation of non-PSG-based drowsiness characterization systems. To evaluate the performance of this automatic PSG-based system, we compared, for a number of participants, the LoD determined by our automatic PSG-based system to (1) the LoD obtained by visually analyzing PSG signals, (2) the performance in the accomplishment of a PVT.

To evaluate the performance of the two systems that we developed, we conducted two experiments. In Experiment A, 30 healthy volunteers performed three visual PVTs under increasing sleep deprivation over two days. In Experiment B, 12 healthy volunteers performed three Driving Sessions in a professional driving simulator under increasing sleep deprivation over two days. During each experiment, we recorded, for each participant, images of the eye, PSG signals, performance data (related to the task performed by the participant) and the self-assessed LoD using the KSS.

Results show that:

- our POG-based drowsiness characterization system has a significant potential for reliably determining the LoD of individuals accomplishing a task and ultimately for preventing drowsiness-related accidents. Furthermore, our POG-based system has the advantage of being non-invasive, usable in any condition, and of requiring no intervention from the individual.
- our PSG-based drowsiness characterization system has a significant potential to become a general reference for drowsiness characterization and to help scoring experts save time.

Acknowledgments

This work could not have been performed without the help of a large number of people. I would like to thank all those who have been involved in the realization of my thesis.

In particular, I would like to thank my supervisor, Jacques Verly, for having given me the opportunity to realize this thesis. Thank you for your availability, your wise and relevant advises (on scientific and technical aspects but also on how to present my research to the scientific community) through all these years. Thank you for your enthusiasm, your encouragements, and also for the careful reading of this manuscript.

I am grateful to Anne-Sophie Nyssen for accepting to be my co-promoter. Thank you for all the interesting discussions we had and that opened my mind to other considerations and perspectives. I hope I succeeded in including some of this knowledge in this manuscript.

Thank you also to the members of my jury for accepting the hard task to evaluate this work. I am very honored to have you all on my jury. Thank you especially to Dr. Poirrier who followed my work, always encouraged me and brought me all his expertise in the domain of sleep and drowsiness.

Many thanks to all my colleagues from the University of Liège (especially Quentin Massoz and Philippe Latour) for their encouragements and all the rich discussions we had, and for bringing a good working atmosphere. I also thank the initial team of the project that led to the development of the innovative POG-based drowsiness characterization system that is presented in this doctoral thesis. Thank you also to the whole Phasya team for their support and especially to Eva and Jérôme for the proofreading of this manuscript. Jérôme, a particular thank you for all your support, your kindness, and for your courage in all the steps and challenges we have gone through. I am proud to be your partner in Phasya and to have you as a friend.

I also wish to express my gratitude to all the participants who participated in our study and all the students who contributed through their Master thesis work.

I also wish to thank all my friends from the university of Liège and others (especially Sophie, Capucine, Charlotte, Stéphanie, Julie, Dorothée, Mélissa) for their eternal encouragement and their wonderful friendship during the realization of this work.

Last but not least, I thank my family (my parents, my sister, my in-laws, Xavier and my son, Maxence) for always believing in me, for supporting me and bringing all the love I needed to get to the end of this job. Xavier, thank you especially for your patience, your unconditional support and for sharing my life for 8 years. Your presence on my side gave me the strength I needed to do this job. A particular thank you also to my son, Maxence, for being the greatest happiness of my life and giving me the motivation and the energy I needed to complete this work.

Thank you to everyone who supported me so well.

Contents

List of Figures

List of acronyms

AASM American Academy of Sleep Medicine
ANN Artificial Neural Network
AUC Area Under the Curve
BCI Brain Computer Interface
BSRT Behavioral Sleep Resistance Task
EDS Excessive Daytime Sleepiness
EEG ElectroEncephaloGram
EMG ElectroMyoGram
EOG ElectroOculoGram
ESS Epworth Sleepiness Scale
FFT Fast Fourier Transform
HHT Hilbert Huang Transform
HR Heart Rate
HRV Heart Rate Variability
HT Hilbert Transform
HVD Hilbert Vibration Decomposition
IMF Intrinsic Mode Functions
IR Infrared
JDS Johns Drowsiness Scale
JTV Johns Test of Vigilance
KDS Karolinska Drowsiness Scale
KSS Karolinska Sleepiness Scale
LoD Level of Drowsiness
MSLT Multiple Sleep Latency Test
MWT Maintenance of Wakefulness Test
NREM Non-Rapid Eye Movement
NTSB National Transportation Safety Board
OA Ocular Artifact
OOG OptoOculoGraphy
OP Ocular Parameter
OSLER Oxford Sleep Resistance Test
OSS Objective Sleepiness Scale
PCA Principal Component Analysis
PERCLOS PERcentage of eye CLOSure
POG PhotoOculoGraphy
PSG PolySomnoGraphy

PVT Psychomotor Vigilance Test
REM Rapid Eye Movement
ROC Receiver Operating Characteristics
RT Reaction Time
SDLP Standard Deviation of the Lateral Position
SEM Slow Eye Movement
SSS Stanford Sleepiness Scale
STFT Short Time Fourier Transform
SVM Support Vector Machine
VAS Visual Analog Scale

Chapter 1

Introduction

1.1 Context

Drowsiness is recognized as being a major cause of several types of accidents, especially for drivers, operators of equipment, and supervisors of large industrial complexes. Drowsiness is indeed an unintentional and uncontrollable need to sleep that can occur at any time of the day and that can affect everyone. There are, however, some specific factors that promote onset and evolution of drowsiness. The appearance of drowsiness during the day depends strongly on the alternation of the sleep/wake cycle, which is regulated by two physiological mechanisms/processes: the circadian process (our internal biological clock) and the homeostatic process (our sleep debt). There are thus some moments more conducive to drowsiness than others. Moreover, drowsiness can also be promoted, among others, by a lack of sleep, taking certain medication, alcohol, or drugs, and suffering from a sleep pathology.

Drowsiness induces hypo-vigilance, which consists in a decrease in reactivity to the environment. The drowsy or hypo-vigilant individual is thus affected by a deficit in information processing and a slowing down of the reaction time which can both cause dramatic accidents. More generally, drowsiness impairs an individual's judgment and ability to execute a task correctly [5, 6].

In the particular case of road driving, drowsiness is estimated to be responsible for 20 to 30% of road accidents [7]. It can indeed lead to difficulties in maintaining a constant speed, a correct distance between vehicles, and a proper position on the road, resulting in involuntary lane changes among other things. These various manifestations can unfortunately result in serious accidents. In the USA, more than 72,000 police-reported traffic accidents caused by drowsiness are recorded every year [8]. In France, drowsiness (sometimes loosely referred to as "fatigue") is the main cause of fatal accidents; drowsiness indeed accounts for 27% of fatal accidents [9]. Moreover, it is widely accepted that these figures underestimate the reality. Some drivers indeed die from such accidents, and crash investigators must look for clues that could demonstrate that drowsiness was the, or a, factor leading to the accident, but these clues are not always detectable. Other drivers prefer to hide the real causes because of concern with their vehicle insurance.

In 2007, in France, a study showed that 4% of drivers had at least one “near-crash” due to drowsiness in the past year, which represent 1.5 million motorists [10].

In addition, the results of another study carried out by researchers from the sleep unit of the hospital "Raymond-Poincare de Garches", in France, in 2014, with 375 truckers are alarming. Results show that 30% of drivers felt that they could have an accident due to drowsiness, and that one in ten drivers admitted to having experienced an episode of severe drowsiness that forced them to stop. They also indicate that 28% of drivers have slept less than 6 hours before taking the wheel [11].

In a study by Hakkanen and Summala in 2000 on truckers, 40% of long-haul drivers reported having trouble staying alert for at least 20% of their conduct and 20% admitted dozing at least twice during each trip [12].

Furthermore, the risk of being drowsy while driving is increased by different factors, e.g. if the driver drives in the time interval between 2:00 am and 5:00 am, if the driver suffers from sleep disorders such as narcolepsy or sleep apnea, or if the driver has a sleep debt (as is the case for a driver who slept 5 hours or less on the day before the start of his[1] trip) [13]. In addition, after 24 hours of activity without sleep, the effect of drowsiness is equivalent, in terms of reflexes, to having a blood alcohol level of 1 g/l. In Belgium, the legal limit of alcohol concentration per liter of blood that must not be exceeded to be authorized to drive is 0.5 g/l. This means that being active for more than 24 hours is equivalent to exceeding twice the legal limit.

The most affected are drivers under 25 of age, drivers over 50, and irregular workers. Young novice drivers indeed constitute a very vulnerable population in terms of road safety and, in particular, concerning drowsiness at the wheel [14]. This would be due to poor risk assessment, poor lifestyle (young people pay less attention to sleep), and excessive trust in themselves. By contrast, older, more experienced drivers are more likely to be affected by the deterioration in physical and cognitive abilities related to age.

Drowsiness is therefore a critical safety problem and contrary to what one might think, this problem will not disappear with the emergence of vehicle automation. Indeed, the automatic driving systems already found in today’s vehicles, such as cruise control, further increase the risk of drowsiness while driving. Indeed, by replacing the continuous and stimulating activity of regulating speed by a passive surveillance activity, these systems increase the monotony of the task and thus promote the appearance of drowsiness. Virtually, all vehicles today (2018) are at most at level 2 of vehicle automation on a scale from 0 to 5 (the classification was established by the Society of Automotive Engineers). The car manufacturers are now working on level 3 cars, which are expected to appear around 2018-2020. Level 3 would enable cars to drive alone in certain environments such as the highway but a driver would still have to be present and be able to take back

[1]Throughout, the generic masculin he, his, himself, etc. is used for concisness. He/his also means she/her.

the control of the vehicle in a relatively short lapse of time. In order to ensure that the driver is able to take back the control of the vehicle, technologies for monitoring the state of the driver will become essential. At present, some manufacturers offer systems for monitoring the behavior of the driver and the behavior of the vehicle, such as the detection of lane deviations. These systems are very useful at the current automation level (2), but they will become obsolete at the subsequent stages (3-5). Indeed, when the car will drive autonomously, monitoring its behavior will not give any information on the state of the driver, and technologies that directly monitor the driver by analyzing physiological parameters will become a necessity.

Drowsiness affects not only road transport sector but also other types of transport (rail, air, etc.) as well as industry. An impressive series of transportation and industry disasters occurred during night work and when the level of human vigilance was low, mainly between 2:00 am and 5:00 am. One can cite, without limitation, the nuclear accidents at Chernobyl in 1986 (around 1:30 am), Three Mile Island in 1979, the Bhopal tragedy in 1984 (around 2:00 am), the explosion of the shuttle Challenger in 1986, or the more recent train accident in 2014 in Chicago (during the day).

The National Transportation Safety Board (NTSB) in the USA has been fighting for years to reduce fatigue-related accidents in transportation and has even made it an important item on its most wanted transportation safety improvements for 2017-2018 [15].

More generally, automation and computerization have transformed human activity. We are more and more assisted in our professional and personal activities. We remain in the system to monitor the process, monitor the devices, and occasionally regain control in unexpected situations but the problem is that it deprives us from some physical and cognitive stimulation, and this encourages the onset of drowsiness and thus makes us less able to make good decisions or perform critical tasks. The solution is not to ban computerization and automation because they allow us to increase our skills and intellectual capacities but rather to verify that we are able to work with these technologies while performing the tasks that are still under our responsibility and that are sometimes critical for our safety.

Characterizing drowsiness, monitoring its level, and determining the times when it reaches a dangerous level thus constitute an important “grail” and valuable endeavor in the area of public health and safety.

1.2 Goal of the thesis

The specific problem addressed in this thesis is to study the phenomenon of drowsiness and to develop an automatic and real-time drowsiness characterization[1] system, mainly from physiological parameters, and that would be usable in any operational environment. In order to establish contextualized thresholds that are critical to safety, we also

[1]Throughout, the term characterization is used to denote “monitoring” and “quantification”.

link the characterization of drowsiness from physiological parameters to the evaluation of the performance of individuals in the execution of a task and of the risks associated with a particular situation. Even though drowsiness is a physiological state that can be characterized by physiological parameters, one cannot ignore the context in which such drowsiness arises. It is necessary to be able to determine the state of drowsiness of an individual based on physiological parameters but the level of the associated risk will vary from one application/situation to another.

Another important concept addressed in this thesis is the validation of drowsiness characterization systems. It is indeed essential to prove the effectiveness and the reliability of such systems if one wants to use them in real life. Therefore, we also proposed a validation strategy and used that strategy to validate the drowsiness characterization system that we developed. In addition, we also developed another automatic drowsiness characterization system based on physiological parameters to serve as a reference (i.e. ground truth, gold standard)[1] for the validation of other systems.

Our approach for the work reported here thus consisted in several steps.

First, we started, not surprisingly, with a review of the literature to identify the different methods for characterizing drowsiness. In this review, we defined several criteria in order to propose a classification of all the methods identified and we performed an analysis of the literature that enabled us to:

1. Determine that the use of images of the eye, i.e. photooculography (POG), appears to be the best method for characterizing drowsiness in operational settings since it is completely physiology based, task independent, and non invasive. The eye activity reflects brain activity and some ocular parameters are recognized by the scientific community as very good indicators of an individual's state of drowsiness.

2. Realize that there is no universal agreement on the reference to be used to evaluate the state of drowsiness of an individual and thus to validate drowsiness characterization systems. We therefore decided to use four references to assess the state of drowsiness: a physiological reference, two performance references, and a subjective reference. To concentrate our efforts, we carried out our research with the tasks of psychomotor vigilance and driving in a professional driving simulator. However, the methodology and results used in this doctoral thesis can easily be generalized to other critical processes for the safety of individuals and goods in various sectors such as aviation, maritime, nuclear, and medical.

Second, we worked on the development of an innovative POG-based drowsiness characterization system that is usable in operational environments, and that continuously, objectively, and automatically determines a level of drowsiness (LoD) on a numerical scale. This system is based on several key ocular parameters, the values of which are

[1] Throughout, the term reference is used to denote "ground truth" or "gold standard"

obtained from (video) images of the eye.

Third, we needed to verify that the LoD determined by our POG-based drowsiness characterization system was well related to the several references: (1) the LoD obtained by analyzing visually polysomnographic (PSG) signals (physiological reference), (2) the performance of individuals in the accomplishment of a Psychomotor Vigilance Test (PVT), (3) the performance of individuals in the accomplishment of a driving session in a simulator, and (4) the self-evaluated LoD using the Karolinska Sleepiness Scale (KSS) (subjective reference).

Polysomnography (PSG) is used as a reference because it is considered to be the "gold standard" for quantifying sleep and thus, for most experts, also the best physiological reference to study drowsiness. When experts analyze the PSG signals to determine in which sleep stage an individual is, they visually analyze the signals via consecutive (non-overlapping) time windows. One says that the experts score the signals. This operation of scoring is very time consuming and the results may differ from one scorer to another.

Fourth, we therefore also developed an innovative, automatic, PSG-based drowsiness characterization system using signal processing and machine learning techniques, together with a drowsiness characterization method based on the Karolinska Drowsiness Score (KDS). This system aims to have a common, automatic, and objective reference for the assessment of an individual's state of drowsiness and thus to be able to validate other drowsiness characterization systems that would, in general, be less invasive and better suited for the operational environment. Moreover, this automatic PSG-based drowsiness characterization system could also be used as a diagnostic tool for people with excessive daytime sleepiness (EDS), which may be due to sleep disorders.

We thus conducted several experiments to acquire data in order to help us to develop our two drowsiness characterization systems (i.e. POG-based and PSG-based) and to validate them. In each experiment, the healthy participants were asked to perform three tests over two consecutive days, and they were not allowed to sleep between the first and last test, with the result that the successive tests are at increasing levels of sleep deprivation.

1.3 Personal contributions in this thesis

The present thesis includes the following main personal contributions:

- Proposal of a classification of methods to characterize drowsiness;
- Development of the experimental protocol;
- Participation in the experimentations and data acquisitions;

- Definition of the list of ocular parameters to use in our POG-based drowsiness characterization system (not fully disclosed in this thesis);
- Development of methods to compute ocular parameters from the positions of the eyelids and of the pupil in each eye image (not fully disclosed in this thesis);
- Development of methods to determine a numerical LoD from a set of ocular parameters (not disclosed in this thesis);
- Development of a strategy for the validation of the two drowsiness characterization systems that we developed;
- Development of new criteria for the visual analysis of PSG signals via the Karolinska Drowsiness Scale (KDS);
- Validation of our POG-based drowsiness characterization system by comparison with several references;
- Development of an innovative, automatic, PSG-based drowsiness characterization system;
- Validation of the innovative, automatic, PSG-based drowsiness characterization system;
- Proposal of thresholds based on our POG-based drowsiness scale to alert individuals before they constitute a risk.

The main results of this thesis were presented in journal publications and conference proceedings. The list of all the publications is given at the end of this report.

1.4 Outline of the work

Chapter 2 introduces the physiological background on drowsiness and reviews the different methods for characterizing drowsiness. An analysis and a discussion of the methods are given at the end of the chapter and a justification of the chosen approach for the development of a drowsiness characterization system is also provided.

Chapter 3 presents our data acquisition environment as well as the data that are used later as references to validate our drowsiness characterization systems. The chapter includes a description of the requirements that we established, the protocol that we developed to acquire relevant data, and the analysis of the data related to the references.

Chapter 4 describes the innovative POG-based drowsiness characterization system that we developed. This chapter explains our motivation and gives an overview of the methods that we implemented. We dedicated the main part of this chapter on the evaluation of the performances of our system. We showed the results of (1) the effect of sleep deprivation on our POG-based LoD, (2) the relation between our POG-based LoD and

the references, (3) our POG-based LoD as predictor of failure, and (4) the comparison with well-known ocular parameters. At the end of the chapter, we discuss the results and draw a conclusion.

Chapter 5 is dedicated to the development of the automatic, reference, PSG-based drowsiness quantification system. This chapter gives our motivation and the details of the implemented method. The global method involves a preprocessing step, the extraction of features via signal processing, and a conversion step into an automatic PSG-based LoD using a machine learning technique. The two last parts of the chapter then focus on the results and the discussion.

Chapter 6 summarizes the major contributions and gives perspective for future work.

Chapter 2

State-of-the art

This chapter describes the state of the art in drowsiness and the existing methods to characterize it. Section 2.1 explains what the drowsiness phenomenon is, why it occurs, and how the sleep-wake cycle is biologically regulated. Section 2.2 contains a description of excessive daytime sleepiness (EDS) and of sleep pathologies that can induce drowsiness. Section 2.3 approaches the existing, traditional methods to characterize drowsiness, whether in the sleep clinic or in operational environments. In particular, we distinguish between subjective methods (based on self-evaluation) and objective methods (based on physiological parameters or performance of an individual). Section 2.4 concludes by describing the method selected for the development of our drowsiness characterization system and the methods selected to evaluate and validate it.

2.1 Mechanisms of sleep and drowsiness

The drowsiness of an individual can be seen as a physiological state where this individual is inclined to sleep and has difficulty to stay awake. This state is intermediate between complete wakefulness (where the individual is fully alert) and sleep. Just as hunger and thirst are the instincts that drive us to eat and drink, drowsiness is the instinct that drives us to sleep; it is a physiological necessity [16].

Drowsiness should not be mistaken with fatigue, especially due to physical or mental effort. Indeed, fatigue is an affect, an emotion experienced in the present. It is a feeling of weakness felt after an effort and it may be associated with a biochemical or physiological change in muscle or the brain. Drowsiness is more of a state influenced by various physiological mechanisms, an irresistible tendency to doze. The individual is seized with sleep drive. Drowsiness is characterized by a lack of awareness of the here-and-now, and this is not the case for fatigue. This lack is the reason for the dangers of drowsy driving. We can be drowsy without being tired, and tired without being drowsy, and we can be both tired and drowsy [17, 18].

A sleep specialist at the Centre Hospitalier Universitaire (CHU) of Bordeaux, Philip, said: “Fatigue is the increasing difficulty to accomplish an effort, while drowsiness is

the inability to stay awake" [19].

The term "drowsiness" is often used interchangeably with the words "sleepiness" or "somnolence", which are considered synonyms in this thesis.

The consequences of drowsiness are numerous: shortened sleep latency, increased attention deficits, increased slowed cognitive functions, and increased reaction times with consecutively impaired performance, leading to work or motor accidents. From the perspective of the individual, drowsiness reduces not only the efficiency at work, but it also leads to problems with concentration, memory, and mood, all of which also affect performance and quality of life [5, 20, 21].

In addition, several physiological mechanisms that regulate periods of wake and sleep, such as the circadian and homeostatic processes, as well as sleep inertia, may favor the appearance of drowsiness at certain times during the day. These three mechanisms/processes are described in more details in subsections 2.1.1, 2.1.2, and 2.1.3.

Our body indeed needs sleep to maintain good functioning and health. We are programmed to sleep every night like a means to restore our body and our mind. Sleep pressure, or the need for sleep, varies along the day and the alternation between wake and sleep needs to coincide with certain biological processes. Man is indeed a diurnal animal, which is active during the day and inactive during the night. Sleep pressure is thus high at night and increases as we approach the moment of falling asleep. Conversely, after waking up, the body evacuates its need for sleep [22]. There is therefore a time period conducive to sleep. Two systems interact together to determine the time of the transitions between wake and sleep: the circadian process and the homeostatic process. These processes were modeled and described in the model of Borbely et al. in 1982; he called them Process C (circadian) and Process S (homeostatic), respectively [23]. This model can be illustrated by a balance with, on the one hand, the weight of the circadian process and, on the other hand, the weight of the homeostatic balance. The two processes also explain why, in normal conditions, we stay awake during the day and we sleep during the night. Waking and sleeping alternate in a periodic cycle throughout an individual's life.

In the brain, when the regions responsible for wakefulness are active, they inhibit the activity of the regions responsible for sleep, and conversely. Several regions in the brainstem and the hypothalamus promote wakefulness by sending signals to the cerebral cortex. These signals are called neurotransmitters. When the neurons in the wakefulness regions are active, neurotransmitters are sent to the cortex which is activated and we are awake. Examples of neurotransmitters that are involved in controlling wake and sleep are histamine, dopamine, serotonin, and orexin. Histamine is released by neurons present in a region promoting wakefulness (the tuberomammilary nucleus (TMN)). Interestingly, several antihistaminic medications block wakefulness and induce drowsiness. Serotonin is notably used by our body under the control of our circadian clock to produce melatonin, also called the "sleep hormone". The production of melatonin

increases in the evening, thus causing drowsiness [24, 25, 26].

Figure 2.1 shows the interaction of the circadian process and the homeostatic process, together with the secretion of melatonin.

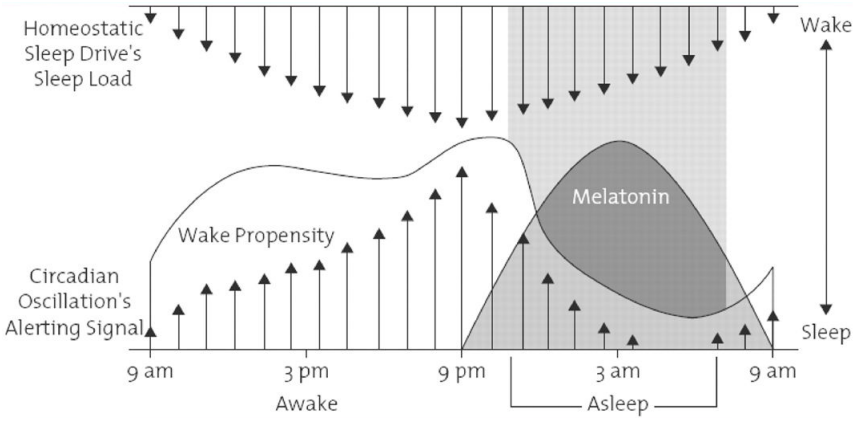

Figure 2.1: Sleep wake cycle [1].

2.1.1 Circadian process

A variety of biological functions - food intake, hormonal secretions (e.g. melatonin, cortisol), body temperature - including the sleep-wake cycle obey a circadian process. The term comes from the latin "circa" (meaning "around" or "approximately") and "dies" (meaning "day"), and refers to functions that operate in a cycle of about 24 hours. These circadian processes are synchronized and coordinated by an internal biological clock that plays the role of conductor.

Our biological clock is located in the heart of the brain, in the hypothalamus organ. This clock has a spontaneous activity and it regulates circadian processes through its connections to various parts of the body. It is our internal clock that determines the "sleep windows", i.e. the favorable moments to sleep.

This biological clock thus has an endogenous activity, but it must still be synchronized to 24 hours because, in reality, our biological rhythms last more than 24 hours, and specifically 24h10min. For example, in 1962, Michel Siffre, a French speleologist, spent two months at the bottom of Scarasson abyss, without any watch and without any daylight. His cycles shifted daily, and at the end of two months, he took his breakfast around 7:00 pm and went to bed in the late morning [27]. Therefore, our clock needs to be reset to 24 hours and this is the role of the zeitgebers (time donors) such as light and social factors (eating meals, working hours, etc.). The zeitbegers are indeed exogenous environmental cues that help to synchronize our internal clock [26] to the evolution of time in the physical world. The most powerful zeitgeber is light because it inhibits the secretion of melatonin. Melatonin is indeed synthesized from serotonin and secreted by

the pineal gland in response to the absence of light. Melatonin levels are high during the night and zero during the day. Melatonin secretion begins 1-2 hours before sleep with a peak of secretion between 2:00 and 4:00 am, and its levels are very low 1-2 hours after awakening. The light that reaches the retina acts on the suprachiasmatic nucleus via the optic nerve. The light information is then sent to the pineal gland, which blocks the secretion of melatonin. This hormone is secreted with an intensity four times higher in winter than in summer. The effect of the light depends on the light intensity, its duration, and its spectral composition. In fact, the more the light is intense and the longer it lasts, the greater the effect is. Exposure to a light of 2,000 lux (which corresponds to the light intensity that does not require one to use a flash to take a picture) for 2 hours is sufficient to significantly reduce melatonin secretion, and thus shift the biological clock and the sleep-wake cycle. In addition, a light of 480 nm wavelength (blue) is the most effective in blocking the secretion of melatonin. It is as effective as a white fluorescent light 100 times more intense [28, 25].

Melatonin therefore acts as a biological marker of the internal time. Indeed, it is through its production that the brain is informed about the relative duration of the hours of darkness and light on a period of 24 hours, as well as throughout the year.

The circadian organization (24 hours) of drowsiness is also linked to changes in body temperature. Temperature reaches a minimum in the middle of the night and a maximum late in the day. The cycle of temperature coincides with that of vigilance. The cyclical variation of temperature determines the "sleep windows". Thus, the decrease in body temperature plays a key role in the onset of drowsiness. We fall asleep more easily when our temperature drops, and we wake up a few hours after the minimum temperature. Moreover, what many call "the nap of the afternoon" actually represents a slight drop in temperature curve around 15h. The drop in core temperature occurs twice per 24h and thus determines the periods of drowsiness. Indeed, the curve of the circadian process has two hollows, one in the early afternoon and one at night, that make us more conducive to sleep and drowsiness [22].

Other factors intervene to prepare sleep such as decreased heart rate and decreased cortisol secretion.

All of the above also explains why shift workers or people travelling and changing time zones are more likely to often experience periods of drowsiness because they disrupt their normal cycle of wake and sleep.

2.1.2 Homeostatic process

Sleep is also regulated by homeostatic mechanisms. Specifically, the need for sleep increases during wakefulness and decreases during sleep; the less we sleep, the more we need to sleep. To describe this homeostatic process, one talks about "homeostatic debt". A debt that increases with the wake time and decreases with the sleep time. Thus, the more the time awake increases, the more the sleep pressure becomes significant. When

this homeostatic debt reaches its high threshold, sleep occurs. When we sleep, the debt decreases to the low threshold and awakening occurs. In case of sleep deprivation, the return to equilibrium is manifested by increased drowsiness (proportional to the duration of arousal) and by a boost of compensatory sleep (including an increase of the percentage of slow waves sleep). Adenosine is thought to play a critical role in the homeostatic debt but it may not explain this complete phenomenon [29, 26].

Sleep is especially beneficial for the brain, and acute sleep deprivation (a sleepless night) causes daytime sleepiness and excessive fatigue, mental slowing, and a reduction of attention, concentration, and reflexes. In the long term, sleep deprivation causes an increase in appetite (with subsequent weight gain), and a decrease in alertness, mood, and motivation [6].

Sleep deprivation is unfortunately common in today's society. Technologies have indeed completely changed the rhythms of our lives; we live in a society where productivity is the key word and where sleep has less and less place.

2.1.3 Sleep inertia

Sleep inertia is a physiological state that occurs directly after awakening (from a night's sleep or a day's nap). It is defined as the feeling of grogginess and is characterized by impaired alertness, which can lead to decreased performance in the execution of a task. This state can therefore be dangerous for workers who must take crucial decisions, or act, directly after being awakened [30], e.g. doctors working in hospitals.

Sleep inertia would be caused by the accumulation of adenosine during deep sleep. The duration of sleep inertia varies between 10 and 30 minutes but, in some cases, it can last up to a few hours. The duration and the severity of sleep inertia is influenced by several factors such as the depth of sleep when awakened. If the brain is in the slow-wave sleep stage when awakened, this indeed encourages sleep inertia and reaction times are very slow. Other factors that influence sleep inertia are the phase of the circadian clock of each individual and caffeine intake. Caffeine blocks adenosine receptors in the brain, thus decreasing greatly sleep inertia [31].

2.2 Excessive daytime sleepiness and sleep disorders

All individuals can be normally affected by drowsiness during the course of the day due to the physiological mechanisms/processes described above, but some people have a higher propensity than normal to drowsiness. Indeed, 6 to 11% of the population report a lot of drowsiness during the day, much more than normal. One says that this part of the population suffers from severe excessive daytime sleepiness (EDS) [32].

EDS is a common complaint encountered in the overall population and in neurological practice. In 2001, a study by Guilleminault reported that, according to the "National

Sleep Foundation 2000 Omnibus Sleep in America Poll", 43% of adults say they are sleepy during the day and that this interferes with their daily activities a few days per month; and 20% of adults report suffering from daytime sleepiness at least a few days a week [33]. In 2009, Powell even showed that a drop in performance due to drowsiness may be worse than that associated with alcohol [34].

It is difficult to find a clear definition of EDS in the literature because it is not a pathology, it is more a symptom of a sleep disorder (particularly of hypersomnia or narcolepsy) or of insufficient time allowed for sleep. In addition, there is no standardized methods to assess EDS. Some experts use well-known questionnaires (e.g. Epworth Sleepiness Scale) but others prefer to use a clinical evaluation in a sleep clinic (using for example the Maintenance of Wakefulness Test). The methods for assessing drowsiness will be discussed in the next section. What is certain is that sleepiness (or drowsiness) is considered excessive when an individual is unable to remain alert during waking hours and in situations where it is crucial to be attentive [21].

There are many sleep disorders that cause fragmented sleep and that manifest themselves as drowsiness. The best known is the sleep apnea disorder that affects a large part of the population. One can also cite narcolepsy, for which the prevalence is much lower than for the sleep apnea disorder, and hypersomnia. Individuals suffering from sleep disorders are therefore at risk when they perform a critical task such as driving.

In 2001, a study by Leger and Vecchierini states that chronic sleepiness (or drowsiness) in adults has a prevalence of 15 to 20% of the adult population when it is moderate and 6 to 11% of the adult population when it is severe [32].

Apart from sleep disorders, drowsiness can also be due, among others, to

- other diseases such as diabetes, obesity, depression, etc.,
- consuming alcohol,
- consuming drugs,
- taking some medications,
- performing a monotonous task.

In the next section, we cover the main traditional methods to characterize drowsiness. The review is not exhaustive but it gives a broad view of what exists.

2.3 Methods for drowsiness characterization

For many years, researchers have conducted studies to better understand drowsiness and to try to characterize and monitor this phenomenon. Several methods exist and there are different ways to classify them. The way we selected is as follows.

- First, we partition the methods in two categories on the basis of the nature of the drowsiness measure, i.e. sleep propensity (specific to the field of medical diagnostics) vs. real-time level of drowsiness (LoD). The methods related to sleep propensity quantify, via a medical evaluation in a sleep laboratory or a questionnaire, the risk that an individual falls asleep at some future time. The methods related to the real-time LoD provide, via an analysis of physiological parameters or an analysis of performance parameters or a questionnaire, an indication of the LoD of an individual in the present. The criterion of speed of delivery of the measure ("timescale") is thus important to discriminate the two categories.
- Second, in each category, we consider the dichotomy of objective methods vs. subjective methods. Subjective methods are mainly based on questionnaires and refer to self reported mesurements while objective methods are mainly based either on performance measures or on physiological measures.

2.3.1 Methods related to sleep propensity

Figure 2.2 shows the block diagram of the different methods related to sleep propensity to quantify drowsiness (represented by their acronyms). All methods are detailed hereafter.

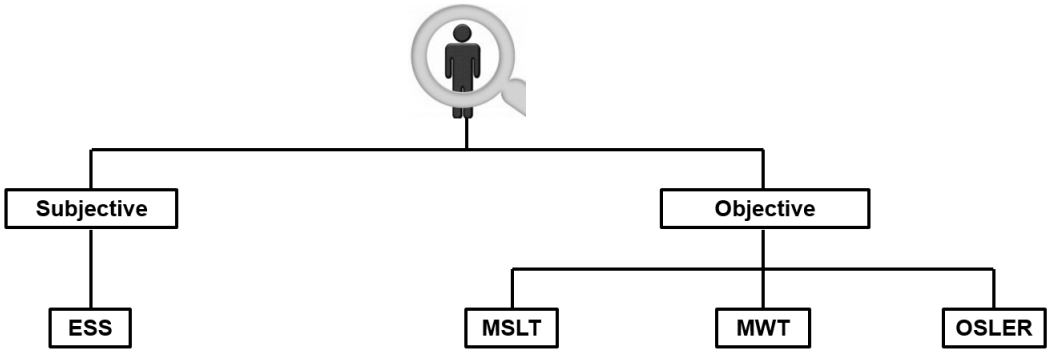

Figure 2.2: Block diagram of the methods related to sleep propensity.

Subjective method

The subjective method described below yields a sleep propensity via the use of a questionnaire. The answers to the questionnaire are provided directly by the individual

tested.

Epworth Sleepiness Scale (ESS)

The Epworth Sleepiness Scale (ESS) is a subjective scale that was proposed by Johns in [35]. It measures sleep propensity by asking the individual about the likelihood of dozing off in particular situations of daily life. It does not refer to the subjective feelings of drowsiness at the present time.

The ESS thus provides a propensity to sleep, but neither a state or LoD, nor a sleep latency.

Contrary to what is done for other subjective scales (such as the Karolinska Sleepiness Scale), the ESS is not repeated several times since the answers of the individual tested should in principle remain the same over some extended period of time and should be independent of the present state of drowsiness of the individual.

The ESS test consists in asking the individual to evaluate the chance that he would doze in 8 different situations such as sitting and reading, watching TV, etc. The chance is entered as a number according to;

- 0: no chance of dozing;
- 1: slight chance of dozing;
- 2: moderate chance of dozing;
- 3: high chance of dozing.

The drowsiness score of the individual is obtained by adding all numbers entered. The interpretation of the score is as follows:

- 0-5: lower normal daytime sleepiness;
- 6-10: higher normal daytime sleepiness;
- 11-12: mild excessive daytime sleepiness;
- 13-15: moderate excessive daytime sleepiness;
- 16-24: severe excessive daytime sleepiness.

The ESS is highly reliable, easily conducted, and widely used [36]. However, it has the disadvantage of proposing situations that may not be common to everybody, and this could therefore distort the result.

Objective methods

The objective methods described below yield a sleep propensity via the use of clinical tests. These clinical tests use physiological parameters to determine sleep propensity.

Multiple Sleep Latency Test (MSLT)

The Multiple Sleep Latency Test (MSLT) is a diagnostic tool to identify sleep propensity which is used to diagnose sleep disorders. For example, it is used to diagnose narcolepsy and breathing disorders (as well as the effectiveness of their treatments) [37].

The MSLT was proposed in 1977 by Dement and Carskadon [38].

The MSLT's main goal is to provide an objective measure of drowsiness. The measure provided is that of sleep latency. Sleep latency is defined below in the description of the MSLT test.

The MSLT is generally performed during the daytime, immediately following an overnight sleep with a control polysomnogram to ensure that the individual had adequate sleep during the night and to make sure that there was no sleep disruption that would be the cause of artifacts in the MSLT test.

The test consists of four to five sessions. For each session, the individual is placed in a lying-down position in a quiet, comfortable room. At the beginning of each session, the individual is asked to close his eyes and the experimenter turns the lights off. These conditions are designed to be conducive to sleep. The individual is asked to let himself fall asleep, i.e. without trying to fight sleepiness. In any case, he is awaken after 20 minutes. The test is repeated every two hours (for a total of four or five sessions). For each session, the time between "lights out" (with eye closed) and the moment when the individual falls asleep unequivocally is recorded. This moment presumably corresponds to the end of sleep stage N1, that will be discussed later, and this duration is called the sleep latency. The four or five values obtained for the sleep latency can be used individually, but their average is usually considered to obtain a single value of sleep latency for the individual.

The moment when the individual falls asleep is determined primarily from polysomnographic (PSG) recordings, including electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) [39, 40].

The test is based on the idea that the drowsier an individual is, the faster he will fall asleep. Therefore, people with EDS are expected to fall asleep quite quickly in each session of this test [41]. However, there is a wide range of normal MSLT sleep latencies, so that a short one is not necessarily an indication of a sleep disorder.

Typical MSLT sleep latencies lead to the following classification of drowsiness states

[42]:

- severe drowsiness: corresponding to a sleep latency lower than 5 min,
- moderate drowsiness: corresponding to a sleep latency between 5 min and 10 min,
- normal drowsiness: corresponding to a sleep latency higher than 10 min (normal).

The MSLT is used in many research protocols.

As indicated above, the MSLT was conceived to be a diagnostic tool and to provide an objective measure of drowsiness in fixed, laboratory conditions. Because of this, the measure that it provides (sleep latency) can be used for the initial diagnosis of a sleep disorder, and to monitor the evolution of a possible pathology and of the effect of a corresponding treatment.

The MSLT provides a measure of sleep latency in the standard conditions of the test, and thus only provides a rough indication of how fast an individual might fall asleep in other conditions.

For example, the laboratory conditions in which the test is conducted (including lying down, with eyes closed, and in darkness) are not representative of, say, driving conditions. It is thus unreasonable to try to draw, from an MSLT, precise conclusions about when the individual might fall asleep while driving. However, it is clear that the diagnosis provided by the MSLT is useful in identifying individuals suffering from sleep disorders and being thus at risk of falling asleep at the wheel.

Because each MSLT takes a significant amount of time, and because it requires a specialized infrastructure, this test is rather heavy to use, which means that it only makes sense to use it in clinical settings to address potentially significant health issues.

The MSLT is also used to determine whether an individual with short sleep latency suffers from narcolepsy. The rational is the following. This rational uses the concept of REM sleep (i.e. rapid eye movement sleep) stage which is a particular deep sleep stage that will be discussed later. In a normal individual, REM sleep stage first occurs about 60 to 90 minutes after sleep onset, and then periodically after that. After a good night sleep, a normal individual would never enter REM within 5 minutes of falling asleep. Therefore, to determine whether a patient might suffer from narcolepsy, the sleep technician just needs to determine whether the patient enters into REM sleep stage within five minutes after falling asleep in a given session. The time between sleep onset and REM sleep is called “latency to REM”. Although even a single early REM occurrence is abnormal, the standard protocol requires latency to REM of less than five minutes at least twice over the set of the four or five sessions for the patient to be declared to suffer from narcolepsy [40, 43].

The MSLT is often considered to be some kind of a gold standard.

Maintenance of Wakefulness Test (MWT)

The Maintenance of Wakefulness Test (MWT) is also a diagnostic tool to identify sleep propensity but, contrary to the MSLT, the MWT measures the ability to stay awake and not to fall alseep [44]. This test is thus typically used for the analysis of sleepiness during waking hours.

The MWT is used to diagnose EDS and difficulties in staying awake.

The MWT's main goal is to provide an objective measure of drowsiness. The measure provided is that of a sleep latency [37].

The MWT is generally performed 1.5 to 3 hours after the individual wakes up, so he is presumably well-rested.

The test consists of four sessions. For each session, the individual typically sits upright on a bed with his back and head supported. The room is dimly lit, with the source of light just behind the individual's head and out of his field of vision. The individual is asked to stay awake as long as possible. The session ends when the individual falls asleep or after 40 minutes if he does not. The test is repeated every two hours (for a total of four sessions). The duration between the start of the session and the moment the individual falls asleep is recorded: this duration is called the sleep latency. The four values obtained for the sleep latency can be used individually, but their average is usually considered to obtain a single value of (MWT) sleep latency for the individual [40].

Identically to the MSLT, the moment where the individual falls asleep is determined primarily from PSG recordings, including EEG, EOG, and EMG.

The test is based on the idea that the drowsier an individual is, the faster he will fall asleep, even if he tries to fight sleepiness.

Healthy people fall asleep after about 30 minutes and some may stay awake through the test. Therefore, MWT latencies of less than 8 minutes are considered abnormal. They may be indicative of EDS [33].

While MSLT and MWT both measure sleep latencies, either is not a substitute for the other. Both tests will typically give different latencies for a given individual.

The MWT is often used for legal considerations, e.g. for reinstating a driver's license.

Many of the other remarks made concerning the MSLT apply here as well.

Oxford Sleep Resistance Test (OSLER)

The Oxford Sleep Resistance Test (OSLER) is a simplified version of the maintenance of wakefulness test (MWT) but it uses behavioral/performance-based parameters rather than EEG-based parameters to assess the sleep latency [45]. The test is performed on an individual in an environment conducive to sleep (dark room isolated from noise). The individual is asked to lie down and to try to stay awake. He must respond to a light stimulus produced by an LED placed at eye level 2 meters away from his head. The test lasts 40 minutes (or, sometimes, only 20) and the LED flashes for 1 second every 3 seconds. The individual is asked to react to the light stimulus by touching a button. The test ends after 40 minutes (or 20) or after 7 unanswered light stimuli (equivalent to 21 seconds of inattention). The analysis of the sequence of reactions to the stimuli allows one to estimate the sleep latency [46].

The OSLER shows good correlation with the MWT based on EEG for the estimation of the sleep latency [45].

The advantages of the OSLER are its simplicity and its low cost. The disadvantage is that, most of the time, the individual performs the test without error. To overcome this limitation, an extended measure of the OSLER has been defined to take into account the changes in the error sequence [46].

The OSLER is also used as a secondary reference in the analysis of drowsiness.

Some authors have also proposed a software-based alternative for the OSLER, the Behavioral Sleep Resistance Task (BSRT) [47].

2.3.2 Methods for operational and real-time characterization of drowsiness

The main traditional methods for characterizing the LoD in real-time can also be divided in two categories on the basis of their objective or subjective character. Objective methods are based on the direct measurement of physiological parameters of the individual or on the performance of the individual in the accomplishment of a task. Subjective methods are mainly based on questionnaires, interviews, or self reports from the individual, and they do not use physiological parameters.

Figure 2.3 shows the block diagram of the different methods related to the operational and real-time characterization of drowsiness (sometimes represented by their acronyms). All methods are detailed hereafter.

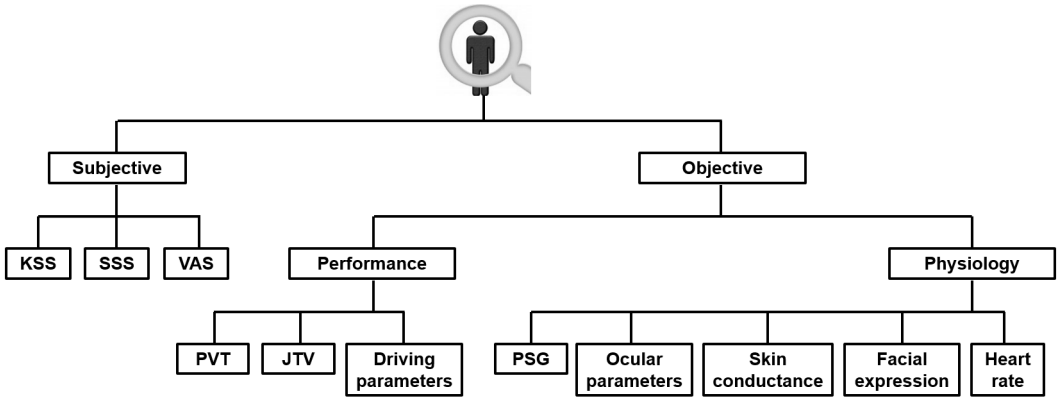

Figure 2.3: Block diagram of the methods related to the operational and real-time characterization of drowsiness.

Subjective methods

The subjective methods described below yield a state of drowsiness or an LoD, and not a sleep propensity. This is thus in contrast with the ESS. The answers to the different scales/questionnaires must be provided by the individual tested.

Karolinska Sleepiness Scale (KSS)

The Karolinska Sleepiness Scale (KSS) test asks the individual to choose, among the nine states proposed, the one corresponding the most closely to his perceived state of drowsiness over the last few minutes [48].

The KSS is a Likert-type scale with the fifth, middle state “Neither alert nor sleepy” being effectively neutral. Despite its characteristic, the scale is not fully symmetrical in its wording.

Since the answer to the test can be provided quickly, the test can be taken repeatedly and provides only a minor distraction to the individual. Moreover, it is very easy to administer.

The KSS is used as the means of quantifying the state of sleepiness in many experiments and it is the most prevalent scale in the drowsiness-related literature.

Karolinska Sleepiness Scale	
Rate	State
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to remain awake
8	Sleepy, some effort to stay awake
9	Very sleepy, great effort to stay awake, fighting sleep

The KSS was validated against EEG features extracted from individuals performing a driving test in a simulator with sleep deprivation [49]. Another study also showed good correlation between high KSS scores and the risk of accidents [50]. However, the KSS showed low reliability for the diagnosis of sleep disorders [51].

Stanford Sleepiness Scale (SSS)

The Stanford Sleepiness Scale (SSS) [52] is similar in spirit to the KSS. One difference is that the SSS uses seven states rather than nine.

It is not clear whether the SSS is truly a Likert-type scale. Indeed, it is not obvious that its fourth, middle state is neutral. The wording of the other states certainly does not exhibit any symmetry.

Murray W. Johns indicates in [37] that the SSS has the disadvantage of using some words that are either not universally known (such as “woozy”) or that are quite vague (such as “vital” and “foggy”). We agree that it is difficult to grasp quickly the exact meaning of each of the SSS “degrees”.

Similarly to the KSS, since the answer to the test can be provided quickly, the SSS can be taken repeatedly and provides only a minor distraction to the individual. However, it is difficult for an individual to keep a precise picture of the scale in his head. The SSS was validated and showed high reliability but, as for KSS, not for the diagnosis of sleep disorders, notably the obstructive sleep apnea syndrome [53].

Stanford Sleepiness Scale	
Rate	State
1	Feeling active, vital, alert, or wide awake
2	Functioning at high levels, but not at a peak; able to concentrate
3	Awake, but relaxed; responsive but not fully alert
4	Somewhat foggy, let down
5	Foggy; loosing interest in remaining awake; slowed down
6	Sleepy, woozy, fighting sleep; prefer to lie down
7	No longer fighting sleep, sleep onset soon; have dream-like thoughts

Visual Analog Scale (VAS)

A Visual Analog Scale (VAS) is a universal scale that can be used easily and anywhere as it is a simple, linear, visual scale [54]. The VAS is a horizontal line segment of 10 cm in length. The two extremities of the segment correspond to two extreme opposites. For example, the extreme could be "very satisfied" and "very dissatisfied". In the context of alertness and drowsiness, the extremes could be "very alert" and "very drowsy".

The individual must make a mark at a point that he feels corresponds to his level of alertness/drowsiness in the last few minutes. The position of the mark can then be converted into a score, for example on a scale from 0 to 10, possibly rounded to an integer. Instead of making a mark, one could imagine that the individual simply picks a number (or an integer) from 0 to 10. This value could be provided in writing, orally, etc.

The major advantage of the approach is that one does not need to think about, and interpret, the description, sometimes complicated, of the "states" (as in the SSS). To the extent that the "states" of the KSS and SSS are linear in the sleepiness, a VAS is, for all practical purposes, virtually equivalent to the KSS and the SSS [33, 37].

Again, since the answer to the VAS test can be provided quickly and easily, the test can be taken repeatedly and provides only a minor distraction to the individual. The VAS has the significant advantage of being instantaneously obvious. However, a disadvantage of the VAS is that the terminology used by different experts for the extremities of the scale may vary, thus limiting the reproducibility of results from one study to another.

Objective methods

Among the objective methods, we first consider the ones that are independent of the task performed by the individual, i.e. the physiology-based methods (polysomnography, ocular parameters, etc.), and then the ones that are dependent of the task, i.e. the performance-based methods (Psychomotor Vigilance Test, driving performance, etc.).

Physiology-based methods

Polysomnography (PSG)

Polysomnography (PSG) may prove to be another method for characterizing drowsiness. PSG is indeed a recording of different physiological signals that enable one to detect changes during wake and sleep. These changes are refered here as activity. We indeed use the term activity to refer to changes in certain frequency bands over time. PSG monitors, via the use of electrodes, several body activities such as the brain (electroencephalogram (EEG)), the eye (electrooculogram (EOG)), the muscles (electromyogram (EMG)), and the cardiac activity (electrocardiogram (ECG)).

PSG is the reference for the study of sleep. Therefore, we first present the characterization of sleep using PSG and, second, we present the characterization of drowsiness using PSG.

Characterization of sleep using PSG

Before the era of modern sleep research, in the 1920s, scientists regarded sleep as an inactive brain state. In 1924, Hans Berger recorded the first human EEG. From EEG recordings, scientists then discovered that during sleep, the brain stays active with a dynamic behavior. Each EEG recording indeed measures electric potentials, i.e. voltages, between two electrodes, which is the result of the electrical currents generated by synaptic activity (neurons) in the brain. The EEG activity is characterized by its shape, amplitude, and frequency. There are four frequency bands of interest for the study of sleep:

- Delta (δ): 0.5-3 Hz
- Theta (θ): 4-7 Hz
- Alpha (α): 8-12 Hz
- Beta (β): 13-25 Hz.

Over time, research revealed two main types of sleep. These were defined by some particular electrical patterns in the brain of a sleeping individual and by the presence or absence of eye movements. The two main types of sleep are the Rapid-Eye-Movement sleep (REM sleep) and the Non-Rapid-Eye-Movement sleep (NREM sleep). In 1968, Rechtschaffen and Kales described the structure of sleep using several stages [55]. This has been the reference in sleep labs until the arrival, in 2007, of the new scoring manual published by the American Association of Sleep Medicine (AASM) [56].

The AASM nomenclature gives three categories of sleep stages:

- Wake stage;

- Non-Rapid Eye Movement (NREM) sleep stage;
- Rapid Eye Movement (REM) sleep stage.

The wake stage represents almost two thirds of an individual's life time. This stage can be further partitioned into two (sub)stages:

- Active wake, eyes open;
- Passive wake (or calm), eyes closed.

During the active wake, the electroencephalogram (EEG) activity is fast and of low voltage. It is usually characterized by beta frequencies (13-25 Hz) or even gamma frequencies (25-100 Hz) depending on whether the individual is very active or not. There is also presence of body movements and eye movements that often create artifacts.
The passive wake can be associated with a state of rest, relaxation, eyes closed. Under these conditions, the EEG activity is characterized by the presence of alpha frequencies (8-12 Hz) which are higher in the occipital regions. Normally, there is no body movement but muscle tone is still present, especially at the chin, where the tone is of great amplitude on the EMG.

The NREM sleep represents, on average, each night, 80% of total sleep. It is divided into three stages of increasing depth of sleep denoted by N1, N2, and N3. In the progression from N1 to N3, brain waves become slower, higher in amplitude, and more synchronized, and the eyes remain still. We now give descriptions of these three stages.

- N1 stage (NREM 1):
  The N1 stage includes an EEG activity slightly slower than in the wake stage. The EEG activity consists of mixed frequencies, usually in the alpha and theta frequency bands with a strong increase of activity in the theta band. The EOG usually shows slow eye movements, as opposed to the blinking of the wake stage. In the EEG, specific patterns also appear under certain channels and the EMG on the chin decreases because the individual feels more relaxed. The N1 stage is usually associated with the state of drowsiness.

- N2 stage (NREM 2):
  The N2 stage is also characterized by a mixed frequency EEG activity. Special patterns appear under the central EEG channels. These are the "spindles" (which are sinusoidal waveforms with frequency between 12 and 16 Hz) and the "K complexes", not associated with awakenings (which are biphasic waveforms with a large negative wave followed immediately by a positive wave). There is no longer any EOG activity and the EMG is still existent but reduced.

- N3 stage (NREM 3):
  The N3 stage (corresponding to Stage 3 and Stage 4 of the Rechtschaffen and

Kales scoring manual) includes a slow wave activity (SWA) on 20% or more of the EEG. Taking a standard window of 30 seconds, the slow wave activity must be of a duration greater than or equal to 6 seconds to be scored as stage N3. These waves are in the delta frequency band (0.5 to 3 Hz), and have a magnitude greater than 75 microvolts. K-complexes as well as spindles may also appear on the EEG. The EOG activity present at N2 has disappeared, and the EMG activity present at N2 is still present, but with reduced amplitude. This stage is referred to as "deep" or "slow-wave" sleep.

The REM sleep stage is mainly characterized by the presence of rapid eye movements (REM), in bursts, which are different from the blinks of the wake stage. The EEG has a low amplitude activity with mixed frequencies, close to those of the N1 stage, but the EEG is associated with alpha bursts (often 1-2 Hz slower than the frequency of the wake state) and theta waves with triangular or pointed shape (between 2 and 6 Hz). The EMG of the chin shows a low muscle tone, lower than in all other stages of sleep. However, a rapid activity symbolized by irregular flashes sometimes occurs on the background of this hypoactivity. This stage is often referred to as "active sleep".

Characterization of drowsiness using PSG

In this thesis, as we are particularly interested in the characterization of drowsiness of an individual, we focus on stage N1 that has been described above. However, the characterization of the N1 stage is different for an individual lying in a bed with eyes closed or for an individual doing any other particular task with eyes open. The N1 stage is particularly characterized by an increase of the spectral power in the theta band. In 2005, Lal indeed showed that there was a high reproducibility of drowsiness detection based on theta band [57]. However, several studies showed that, for individuals with eyes open, an increase of the spectral power in the alpha frequency band is also observed when drowsiness occurs [58, 59]. This was confirmed by several studies that highlighted that drowsiness is characterized by an increase in electrical activity in the theta and alpha frequency bands [48, 60, 61]. Otmani *et al* showed that sleep deprivation and driving duration have an effect on alpha and theta activities [14]. In 2007, Papadelis demonstrated that severe driving errors are linked to the occurence of brief bursts of alpha activity [62]. Hori *et al* also demonstrated that the alpha frequency band moves from occipital regions towards the frontal regions in a state of drowsiness [63]. Moreover, drowsiness in active individuals also generates slow eye movements. Some studies were also performed in driving simulators and their results are indisputable: alpha and theta activities, slow eye movements, and lateral variability increase before driving off the road [64, 61, 14, 65]. Sandberg and his colleagues also investigated the effect of drowsiness during real driving and they confirmed that alpha and theta activities and blink duration were increasing with drowsiness [66].

The presence of micro-sleep episodes, reflected in the EEG by brief episodes of theta waves or diffused delta waves, is also often considered as a parameter indicative of the onset of sleep. During these episodes, a loss of involuntary attention comes and that

can lead to an inability to correctly react or respond to a stimulus. These micro-sleep episodes were observed to be correlated with poor performance at the wheel of a driving simulator [67].

However, Thomas et al. conducted a study of 66 individuals, partially or totally sleep deprived, at the wheel in a driving simulator with EEG monitoring [68]. Of the 619 accidents recorded, only 1% had at least a micro-sleep occurence in the EEG and 14% of them presented a micro-sleep during the one minute period before the accident. Thus, the appearance of a micro-sleep cannot be a unique marker of drowsiness as it may come too late.

For the development of a drowsiness characterization system, the ideal would be to have the equivalent of sleep states/stages for the period of drowsiness. However, drowsiness has been less studied and less standardized, as compared to sleep in the literature, and there are no universally-recognized stages for drowsiness, as there are for sleep.

However, several attempts were made to try to partition the drowsiness state into several stages. In 1983, Valley and Broughton were among the firsts to determine states of drowsiness from PSG parameters [69]. They divided the N1 sleep stage in two parts.

In 1994, Hori and Tanaka proposed nine (9) states designated H1 to H9 [2]. Figure 2.4 shows these nine (9) states: H1 and H2 correspond to the wake state, H3-H8 to the N1 stage, and H9 to the N2 stage. According to the opinion of several experts, it seems that the drowsiness usually passes through this sequence of states (H1 to H9) from wake to sleep.

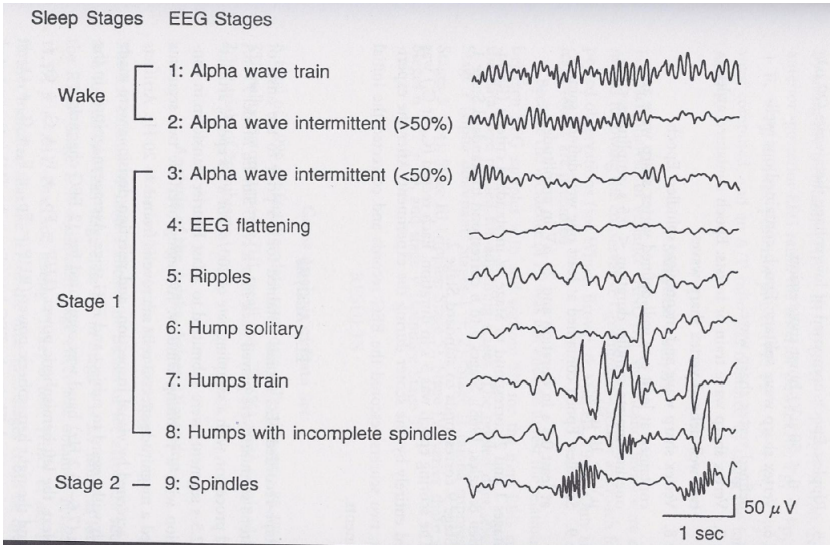

Figure 2.4: Illustration of the nine (9) EEG-based states proposed by Hori and Tanaka and their correspondence to the standard sleep stages [2].

However, because the experiment described begins with an individual lying down and having his eyes closed, and because it uses an auditory stimulus, Johns doubts about the overall value of this model for the characterization of drowsiness for individuals performing a visual task such as driving a car [70].

These two attempts at introducing the notion of stages for the condition of drowsiness, represent a major advance in knowledge but their relevance in characterizing drowsiness while executing a task is not certain.

From the late 90s, other "scales" were also developed to determine the state of wakefulness or drowsiness of an individual in "active situations" (operational settings) and at a given time: the Karolinska Drowsiness Scale (KDS) and, later, the Objective Sleepiness Scale (OSS).

The KDS was developed in Sweden in 1996 by Gillberg *et al* [60]. In this scale, the EEG is used to detect the presence of alpha activity and theta activity, and the EOG is used to detect slow eye movements (SEMs). The KDS scoring method is based on the scoring rules from Rechtschaffen & Kales (1968) [55]. The method consists in dividing the data in successive windows of 20 seconds and visually determining a (KDS) score. Each KDS score is based, as explained above, on the visual detection of signs of drowsiness (i.e. the presence or absence of alpha activity and/or theta activity in the EEG recording and/or SEMs in the EOG recording). Specifically, an expert manually/visually (as opposed to automatically) analyzes the EEG and EOG signals to detect the presence of signs of drowsiness and to determine a KDS score. The KDS score determined for each 20-second window is a numerical value between 0 (well awake) and 100 (very drowsy).

According to a study made by Anund and her colleagues, the KDS scale is a reliable scoring method to determine drowsiness-related performance decrements, particularly in driving situations [65].

The OSS was developed in France in 2003 by Muzet and it is composed of 5 stages. As for the KDS, the OSS scale uses the EEG to detect the presence of alpha activity and theta activity, and the EOG to detect slow eye movements (SEMs). The difference with the KDS is that the OSS scoring method is based on the duration of alpha and theta activities in a 20-second window as well as on the speed of eye movements in the 20-second window [71]. The OSS score determined for each 20-second window is a numerical value between 0 (well awake) and 4 (very drowsy).

Since the KDS appears to be more accurate (11 LoDs), and since it includes all the relevant drowsiness detection criteria discussed above, we choose, in this thesis, the KDS as the reference for the visual analysis of drowsiness from PSG signals.

Ocular parameters (OPs)

The activity of the eyes and their surrounding features (such as the eyelids) are the re-

flection of the activity of the brain, as a consequence of the fact that the behavior of all these entities is related to the central nervous system. Therefore, the measurement of ocular parameters (OPs) - related to the movement of the eyes and surrounding features - appears to be the basis of one of the most appropriate methods for characterizing (the level of) drowsiness of an individual. It is indeed well-known that several OPs are indicative of the LoD of an individual [72]. For each eye, most of the OPs fall into one of two categories: those related to the motion of the eyelids (including blinks), and those related to the motion of the eyeball (including saccades). Other relevant OPs exist, such as the size of the pupil.

In 1994, Dinges and Wierwille were among the firsts to study and test drowsiness characterization techniques appropriate for driving based on OPs, and particularly on the PERCLOS (PERcentage of eye CLOSure) [73, 74, 75, 76]. The PERCLOS, which computes the percentage of time that the eye remains closed more than 80% or 70% (depending on the definition found in the literature or chosen arbitrarily), does appear to be a relevant indicator of drowsiness [77, 78]. In 1994, based on the deterioration of the driving performance under sleep deprivation, Wierwille and colleagues divided the range of values of the PERCLOS parameter into three (3) different states: alert, questionable, and drowsy.

Since the time of the investigation by Dinges and Wierwille, many other OPs were found to be relevant indicators of drowsiness, especially, the quantities related to blinks [79, 12, 80, 81]. Indeed, several researchers showed that drowsiness increases the duration of blinks and their frequency. In 2008, Schleicher and his colleagues confirmed this finding after conducting a review of many OPs indicators of drowsiness [72]. They presented a driving test method using a driving simulator where the user was asked to give a feedback on his level of vigilance. The results of their study is a ranking of the most relevant OPs to characterize drowsiness; the duration of blinks is the highest ranked.

The parameters related to the closing and reopening speeds of the eyelids, as well as the duration of the reopening of the eyelids, also seem to be a wise choice to characterize drowsiness [82, 72].

In 2006, Johns proposed to use an amplitude-to-velocity ratio as a decisive measure to characterize drowsiness [83]. This parameter is the ratio between the amplitude and the maximum velocity of the closing phase or, of the reopening phase, of a blink. Through his patents [83, 84, 85, 86], he presented a new drowsiness scale composed of 11 stages and based on a weighted combination of OPs measured by infrared reflectance oculography, called, in this thesis, optooculography (OOG).

In 2013, to further confirm the reliability of eyelid-related OPs, Wilkinson and her colleagues conducted a study to demonstrate the reliability of eyelid movement parameters to characterize drowsiness. These parameters were measured at different times and in different rest conditions (including sleep deprivation) and compared to the OSLER test results. The conclusion indicated that OPs, such as the average duration of eye closures,

are promising real-time indicators of drowsiness [87].

Other publications consider OPs that are related to saccades rather than blinks, and that are also good indicators of drowsiness [88, 89] but saccades are still less often considered than blinks.

OPs related to the pupil also demonstrated promising efficiency in the characterization of drowsiness [90]. Indeed, there are spontaneous oscillations in the size (diameter) of the pupil with the appearance of drowsiness [91]. However, the pupil diameter is also influenced by other factors such as changes in brightness, stress, or nervousness, taking certain medications and/or drugs, etc. It is therefore a parameter of interest when the measurement conditions are controlled (in laboratory with a constant light). This parameter indeed becomes unreliable in operational conditions.

There are different techniques for calculating OPs: OOG, POG, and EOG. We have already discussed EOG in the section concerning PSG. OOG is based on the transmission of IR pulses and analysis of the returned signal. This technique is also called the infrared (IR) reflectometry of the eye. The POG is based on the acquisition and analysis of video images of the eye.

Facial expressions

Some facial expressions are also useful indicators of the state of alertness or drowsiness of an individual. Yawning is indeed recognized as a sign of drowsiness and it can be detected through the analysis of face images from a remote camera [92]. Computer vision systems are now sufficiently advanced to detect and recognize a face in an image and to bring out several feature points from the face including points around the mouth, and the eyes, among others. However, yawning does not always occur before a drowsy event, so that it must be used in conjunction with other parameters like inner brow rise, outer brow rise, lip stretch, or eye blink [93]. These last four parameters can be measured using the facial action coding system (FACS), which describes facial expressions in terms of several component movements.
Other techniques, sometimes simpler, have also been developed. In 2013, Mbouna and his colleagues, for example, proposed to use eye state and head pose to monitor drowsiness [94]. In 2014, Nakamura and his colleagues proposed measuring the distance between feature points of the face and textural changes in face images to detect changes in alertness and drowsiness [95].

Skin conductance

The skin conductance, which is also called electrodermal activity or galvanic skin response, is a measurement of the electrical resistance between two electrodes placed on the surface of the skin. This resistance varies with the sweating, which is controlled by the sympathetic nervous system, which itself drives our emotional and cognitive states, including the state of alertness/drowsiness. Indeed, changes in the autonomic sympa-

thetic nervous system activity (e.g. when we move from the alert state to the drowsy state) alter the level of sweat, which in turn affects the skin conductance [96]. The skin conductance is thus an indicator of drowsiness, and it can easily be measured by electrodes placed on the skin but it has the disadvantages of being influenced by the weather including temperature and humidity, and by other cognitive states such as stress and anxiety. In 2009, Bundele and his colleagues showed that the measurement of the skin conductance combined with oximetry can be used as indicators of drowsiness, and they even proposed to combine them with other physiological indicators, like respiration rate, to improve the accuracy of the drowsiness characterization [97].

Heart rate (HR)

The heart rate (HR) is another indicator of the state of alertness/drowsiness of an individual. Indeed, when drivers are drowsy, the HR decreases and the heart rate variability (HRV) increases. The HRV refers to beat-to-beat alterations in the HR signal. The HRV can be analyzed both in the time domain by computing the standard deviation of the beat-to-beat intervals, and in the frequency domain by computing an estimate of the power spectral density of the HR signal. When analyzing the HR signal in the frequency domain, one can indeed distinguish between a low frequency (LF) part [$0.04 - 0.15$ Hz] and a high frequency (HF) part [$0.15 - 0.4$ Hz]. The ratio between the LF part and the HF part was found to decrease with an increasing LoD [98, 99]. However, the HR is also influenced by factors other than drowsiness, like age, health condition, and body movements.

Performance

The performance in the execution of a task refers to how well, or poorly, an individual executes a task at a given time.

Just as there are PSG and POG parameters that are indicative of drowsiness, there are parameters that are indicative of the quality (goodness or badness) of task performance. We refer to them as task performance parameters.

While there are ways to describe the state of drowsiness via a single number, it seems that, at the time of this writing, one cannot find in the literature any paper/report describing a method for characterizing the task performance via a single number, including for the task of driving.

In the literature, however, one can find some tests/tasks that enable one to evaluate the performance of individuals and to use these tests/tasks as a reference in order to assess the state of an individual. Among them, we find mainly the Psychomotor Vigilance Test (PVT) and the Johns Test of Vigilance (JTV).

Psychomotor Vigilance Test (PVT)

The Psychomotor Vigilance Test (PVT) is a performance test that estimates the ability of an individual to detect and respond to a small change in an environment conducive to sleep. The PVT requires the individual to react to a visual stimulation. It is based on the assumption that drowsiness increases the reaction time (RT) and the number of omissions (i.e. failure to respond in time to a stimulus). The extracted values are generally the statistics of omissions and lapses (with a lapse usually defined as a RT greater than 0.5s) and the RT [100, 20, 101].

The PVT consists in presenting, for 10 minutes, a visual stimulus (a yellow counter) on a screen at random time intervals, between 2 and 10 seconds. The individual must respond by pressing a button (or clicking on a "mouse") as soon as the stimulus appears. In general, the entire test is performed repeatedly, at regular intervals of time, for a given individual, in order to see the expected increase in RT due to sleep deprivation. It has indeed been shown in the literature that the PVT is very sensitive to sleep deprivation, extended wakefulness, and time on task. Moreover, the PVT has the advantage of being unaffected by inter-individual variability and by learning [102].

The PVT is thus a simple test that is very popular and very reliable, even when performed repeatedly. Furthermore, this test was validated against the EEG. Therefore, because of its simplicity of implementation, the PVT is often used as secondary reference.

Johns Test of Vigilance (JTV)

In 2007, Johns introduced the Johns Test of Vigilance (JTV) [85, 103]. The JTV is based on a modified version of the PVT. This modification makes the new test suitable for use with a system that records eye and eyelids movements.

The JTV is a reaction time test. In the JTV, the individual must push a button that he holds in his hand (left for left-handed, and right for right-handed) as quickly as possible after the onset of a visual stimulus. The visual stimulus consists in a change in the shape of an object. In this case, three circles become squares or diamonds for 400 ms. The three circles appear at random time intervals, between 5 and 15 seconds. The JTV usually lasts between 10 and 15 minutes and includes, on average, between 65 and 90 stimuli.

In the JTV, two types of error are identified:

- The absence of answer (omissions or lapses): the lack of response 2000 milliseconds after the onset of the stimulus;
- The late answer (or moderately delayed answer): a response after the disappearance of the stimulus, i.e. 500 to 2000 milliseconds after the onset of the stimulus.

Driving performance

Since this thesis focuses partly on driving, we also examine the driving parameters often used in the context of drowsiness at the wheel. By driving parameters, we refer to parameters that reflect the behavior and performance of the driver while driving. For example, drowsiness due to sleep deprivation was shown to induce greater variability in speed [104, 66, 105].

Another important measure is based on the movements of the steering wheel [106]. Researchers demonstrated that a drowsy driver makes fewer movements and more large movements to correct lane deviations [104].

In 2009, Liu and colleagues showed that the measure of the deviation/variability of the vehicule position on the road, also called standard deviation of the lateral position (SDLP) of the vehicule on the road, is the driving parameter that had the greatest correlation with drowsiness [104]. This parameter indeed measures the ability of a driver to maintain his trajectory, which is an indication of alertness. In 2013, Forsman and her colleagues evaluated 87 different metrics to detect moderate drowsiness and, thus, to be able to prevent the driver from having a high risk of an imminent accident [107]. They found that the most significant parameters were related to the variability of the motion of the steering wheel and to the SDLP. The latter had the greatest correlation with performance when performing a PVT. They also found that the SDLP could be derived from the steering angle changes via a transfer function. This interesting finding led, in 2004, to a US patent [108].

2.4 Analysis of the state of the art and justification of our approach

The main objective of this thesis is to develop an automatic, objective, and real time drowsiness characterization system that is usable in an operational environment, such as driving a car or a truck, without being invasive for the individual. Now that we have detailed the traditional methods for characterizing drowsiness, we justify the approach chosen to develop our system.

Below, we often refer to “present time”. This, of course, refers to the “present” in an actual scenario. It is also a synonymous for “now” in such a scenario. We also often use the term “real-time” to refer to the fact that we are interested in determining the LoD almost instantly, at this “present time”, i.e. “now”.

Sleep propensity methods such as MSLT, MWT, and ESS quantify the risk that an individual falls asleep at some future time under certain conditions. They are useful in determining whether an individual is likely to suffer from sleep disorders but they do not measure an LoD at the present time in operational conditions. Therefore, these methods cannot be used for real-time characterization of drowsiness.

Subjective methods such as KSS, SSS, and VAS have the advantage of being cost effective and, furthermore, they can be completed in a minimum of time and they provide a directly available (real-time) measure. However, these three methods require people to self-assess their drowsiness and, therefore, inevitably involve their participation, which is almost certainly distracting. Above all, these scales are based on a subjective assessment of the individual, and this self-assessment cannot be considered reliable for the following reasons. First, the assessment is influenced by the drowsiness experience/history and by other factors such as posture or activity. Second, individuals have sometimes difficulties to distinguish between drowsiness and fatigue, which are distinct phenomena. Third, if each individual could properly assess the moment when he becomes drowsy and no more able to perform a task, there would be no accidents related to drowsiness. It is furthermore obvious that a driver will not apply such a method when he is at a high LoD and seconds before a potentially deadly crash.

If we focus our efforts on improving safety, the key issue is to what degree the present performance or, more generally, its evolution over the recent past (up to present) is risky. As we have seen, there are different methods to measure the performance of an individual while he is performing a task. Some of these methods are related to the execution of psychomotor tests (e.g. PVT), which also require the participation of the individual. These tests are very relevant for assessing an individual's condition in a lab, such as for research purposes, but they constitute an interruption/distraction in relation to a main task that the individual should perform (driving, flying, navigating, monitoring critical instruments, etc.) and this is not really feasible in an operational situation. In addition, these tests always take a significant time (minimum 10 minutes). Another approach in the analysis of performance focused on driving through the measurement of various parameters related to the vehicle such as the SDLP or steering wheel movements. This approach is very useful to avoid some accidents but the driving parameters can be influenced by many factors other than drowsiness (e.g. traffic conditions, weather conditions, age, driver experience, etc.). In particular, there are lots of roads that have poor lane markers or no lane markers at all. And, in any case, the markings are invisible/undetectable when covered by snow and the like. But, even more fundamentally, this approach does not get down to the fundamental roots of "true" physiological drowsiness. Moreover, with the progressive appearance of increasing autonomy in vehicles, these parameters will become obsolete when the car drives on its own. Finally, there is no marking in the sky to measure some flying version of an SDLP.

This leaves us with methods based on physiology, which is perfectly logical since drowsiness is, after all, a physiological phenomenon.

PSG - consisting mainly of EEG and EOG - is considered by many experts in the world to constitute the reference. However, the use of these techniques for analyzing drowsiness requires the placement of electrodes on the individual. This process usually takes time, is binding, and requires a qualified individual to be able to do so. PSG signals are also often tainted by artefacts due to body or eye or electrode movements, for example.

In conclusion, at present, methods based on PSG are not very practical for a real-time drowsiness characterization system.

According to the literature, the measurement of ocular activity (at least one eye) seems to be one of the most suitable methods for characterizing an individual's LoD in operational settings, such as driving. Indeed, as detailed in Section 2.3.2, some OPs are scientifically recognized to be significant indicators of drowsiness.

Besides EOG, which is part of PSG, two approaches were presented in the OPs section: OOG and POG. Both techniques have the advantage of requiring neither any intervention from the individual, nor any invasive physical contact with him, making these two techniques ideal candidates for a drowsiness characterization system. However, the behavior of the eyes (movements of the eyeball and of the eyelids) is a phenomenon eminently spatio-temporal. In this context, it should be noted that OOG can have a very high temporal resolution (with up to 1000 or 2000 pulses per second) but no spatial resolution (i.e. no localization of the eye), and that POG has a very good spatial resolution (thousands of pixels) and good temporal resolution (up to a few hundred frames per second today, which is more than enough). Moreover, in 2013, Anderson et al. presented a study in which they assessed drowsiness based on OOG [109]. During their study, they faced losses of data mainly due to poor signal quality with the OOG technique or improper positioning of the OOG-based data acquisition system. The POG approach has the major advantage that it enables one to determine the exact positions of the eyelids and of the eyeball and thus to compute a lot of different OPs that are indicative of drowsiness. The best solution is therefore to use POG, i.e.to use video (image) sequences of at least one eye.

In the preferred context of POG-based systems, we must address the differences between head-mounted systems (as in the form of a pair of eyeglasses) and remotely-placed systems. The advantage of remote systems is that they do not require physical contact with the individual. However, head-mounted systems have their own advantages including:

- the ability to monitor the eye continuously, making the assessment of drowsiness much more reliable;
- much more spatial resolution on the eye given the small eye-camera distance;
- much easier control of lighting, and elimination of lighting disturbances;
- the mobile nature of the glasses (ease of transport, e.g. from one vehicle to another).

Even though, some people might object wearing a head-mounted system, such as specialized eyeglasses, one should note that they do not object to wearing sunglasses and even 3D glasses in cinema.

There are still other methods for assessing drowsiness such as the ones based on facial expressions, skin conductance, and heart rate. The last two are also interesting, but the

parameters they measure are not sensitive enough to determine a precise progression from alertness to drowsiness; it is often necessary to combine several sensors and measurements to obtain a good result. In addition, the parameters are often influenced by other factors such as age, climate, and activity. Facial expressions can provide useful information but, in most studies, it is the eyes that are analyzed to characterize drowsiness. There are also drowsiness detectors based only on head nodding. The occurence of such a head movement probably indicates that the individual has already had a micro-sleep ("microsleep") or is asleep. A true drowsiness characterization system cannot afford to wait to detect such a sign.

In summary, the best solution for achieving our goal, i.e. an automatic, objective, real-time, and reliable drowsiness characterization system, is to develop a head-mounted system based on images of at least one eye.

Now that the approach has been identified, we also need to examine and discuss how to evaluate/validate our drowsiness characterization system.

We were quite surprised to read in the literature, and to learn from experts, that there does not appear to be a universal agreement/consensus on the reference to be used to characterize drowsiness and thus to be able to evaluate drowsiness characterization systems.

Sleep experts would say that EEG is the reference since it is already the gold standard for sleep and since drowsiness is a physiological state between wake and sleep. However, cognitive psychologists will most likely argue that the goal of a drowsiness detection system is, in the case of driving, to avoid driving accidents, so that such systems should be evaluated against their ability to prevent accidents. Continuing with the case of driving, the system will be successful if it warns a driver just before he crosses the lines of his driving lane.

The above remark seems to suggest that drowsiness characterization systems, whether video-based or not, should be validated against their ability to prevent accidents, initially in a driving simulator for obvious practical and safety reasons.

We thus decided to compare our system against four different drowsiness references: PSG, PVT, driving parameters, and KSS. This would enable us to verify that the LoD produced by our POG-based drowsiness characterization system is meaningful, from the points of view of both physiology and performance.

Chapter 3

Description of our data and acquisition environment

We present our data acquisition environment as well as the data acquired including the ones that are used later for validation of our POG-based drowsiness characterization system. Section 3.1 describes the requirements that we established in order to obtain a relevant database for the development and the test of our system. Section 3.2 explains the protocol that we developed for the acquisition of data, the equipment used, and the data acquired. Section 3.3 provides a detailed description of the data that are used for validation later in Chapter 4. The data presented in this chapter are also used in Chapter 5 for the development and the validation of our auomatic PSG-based drowsiness characterization system.

3.1 Requirements for the development and test of our real-time drowsiness characterization system

In order to develop and then test (validate) our drowsiness characterization system based on images of the eye, it was essential for us to acquire real data and to have references indicating the state of drowsiness of an individual. The term "real" refers here to data coming from real people put in different conditions of drowsiness, and not data synthesized using a computer. However, this does not mean that the data has been acquired in real situations/environments. For several reasons that will be explained below, all our tests were performed in the laboratory.

The purpose of our system being to determine the objective, and physiological state of drowsiness, we needed, in addition to acquiring images of the eye, to collect data that can serve for validation to determine the state of an individual at the moment of testing. In Chapter 2, we identified polysomnography (PSG) as the best physiological reference. It is indeed the gold standard for the study of sleep, and thus by some extension, for the study of drowsiness. Acquiring PSG signals requires placing electrodes on the scalp, around the eyes, on the chin, and on the chest. Therefore, we trained at the sleep lab of the University Hospital of Liège in order to equip the participants ourselves.

In addition, as the main application of our system is to improve safety, we also wanted to acquire data related to the performance of individuals in the execution of a task. We indeed wanted the level of drowsiness (LoD) produced by our system to be well correlated with performance decrements so that our system could prevent these performance decrements when performing a risky task and alert the person. We thus also needed data that can serve for validation to determine drowsiness-related decrements in performance in the execution of a task. In Chapter 2, we also identified Psychomotor Vigilance Test (PVT) and driving in a simulator as the best performance references. As a reminder, the PVT is a validated test that is very sensitive to sleep deprivation and, that leads to statistics linked to reaction times, the latter being known to increase with drowsiness. We thus implemented a version of the PVT that is based on the one described by Dinges [100]. Regarding the driving task, the ideal would have been to ask participants to perform a task in real conditions, such as driving a real car or truck, but from an ethical point of view, it was not possible to acquire data from fully awake to asleep states without any risk to the participants. Furthermore, driving in a simulator offers the advantage of providing participants with identical and controlled driving situations, and providing the experimenter with a selection of many parameters (driving environment, traffic density, time of day, etc.). In particular, we decided to use a monotonous driving scenario without stimulation to encourage/induce drowsiness.

We thus decided, logically, to acquire eye image sequences, PSG signals, and performance data. To make relevant comparisons, all data had to be acquired synchronously. This aspect is detailed later.

In order to acquire drowsiness data, we analyzed the sleep/wake cycle (the interaction of the circadian and the homeostatic processes) to define the moments of the day the more conducive to drowsiness. This analysis showed that the beginning of the afternoon seemed like an excellent moment to carry out a data acquisition because the sleep pressure is higher due to the circadian process and, in addition, a data acquisition at that time was not difficult to organize. The night was also found to be a judicious moment as the sleep pressure is very high due to both the circadian process and the homeostatic process, even if this implies recruiting volunteers who agree to stay awake a few hours during the night.

Initially, we conducted several small data acquisition sessions with volunteers who spent the afternoon or the night in our laboratory to perform some tests (PVT or driving session in a simulator). These data enabled us to advance in the development of our system but we quickly realized that it was necessary to impose more constraints in order to have a complete and valid database to finalize our developments and validate our system.

Here is the list of requirements that we established:

Requirements related to the data acquired

We should acquire data with both low and high levels of drowsiness for a same person in order to get an interesting range of alertness and drowsiness data in our database. We thus decided to impose several tests for a same person at different times of the day and, especially, after both a good night of sleep and a long sleep deprivation without any consumption of stimulant.

We must know the sleep-wake "profile" of each participant. To do so, we decided to ask each participant to complete a sleep diary to estimate the average number of hours of sleep in a week period, and to answer the Epworth Sleepiness Scale (ESS) to evaluate his propensity to daytime sleepiness. These data are very useful both to categorize participants and to provide additional information for our analysis of results.

Assessing subjective drowsiness is also very important to find out how people assess themselves in different LoDs. Indeed, there is evidence that, at a certain LoD, individuals misjudge their levels of drowsiness. Many experts and authors like to use the KSS as a reference to determine the LoD. The KSS is indeed an instantaneous subjective drowsiness scale that is quite instinctive and easily assimilated by the participant. We thus decided to collect the KSS data and to use it as a comparison point with other references (polysomnography, PVT, and driving parameters), and with our POG-based drowsiness characterization system.

In order to compare data from more than one measurement system, the synchronization is very important. Studies like ours indeed often utilize a variety of specialized systems that are each specifically designed for one type of measurement. For example, in our case, we need to record PSG data as well as images of the eye. It is thus important to ensure that events detected in EEG recordings are identically time-stamped with the corresponding events in images of the eye. We must ensure that the acquisition systems start at the same time and remain time-aligned so that there is no delay between them even if, from a practical point a view, a slight delay may be acceptable depending on the application. In the present context, we know that a synchronization delay of 8 ms (corresponding to 125 Hz) would be virtually undetectable because the lowest sampling frequency is the one for the acquisition of images of the eye and is 120 Hz while the sampling rate for PSG signals is 512 Hz. However, as we analyze drowsiness using windows of 60 seconds, delays up to one or two hundred milliseconds may be acceptable. We have thus synchronized the clocks of all the computers that we used for our study (i.e. one computer for the recording of images of the eye, one computer for the recording of PSG signals, and one computer for the recording of performance parameters). We also implemented a "triggering system". This is accomplished by sending one electrical pulse to one channel in the PSG data acquisition system each time a set of 120 images (1 second) is recorded by the POG-based data acquisition system. As we also record the time stamp for each image, we can easily determine if there is any delay between the two acquisition systems and correct it.

Requirements related to the testing environment

The environment should be the same for all participants. During the tests, we decided that each participant had to stay alone in a quiet room with no temporal cues (watch, smartphone, etc.). In order to induce more drowsiness, we also decided to turn the lights off in the room for the "sleep-deprived" tests. Before each test, it is also important to control the conditions under which the participant is in order to minimize the influence of external factors when analyzing the data. The participant was thus instructed to remain calm without distractions for half an hour before each test.

Based on all these requirements and on our experience with other small data acquisition sessions, we developed a data acquisition protocol. This protocol was used for two experiments, called A and B. In Experiment A, each participant conducted three PVTs at different times of the day and of the night under progressive sleep deprivation. In Experiment B, each participant performed three driving sessions in a driving simulator at different times of the day and of the night under progressive sleep deprivation. The resulting protocol is described in the following section.

3.2 Data acquisition in laboratory

3.2.1 Protocol of Experiment A

We recruited thirty healthy participants by placing announcements on social networks and on notice boards at the University of Liège. We selected the participants based on several criteria:

1. they should be between age from 18 to 35 years,
2. they should have no drug or alcohol addiction,
3. they should have no sleep pathology,
4. they should have no jet lag during the two preceding weeks,
5. they should not be shift workers.

These criteria were put in place to analyze a control and healthy population whose state of alertness and/or drowsiness is not influenced by other factors. We actually wanted to obtain data from individuals having a normal sleep/wake cycle.

In addition, each participant should be able to conduct all phases of the experiment without using prescription glasses. The protocol indeed included wearing a data acquisition equipment in the form of a pair of eyeglasses that does not fit with prescription glasses. However, the contact lenses were allowed.

The experiment consisted in performing three PVTs - each of 10 minutes in duration - in different sleep deprivation conditions over two consecutive days. Each participant

was progressively sleep deprived from the first PVT until the end of the study.

Prior to the study, each participant received an information sheet on the study, instructions, and three documents to fill in:

- a questionnaire to verify that he fits the inclusion criteria to participate in the study,
- a sleep diary to fill in for the week before the study,
- the Epworth questionnaire (ESS).

Each participant received also the brochure "Somnolence au volant - Une étude pour mieux comprendre (ASFA, June 2010)" to raise awareness of the issue of drowsiness at the wheel.

For ease of explanation, the experiment can be viewed as consisting of the succession of Night 1, Day 1, Night 2, and Day 2, and of three PVTs (PVT 1, 2, and 3). Figure 3.1 provides an illustration of the protocol. The participants provided a written, informed consent at their arrival to our laboratory.

On Night 1, each participant slept at home and was asked to have a full night sleep (of 7 to 8 hours at least). Then, the participant was not allowed to sleep from the time he woke up on Day 1 until the end of the study (noon on Day 2). At 8:00 am on Day 1, the participant arrived at our laboratory and was equipped with the PSG electrodes. Between 9:00 and 10:00 am, he performed PVT 1. He was then free to leave the laboratory to carry out his normal activities but was equipped with an actimeter in order to check that he had not slept. The participant came back to our laboratory at 8:30 pm on Day 1 and was again equipped with PSG electrodes. On Night 2, he performed PVT 2 between 2:00 and 3:00 am and, after breakfast on Day 2 (provided by us), he performed PVT 3 between 11:00 am and noon (and after at least 28 h of sleep deprivation). At the end of the study, the participant was sent back home. He was asked not to consume any stimulant (coffee, tea, etc.) from noon on Day 1 until the end of PVT 3. When the participant did not perform a PVT and was present in the laboratory, he could go about his business and do as he wanted in the common room. However, he remained mostly in a sitting position with controlled brightness in the room. We instructed him to remain calm, without any computer, telephone, or any too distracting object during the half hour preceding each PVT test.

Before each test, the participant was also asked to evaluate his LoD using the Karolinska Sleepiness Scale (KSS).

During a PVT test, the participant was alone in a quiet room without any temporal cues (watch, telephone, etc.). For the first PVT, the light in the room was on. However, for the other two tests, the light was off.

At the end of the study, the participant was in a sleep-deprived condition, so that there was a risk if he drove a vehicle in the hours following the last PVT. We informed him of the risk and strongly advised him not to drive to return home. In addition, we covered public transportation costs (bus, taxi within 20 kms, train, etc.) to return home at the end of the study.

The above protocol was approved by the Ethics Committee of the University of Liège.

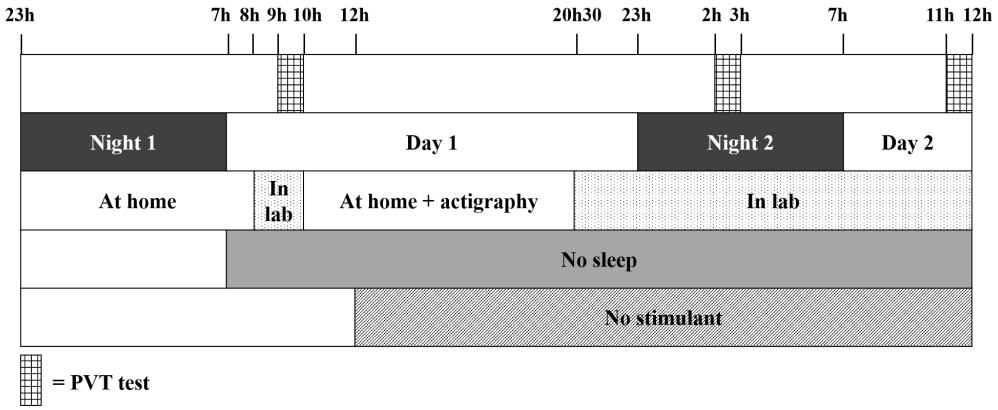

Figure 3.1: Data acquisition protocol (Experiment A).

The condition during the morning of Day 1 (PVT 1) is referred to as "not-sleep-deprived", and the conditions during and after Night 2 (PVTs 2 and 3) as "moderately-sleep-deprived" and "sleep-deprived", respectively.

3.2.2 Protocol of Experiment B

For Experiment B, the protocol was identical to the one of Experiment A except that, instead of performing a PVT, each participant was subjected to a driving session in a professional driving simulator. The experiment thus consisted in performing three driving sessions in different sleep deprivation conditions over two consecutive days. Each participant was progressively sleep deprived from the first driving session until the end of the study. The first driving session on the morning of Day 1 (at the same hour of the day as in Experiment A) lasted 45 minutes and, as for the PVT tests, the light was on. The second driving session was performed during Night 2 at the same hour of the day as in Experiment A) with a duration of 1 hour and with light off. Finally, the third driving session was done at the end of the morning of Day 2 at the same hour of the day as in Experiment A) for 45 minutes and with light off.

We recruited twelve healthy participants in Experiment B and we selected them based on the same criteria as in Experiment A.

3.2.3 Material and data acquired

For both Experiment A and Experiment B, the material and the data acquired included:

- Images of the right eye at a rate of 120 Hz using a pair of eyeglasses developed by our team at the University of Liège and then transferred to the company Phasya. These eyeglasses include:
  - an infrared (IR) illuminator (LED),
  - a hot mirror that reflects IR light/radiation,
  - a small high speed camera sensitive in the IR.

  Thanks to the use of IR, this image aquisition device is usable in any lighting conditions. The pair of eyeglasses is linked to a computer via a USB cable to record the images.
- PSG signals at a rate of 512 Hz using a PSG acquisition system called Embla Titanium, with related electrodes. The Embla Titanium is a type of compact ambulatory PSG system that is easily transported and that enables the recording of 34 channels (EEG, EOG, EMG, ECG, ect.). The device operates in either of two modes: either it sends data directly to the computer (i.e. online recording) or it records up to 33 hours of data on a flash card. In our tests, we performed the acquisition online and we decided to record the following channels:
  - EEG: Fz, Cz, Pz, C3, C4, and A1.
  - EOG: 2 channels for vertical-EOG and 2 channels for horizontal-EOG.
  - EMG: 2 channels.
  - ECG: 2 channels.
- Actimetry data using an actimeter called Actiwatch 2 from the company Philips Respironics. The actimeter is a small device in the form of a watch that is worn on the non-dominant wrist. Inside this device, there are a piezo-electric cell that detects acceleration of movements, an electronic chip that records the pulses generated by these accelerations, and a light sensor that records light intensity. This device thus gives us a range of information on the sleep-wake cycle of the individual thanks to the recorded features. From the end of Test 1 (PVT 1 or driving session 1), the participant was provided with the actimeter, and he wore it until he returned to the laboratory for Test 2 (PVT 2 or driving session 2). The actimeter enabled us to verify that he had no slept during the day outside the laboratory.

In addition, for Experiment A, we also recorded:

- Data from the PVT. This test uses a computer with monitor, keyboard, and mouse. The data extracted from the PVT is the time that the participant takes to react to the appearance of a stimulus on the screen, measured in seconds (s).
- Subjective drowsiness levels using the KSS.

For Experiment B, we also recorded:

- Data from the driving simulator sessions. This test uses a professional driving simulator including driving seat, pedals, steering wheel, gear lever, screen, and computer, on which is installed a professional driving simulation software that we obtained from the French institute IFSTTAR. The data extracted from the driving simulation task is the standard deviation of the lateral position (SDLP) of the vehicle on the road, measured in millimeters (mm).

All data were collected anonymously.

3.3 References for performance evaluation

We present and analyze the different references that we used to evaluate the performance of (i.e. to validate) our automatic and real-time POG-based drowsiness characterization system and that we describe in the next chapter.
Specifically,

- we describe the physiological reference, which is based on the visual analysis/scoring of PSG signals,
- we detail the references related to the performance of each participant in the execution of a task, and
- we present the subjective reference, which is based on the KSS.

According to our conclusion in Chapter 2, these references are the most relevant to assess an individual LoD. To verify this, we demonstrate here the influence of sleep deprivation on these data. To do this, we consider the data acquired for all participants during PVTs and driving sessions and we divide the data according to the three sleep conditions, i.e. not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, as previously defined. We then compute several statistics to show the effect of sleep deprivation on the data. We also perform several analyses of variance to determine whether there are differences between individuals. Since the data coming from all references are not normally distributed, we decided to use the Kruskal-Wallis method to perform the analyses of variance. The Kruskal-Wallis test is a non-parametric alternative to the one-way ANOVA. It is used to compare several samples, and to test the null hypothesis that the samples to be compared are from the same distribution or distributions of the same median.

To visualize the results, we decided to use box plots to illustrate the dispersion of data within groups.

3.3.1 Analysis of physiological reference (visual scoring of PSG signals)

Method

Usually, the study of sleep is based on the recordings of EEG, EOG, and EMG signals. Experts then analyze the signals and try to:

- recognize specific features/patterns contributing to the definition of a sleep stage;
- classify/score a window of 30 seconds according to the following rule: when at least 50% of the window is characteristic of a stage, then the entire window is scored as belonging to this stage. (Successive windows are typically butting, i.e. non-overlapping.)

The EEG is described in terms of frequency activity. As already explained in Chapter 2, this frequency activity is divided into the following bands:

- Delta activity (δ): [0.5-3]Hz;
- Theta activity (θ): [4-7]Hz;
- Alpha activity (α): [8-12]Hz;
- Beta activity (β): [13-25]Hz;
- Gamma activity (γ): $>$ 25Hz.

The study of sleep is relatively well standardized and, today, the rules of the AASM [56] serve as a reference to identify the different stages of sleep. However, for the study of drowsiness, one must proceed differently. One can find some "scales" in the literature but there is no standardized protocol for distinguishing between drowsiness levels as it is the case for sleep.

After a review of literature, we chose to use the Karolinska Drowsiness Scale (KDS) for our data analysis. It is indeed, in our opinion, the most relevant scale as it includes most of the criteria usually found in the literature to characterize drowsiness, i.e. an increase in alpha and theta activity and slow eye movements.

As already mentioned, the KDS was developed by *Å*kerstedt et al. to help to determine the state of drowsiness of an individual in "active" situations (operational settings) at a given time. In this particular scale, the EEG and the EOG are used to detect signs of drowsiness, i.e. the presence of alpha activity and/or theta activity, and/or slow eye movements (SEMs). The KDS scoring method is based on Rechtschaffen and Kales (1968) scoring rules [55], and consists in visually determining a KDS score for each successive (i.e. butting) 20-second window based on the presence of signs of drowsiness. The procedure involves dividing a 20-second window in 10 sub-windows of 2 seconds each. If one detects the presence of at least one sign of drowsiness in any 2-second

sub-window, the score for the window is incremented by 10. The KDS score determined for each 20-second window is thus a numerical value between 0 (well awake) and 100 (very drowsy).

We found it difficult, in our work, to apply the KDS method for the visual scoring of PSG signals. Indeed, the "criteria" described in the literature are rather vague and very much subject to interpretation. We therefore decided to refine the criteria with the help of several experts in sleep scoring in order to obtain a more accurate and robust visual analysis of PSG signals.

The new criteria that we developed for the detection of each sign of drowsiness are as follows:

- Theta activity:
  - frequency: between 4 and 7 Hz;
  - amplitude: at least $25\mu V$ peak to peak;
  - duration: at least 1 second in the 2-second sub-window;
- Alpha activity:
  - frequency: we introduce the notion of rhythm that is different from our point of view from activity. The alpha rhythm consists of sinusoidal-like waves with frequencies in the alpha range [8-12] Hz while the alpha activity consists of nondescript (non-sinusoidal) waves with frequencies in the alpha range. The notion of alpha rhythm is not much detailed in the literature but it is a distinction that we want to make in this thesis. The drowsiness sign is here linked to the presence of alpha rhythm and we will now, throughout, use the term alpha rhythm to refer to the alpha-related drowsiness sign;
  - amplitude: at least $25\mu V$ peak to peak, (as for theta activity);
  - duration: at least 1 second in the 2-second sub-window; (as for theta activity).
- Slow eye movements:
  - amplitude: at least $100\mu V$ peak to peak;
  - duration: at least 500 ms in the 2-second sub-window, which is considered to be a slow movement.

In our study, we trained ourselves, visually analyzed the data, and manually assigned a KDS score to each 20-second window, which we decided to call here a visual PSG-based LoD.

Figure 3.2 gives an example of the grid, called scoring sheet, that we established for scoring the PSG signals during an entire PVT. The duration of each PVT is 10 minutes. There are thus 30 windows of 20 seconds, which we subdivided into 2-second

sub-windows. When a sign of drowsiness is detected in one of the sub-windows, the scorer notes it, and the score of each 20-second window is directly and electronically incremented by 10.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Protocol		PVT			KDS Scoring Sheet				Legend			
2	Subject		1			Scorer				alpha	a	spindle	s
3	Session		1			Scoring date				theta	t	vertex	v
4	Starting time		9:00:00							eye	e	K-comp	k
5													
6	Epoch	Time	Score	1	2	3	4	5	6	7	8	9	10
7	1	9:00:00											
8	2	9:00:20											
9	3	9:00:40											
10	4	9:01:00											
11	5	9:01:20											
12	6	9:01:40											
13	7	9:02:00											
14	8	9:02:20											
15	9	9:02:40											
16	10	9:03:00											
17	11	9:03:20											
18	12	9:03:40											
19	13	9:04:00											
20	14	9:04:20											
21	15	9:04:40											
22	16	9:05:00											
23	17	9:05:20											
24	18	9:05:40											
25	19	9:06:00											
26	20	9:06:20											
27	21	9:06:40											
28	22	9:07:00											
29	23	9:07:20											
30	24	9:07:40											
31	25	9:08:00											
32	26	9:08:20											
33	27	9:08:40											
34	28	9:09:00											
35	29	9:09:20											
36	30	9:09:40											

Figure 3.2: Example of a scoring sheet for PSG signals based on the KDS method.

In order to facilitate the comparison with later results, we decided to divide the PSG-based LoD so obtained for each 20-second window by 10 to obtain LoDs ranging from 0 (well awake) to 10 (very drowsy).

Results

We show the results of the analysis of the PSG signals during the PVT tests, in particular:

- the evolution of the visual PSG-based LoD for the three tests:
  - Test 1: PVT 1, corresponding to the not-sleep-deprived condition;
  - Test 2: PVT 2, corresponding to the moderately-sleep-deprived condition;
  - Test 3: PVT 3, corresponding to the sleep-deprived condition;
- the variations between individuals.

Figure 3.3 shows the evolution of the visual PSG-based LoD as a function of time for the three tests. Each curve in the figure represents the mean value of this LoD for all participants, and we can observe that, on average, the visual PSG-based LoD increases with sleep deprivation (i.e. from Test 1 to Test 2 and even more from Test 2 to Test

3) but only very slightly with time. Figure 3.4 also indicates an increase of the visual PSG-based LoD as a function of the three tests. Moreover, if we compare the mean of all visual PSG-based LoD for all participants, we can observe that this mean increases from 0.72 ± 1.23 in not-sleep-deprived condition (Test 1) to 1.35 ± 1.84 in moderately-sleep-deprived condition (Test 2), to 2.33 ± 2.13 in sleep-deprived condition (Test 3). The mean values of the visual PSG-based LoD during the PVT 2 and PVT 3 seem low compared to the maximum value they can have (10) but this can be explained by the fact that low KDS scores are already found to be associated with performance decrements. Anund and colleagues indeed showed in one of their studies that there was already a significant change in the variability of the lateral position (also called SDLP) of the vehicle on the road for KDS ≥ 2 [65].

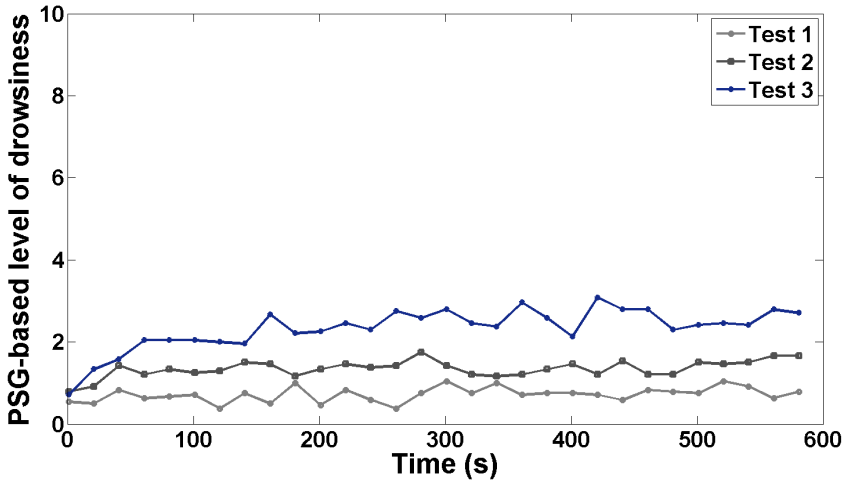

Figure 3.3: Evolution of the mean PSG-based LoD as a function of time for PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively).

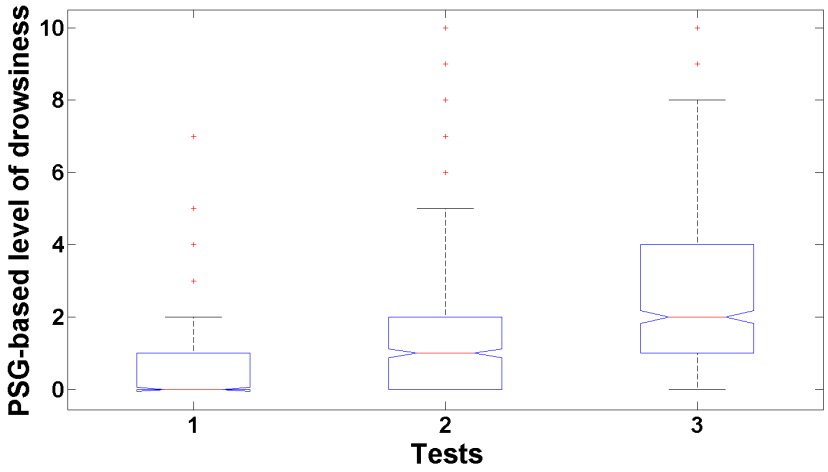

Figure 3.4: PSG-based LoD as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

We also wanted to see the distribution of the signs of drowsiness detected in PSG signals as a function of the three sleep conditions. Figure 3.5 shows the result. We can observe that the proportions of slow eye movements and of alpha rhythm increase, on average, with sleep deprivation, for all participants. However, the theta activity remains very low but this is probably due to the fact that the conditions in which the participants are (i.e. sitting in front of a computer) are not likely to encourage theta activity, as opposed to the case where they are lying in a bed while falling asleep. Moreover, the theta activity was the most difficult sign to detect in our data, probably due to noise in the PSG signals (i.e. artifacts).

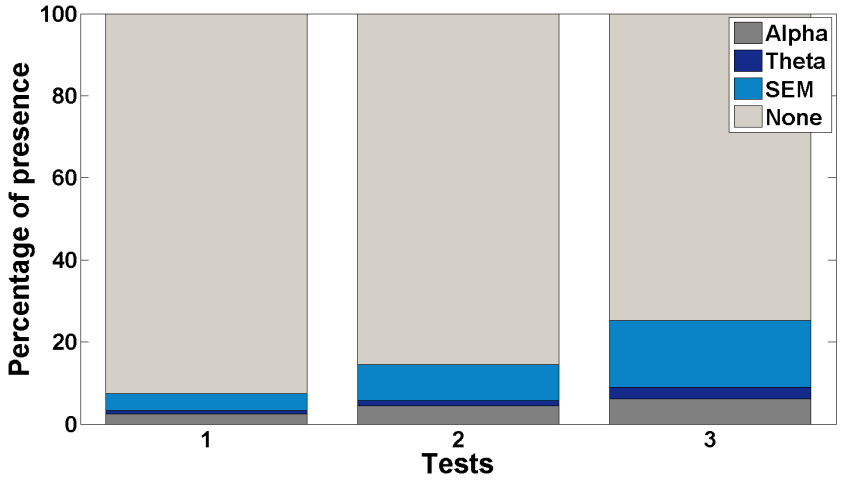

Figure 3.5: Proportions of signs of drowsiness as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively) for all participants. Alpha refers to the proportion of alpha rhythm. Theta refers to the proportion of theta activity. SEM refers to the proportion of slow eye movements. None refers to the proportion of no sign of drowsiness.

We also performed a Kruskal-Wallis analysis on visual PSG-based LoD to distinguish differences between participants. For this analysis, we used data from 24 participants because we faced some losses of data and we needed to analyze the three tests for each participant. This analysis led to the following results. There are significant differences of visual PSG-based LoD between participants, regardless of their sleep condition ($\chi^2 = 677.23; p < 0.01$), and between participants when they are non-sleep-deprived ($\chi^2 = 309.66; p < 0.01$), moderately-sleep-deprived ($\chi^2 = 382.82; p < 0.01$), and sleep-deprived ($\chi^2 = 352.2; p < 0.01$). The boxplot in Figure 3.6 indicates the differences of the visual PSG-based LoD between participants regardless of their sleep-deprivation condition (i.e. for the three tests combined).

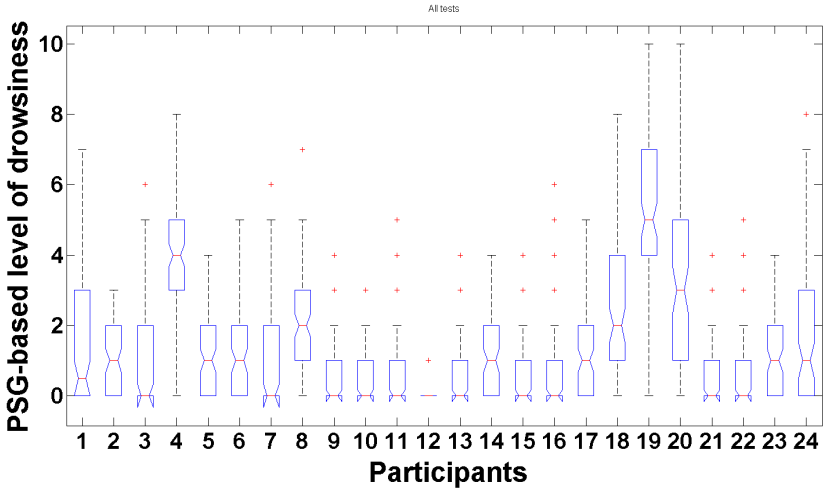

Figure 3.6: Differences of PSG-based LoD between subjects regardless of their sleep deprivation condition. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

3.3.2 Analysis of task-performance references

We successively consider the two performance references:

- the performance in the accomplishment of a PVT;
- the performance in the accomplishment of a driving session.

A. PVT

Method

We assessed the performance of each participant in the accomplishment of a task using the PVT. The PVT is indeed, as already mentioned in Chapter 2, a performance test that is sensitive to sleep deprivation and that estimates the ability of individuals to sustain attention and to respond to visual stimuli in an environment conducive to sleep. The PVT enables the measurement of reaction times (RTs), which increase with drowsiness.

The PVT used in our study is our own implementation of the PVT defined by Dinges [100]. Each participant was instructed to monitor a computer screen presenting a red rectangular box at random places on the screen, and to press a button (here a button on a mouse of a computer) as quickly as possible after detecting a yellow millisecond counter appearing inside this box. Upon pressing the button, the RT was displayed for 1 second, giving feedback to the participant. The inter-stimulus interval varied randomly

from 2 to 10 seconds. The millisecond counter timed out after a 30-second period without any response. Each test lasted 10 minutes.

The data recorded for each stimulus consists in the time when the counter starts, t_s, and the time when the participant responds, t_r. The RT is defined as $RT = t_r - t_s$, where t_r corresponds to the first response, if any, after at least 100 ms from the start of the counter. If the participant responds with a delay of at least 500 ms from the start of the counter, we report the result as a "lapse". If the participant does not respond at all for the 30-seconds period, we report the result as an "omission". In our dataset, we didn't record any omission. For each 1-minute window, we determine the following statistics: the mean RT (including lapses) and the percentage of lapses. The percentage of lapses is defined here as the ratio between the number of lapses and the number of stimuli over the 1-min time window.

Results

We show the results of the analysis of the PVT data, in particular:

- the evolution of the RT and of the percentage of lapses for the three tests:
  - Test 1: PVT 1, corresponding to the not-sleep-deprived condition;
  - Test 2: PVT 2, corresponding to the moderately-sleep-deprived condition;
  - Test 3: PVT 3, corresponding to the sleep-deprived condition;
- the variations between individuals.

Figures 3.7 and 3.8 show the evolution of the RT and of the percentage of lapses, respectively, as a function of time for the three tests. Each curve in both figures represents the mean value of these statistics for all participants, for each test. We can observe that, on average, the RT and the percentage of lapses increase with sleep deprivation but only very slightly with time. Figures 3.9 and 3.10 also indicate an increase of the RT and of the percentage of lapses as a function of the three PVTs, respectively. Moreover, if we compare the mean of all RT for all participants, we can indeed observe that it increases from 0.407 ± 0.099 sec in not-sleep-deprived condition (Test 1) to 0.458 ± 0.197 sec in moderately-sleep-deprived condition (Test 2), to finally 0.553 ± 0.328 sec in sleep-deprived condition (Test 3). Similarly, the mean percentage of lapses increases from 9.48±16.02 % in not-sleep-deprived condition (Test 1) to 18.30±23.74 % in moderately-sleep-deprived condition (Test 2), to finally 29.20 ± 26.92 % in sleep-deprived condition (Test 3).

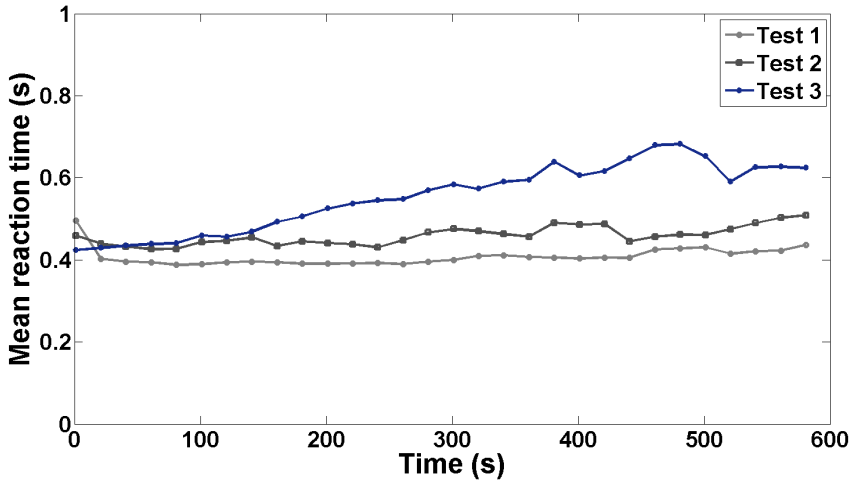

Figure 3.7: Evolution of the mean RT as a function of time for PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively).

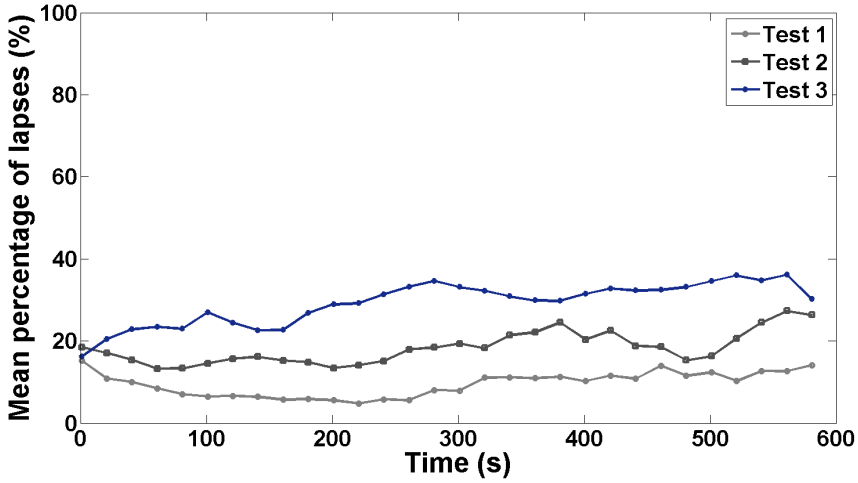

Figure 3.8: Evolution of the percentage of lapses as a function of time for PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively).

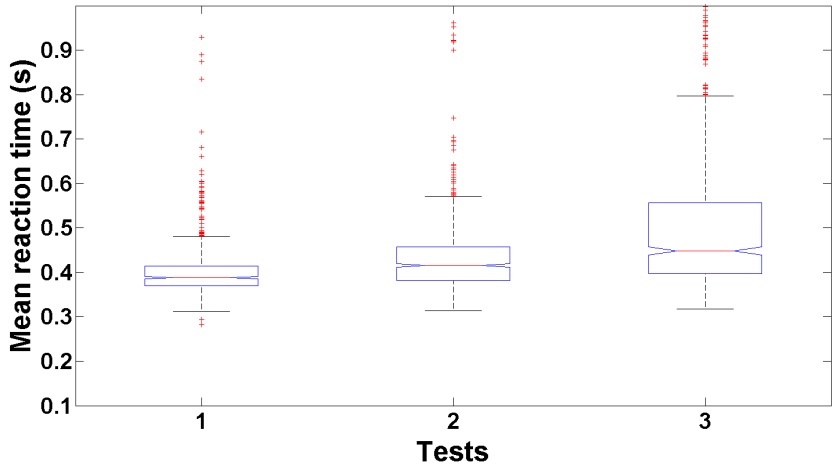

Figure 3.9: RT as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

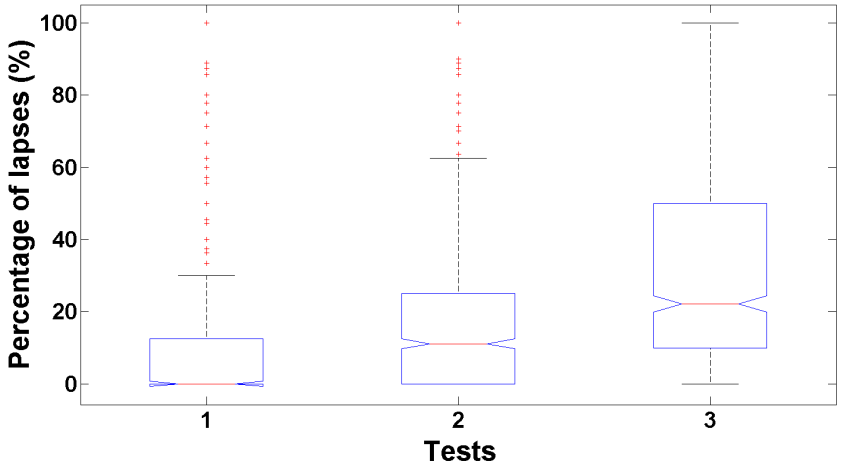

Figure 3.10: Percentage of lapses as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

We also performed Kruskal-Wallis analyses on RT and percentage of lapses to identify differences between participants. For these analyses, we used data from 26 participants because we faced some losses of data and we needed to analyze the three tests

for each participant. These analyses led to the following results. There are significant differences of RT between participants, regardless of their sleep condition ($\chi^2 = 1090.89; p < 0.01$), and between participants when they are not-sleep-deprived ($\chi^2 = 449.31; p < 0.01$), moderately-sleep-deprived ($\chi^2 = 573.34; p < 0.01$), and sleep-deprived ($\chi^2 = 573.69; p < 0.01$). The boxplot in Figure 3.11 indicates the differences of the RT between participants regardless of their sleep deprivation condition, i.e. for the three tests combined. Similarly, we can notice that there are significant differences of percentages of lapses between participants, regardless of their sleep condition ($\chi^2 = 918.61; p < 0.01$), and between participants when they are not-sleep-deprived ($\chi^2 = 317.38; p < 0.01$), moderately-sleep-deprived ($\chi^2 = 499.14; p < 0.01$), and sleep-deprived ($\chi^2 = 563.43; p < 0.01$). The boxplot in Figure 3.12 indicates the differences of the percentages of lapses between participants regardless of their sleep deprivation condition, i.e. for the three tests combined.

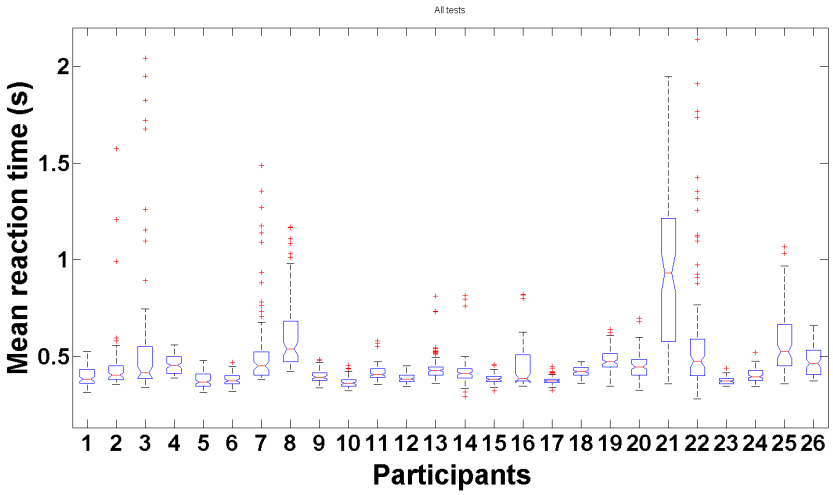

Figure 3.11: Differences of mean RT between participants regardless of their sleep deprivation condition. The red line inside a box represents the median value. The upper and lower borders of a box show the upper and lower quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

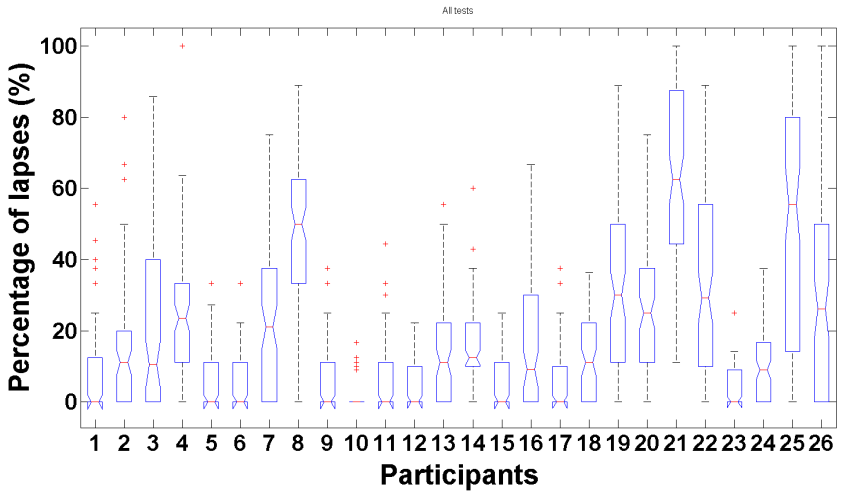

Figure 3.12: Differences of percentage of lapses between participants regardless of their sleep deprivation condition. The red line inside a box represents the median value. The upper and lower borders of a box show the upper and lower quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

In summary, this analysis of RTs and percentages of lapses provides evidence that the participants were indeed sleep deprived to a degree that affected their psychomotor performance.

B. Driving session

Method

We assessed the performance of each participant in the accomplishment of a task using a driving simulator. Driving in a simulator with a monotonous scenario and no environmental stimulus induces drowsiness levels that are even higher than for real driving [110]. According to several authors, it would seem that a driving time between 20 and 40 minutes is sufficient to cause drowsiness [111].

The simulator used in our study is equipped with the SIM^2 simulation software developed by IFSTTAR and with all accessories described in Section 3.2.3, to get as close as possible to actual driving conditions. The scenario that we chose simulates night time driving on a two-lane highway, free from traffic and other disruptive stimuli. The circuit implemented was in the form of a loop of about 20 km, in order for us to be able to impose any driving duration to participants. In this study, we chose to expose the participants to a 45-minute driving time.

Driving performance was assessed using the Standard Deviation of Lateral Position (SDLP). Figure 3.13 illustrates the SDLP. As we have seen in Chapter 2, this parameter

is widely used as an indicator of driving performance as it represents the deviation of trajectory of a vehicle. The SDLP is calculated using:

$$SDLP = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x})^2},$$

where i is the time/sample index, N is the number of samples, x_i is the lateral position of the vehicle on the road at the i^{th} time index, and $\bar{x}$ is the mean lateral position.

The higher the SDLP, the more the vehicle deviates from its trajectory and crosses lines. It is thus an indicator of poor driving performance. The SDLP is a very reliable measure; it varies between individuals but is significant within individuals across time [112].

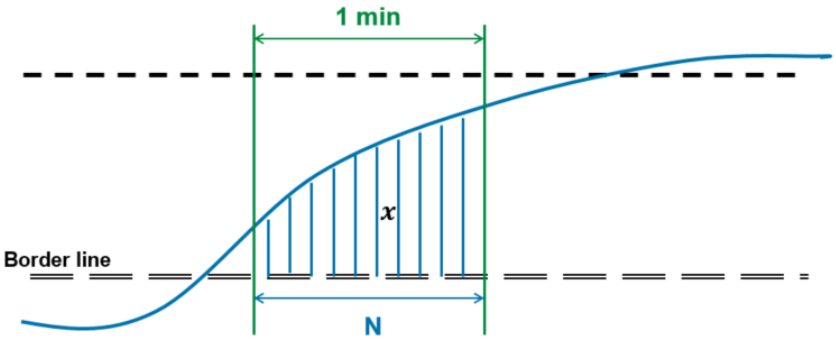

Figure 3.13: Illustration of the Standard Deviation of the Lateral Position (SDLP). The SDLP is a measure of the deviations of the vehicle on the road in a given time window (1 min here).

Results

We show the results of the analysis of the performance of individuals during the driving sessions in the simulator, in particular:

- the evolution of the SDLP for the three tests:
  - Test 1: Driving session 1, corresponding to the not-sleep-deprived condition;
  - Test 2: Driving session 2, corresponding to the moderately-sleep-deprived condition;
  - Test 3: Driving session 3, corresponding to the sleep-deprived condition;
- the variations between individuals.

Figure 3.14 shows the evolution of the SDLP as a function of time for the three tests. Each curve in the figure represents the mean value of the SDLP for all participants for each driving session and we can observe that, on average, the SDLP increases strongly with sleep deprivation and that there are also many more variations of the SDLP over time for Test 3. Figure 3.15 indicates also an increase of the SDLP as a function of the

three tests. Moreover, if we compare the mean of all SDLP values for all participants, we can observe that it increases from 397.49 ± 126.78 mm in not-sleep-deprived condition (test 1) to 628.14±558.75 mm in moderately-sleep-deprived condition (test 2), to finally 888.39 ± 937.10 mm in sleep-deprived condition (test 3).

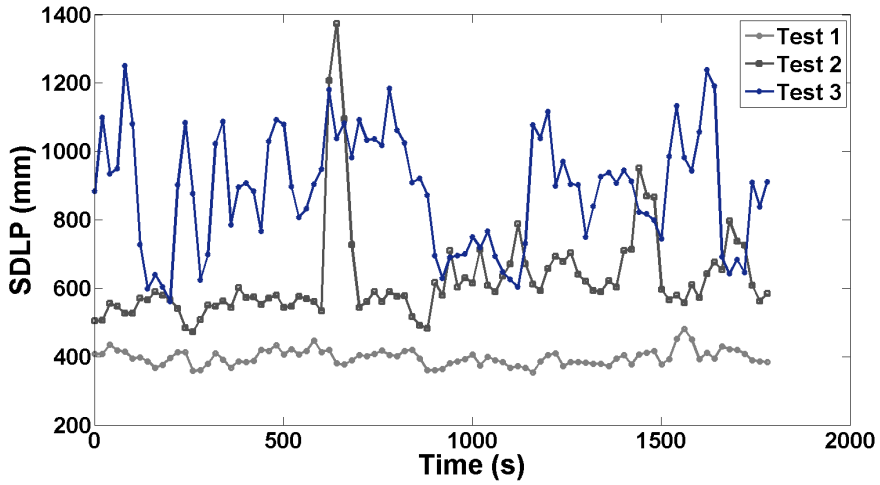

Figure 3.14: Evolution of the SDLP as a function of time for Driving Sessions 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively) and for all participants.

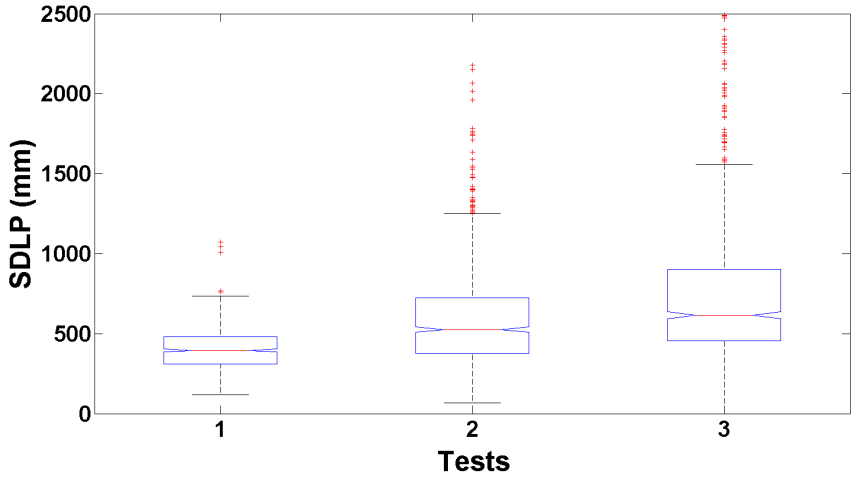

Figure 3.15: SDLP as a function of driving Sessions 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

We also performed a Kruskal-Wallis analysis on SDLP to identify differences between participants. This analysis led to the following results. There are significant differences of SDLP between participants, regardless of their sleep condition ($\chi^2 = 1121.84; p < 0.01$), and between participants when they are not-sleep-deprived ($\chi^2 = 497.89; p < 0.01$), moderately-sleep-deprived ($\chi^2 = 714.1; p < 0.01$), and sleep-deprived ($\chi^2 = 563.4; p < 0.01$). The boxplot in Figure 3.16 indicates the differences of the SDLP between participants regardless of their sleep deprivation condition, i.e. for the three tests combined.

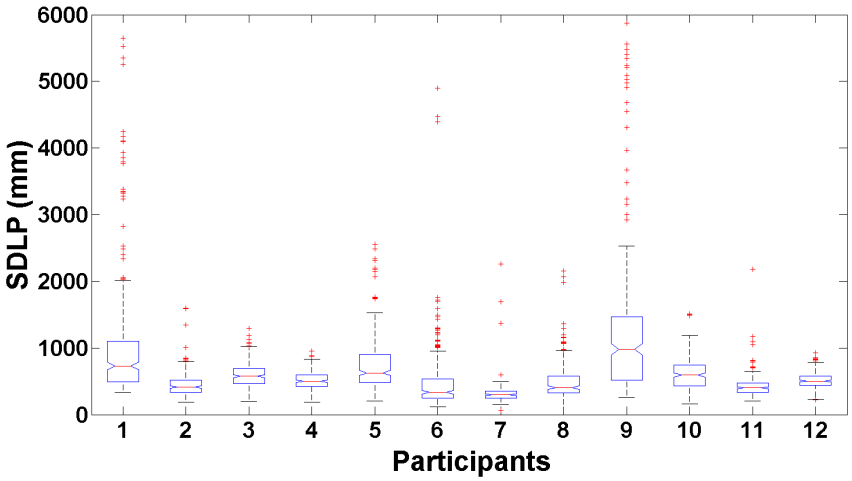

Figure 3.16: Differences of SDLP between participants regardless of their sleep deprivation condition. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

3.3.3 Analysis of subjective reference (KSS)

Method

Before each PVT, each participant was asked to rate his own LoD using the KSS. This scale is composed of nine states (ranging from 1 = extremely alert, to 9 = very sleepy, fighting sleep) and the participant must choose the state that best corresponds to his perceived state of alertness/drowsiness over the last few minutes. As described in Chapter 2, the KSS was validated against EEG and is widely used in the drowsiness-related literature. The KSS is indeed very easy to implement, and it gives a good indication of the state of drowsiness perceived by each participant. We thus obtained one value per test per participant.

Results

We show the results of the comparison of the KSS levels of all participants for the three tests.

Figure 3.17 indicates that the KSS levels for all participants increase as a function of sleep deprivation (across the three PVT tests). If we compare the mean value of KSS for all participants, we can notice that it increases from 2.79 ± 1.36 in not-sleep-deprived condition (Test 1) to 5 ± 1.35 in moderately-sleep-deprived condition (Test 2), to 6.48 ± 1.66 in sleep-deprived condition (Test 3).

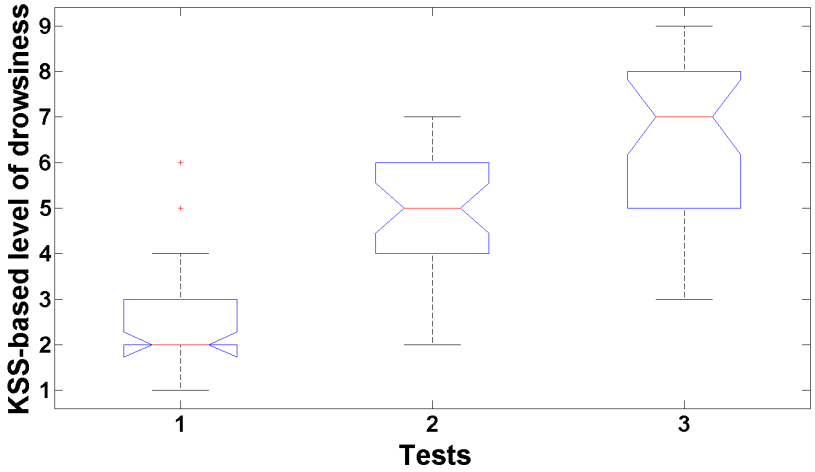

Figure 3.17: KSS as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

3.4 Summary of chapter

We established and implemented a complete test protocol to acquire data under different sleep deprivation conditions. The data collected are essential to develop and then test (validate) the two drowsiness characterization systems described in the following chapters. We also presented and analyzed the "reference" data (i.e. PSG signals, PVT parameters, driving session parameter, and KSS) that are used later for validation. The analysis of these data showed that each reference is significantly sensitive to sleep deprivation but that there are differences between participants.

Chapter 4

Development and test of an innovative POG-based drowsiness characterization system

This chapter describes the innovative POG-based drowsiness characterization system that we developed. This system is based on several key ocular parameters, the values of which are obtained from images of the eye. The use of images of the eye, i.e. phootooculography (POG), appears to be the best method for characterizing drowsiness in operational settings since it is completely physiology-based, task-independent, and non-invasive. Section 4.1 explains our motivation and the choices that we made. Section 4.2 describes the state of the art relating to techniques for the characterization of drowsiness via the analysis of the eyes. Section 4.3 gives an overview of the methods that we implemented in order to develop a complete drowsiness characterization system. Section 4.4 shows a comparison between the results of our system and different references and Section 4.5 discusses them. Section 4.6 describes a first adaptation of our system to process face images instead of eye images. Finally, Section 4.7 summarizes the chapter.

4.1 Introduction

The main goal of this chapter is to develop an innovative drowsiness characterization system mainly on the basis of the analysis of images of the eye and that can be used in operational settings. The activity of the eye is indeed an excellent indicator of an individual's state of drowsiness as it reflects what happens in the brain. Moreover, by just looking at the eyes of a driver or of a passenger, we can generally tell that the individual is getting more and more drowsy. Specifically, we would not consider looking at any other body part of the individual for this. It is thus completely natural to try to develop a drowsiness characterization system that is mostly based on images of the eye.

The use of eye images brings many essential advantages (as already detailed at the end of Chapter 2):

- to obtain an objective characterization (as opposed to a subjective characteriza-

tion) of the level of drowsiness (LoD) of an individual;

- to obtain a measure based on physiological parameters (as opposed to based on performance or on behavior parameters) that is independent of the task and thus usable in any situation/application;
- to determine an LoD in real time and automatically;
- to be non-invasive and not to ask for the participation of the user (i.e. non-disturbing);
- to use the system in operational settings.

We first focused our attention on a camera mounted on the head of an individual, as opposed to being mounted remotely, such as on the dashboard of a vehicle, because a head-mounted camera enables to be more accurate and reliable. As already mentionned in Chapter 2, we can indeed have a better resolution of the eye in the acquired images and a permanent tracking of the eye even if the individual moves. Much can be done with just one eye, so that we also concentrated our efforts on the analysis of a single eye throughout.

Even if we focused on a head-mounted camera, we propose, at the end of the chapter, an adaptation of our system to process face images with low resolution and low frame rate instead of eye images with high resolution and high frame rate. Wearing glasses can actually seem a bit binding and the automotive industry is more in favor of integrating a camera into the dashboard of vehicles. Therefore, we wanted to show that the algorithms that we developed for images acquired by a camera placed on the head of an individual can be adapted to the case of images acquired remotely.

Our general strategy is to follow the motion of the eyelids and of the eyeball (of one eye), to quantify a number of ocular parameters (OPs) that are known to be indicative of drowsiness, and to derive an LoD.

Our system comprises the following steps:

- to develop smart image processing algorithms to detect the position of the eyelids and the position of the eyeball for each eye image. This was done with the help of another researcher, Thomas Hoyoux [113].
- to identify the OPs that are indicative of drowsiness and to understand how these parameters evolve as the LoD increases. To this end, we made several searches of the literature on this specific subject and we established a complete list of existing OPs, and we also created new ones.
- to determine the most appropriate duration of a time window for the computation of OPs.

- to develop algorithms to compute the OPs from the positions of the eyelids and of the eyeball.
- to develop a method that processes all the values of the selected OPs to derive a single number that is the LoD of the individual being analyzed on a numerical scale from 0 (well awake) to 10 (very drowsy).

We cannot address here the detailed design of the methods used to extract the positions of the eyelids and of the eyeball and the methods to convert the OPs into an LoD (for contractual reasons of confidentiality). The thesis thus concentrates on the analysis of the performance of our complete system by comparing some of our computed OPs and our computed LoD with several references. The purpose of this comparison is to validate our system and to show that, while the methods are, by necessity, kept confidential, they are relevant and make sense for determining an individual's state of drowsiness. We particularly wanted to verify that the LoD determined by our POG-based drowsiness characterization system is well related to several references (already described in Chapter 3, Section 3): (1) the LoD obtained by analyzing PSG signals, (2) the performance in the accomplishment of PVTs, (3) the performance in the accomplishment of a driving session in a driving simulator, and (4) the self-estimated LoD using the KSS.

In order to determine the best threshold among our POG-based LoDs to alert individuals before they become dangerous, we also performed an analysis using confusion tables and ROC curves.

4.2 State of the art

This section presents the state of the art in the relevant literature of techniques that take OPs as input (extracted from signals or images) to determine the state of drowsiness of an individual. This review is not exhaustive but it enables to compare the performance of our system to what exists in this literature.

The literature related to drowsiness characterization techniques based on OPs began around 1998 with PERCLOS and its calculation via a semi-automatic eye tracking measurement system [77]. Since then, several research groups have worked on eye or face images to obtain OPs and to determine an LoD but, generally, they only detect the state of the eye, i.e. open or closed, to deduce the state of the individual. In 2002, Hayami et al compared the PERCLOS with vertical eye movement frequency to distinguish wakefulness from drowsiness but the results were not clear [114]. In 2008, Tabrizi et al developed a four-steps system that included face detection, eye detection, eye state detection using a chromatic-based algorithm, and drowsiness detection using PERCLOS. They showed that their method of eye state detection reached better results in comparison to traditional ones based on eyelids distance but they did not evaluate the drowsiness detection [115]. In 2009, Bhowmick et al also reached good classification of eye states using eye segmentation methods and the computation of shape features to train a Support Vector Machine (SVM) classificator but there was no assessment of the drowsiness

detection [116]. In 2010, Picot et al performed a study with face images analysis, in which they used data mining to select the most relevant blinking features, and fuzzy logic to detect drowsiness [117]. They achieved a little more than 80% of correct detection of the drowsy state with 13% of false alarm, but they only distinguished two states, i.e. alert or drowsy. Also in 2010, Liu et al proposed three criteria to quantify drowsiness from eye closures and eye blinks in images [118]. They reached 98% of accuracy but only on 10 individuals that were furthermore asked to imitate alertness and drowsiness blinking behaviors. In 2015, Rahman et al proposed an eye blink detection system to determine the open or closed state of the eye and activate an alarm if the closure lasts 2 seconds, or more, which may turn out to be too late [119].

Other research groups have instead decided to analyze the EOG to detect the behavior of blinks and thus determine an LoD [81]. In 2009, Shuyan and his colleagues decided to analyze the EOG with an SVM to discriminate 3 states of drowsiness. They reached accuracies of over 80%, with the best accuracy for the very sleepy state [120]. In 2014, Ebrahim et al tried to reproduce the same kind of result by extracting 18 blinking features from the EOG and by using different machine learning methods but they only achieved 66% accuracy for 3 output classes [121].

In 2007, Johns proposed a new 11-level drowsiness scale, the Johns Drowsiness Scale (JDS), based on the analysis of OPs extracted from an optooculography system (the Optalert system) [122]. This system is integrated in the form of a pair of eyeglasses that uses infrared (IR) reflectometry of the eye, where short pulses of IR light are sent to one or both eyes and the reflected intensities are measured. In 2007, Johns compared his scale to a specific driving parameter (i.e. two wheels and half the car out of the lane in a driving simulator) and he showed a sensitivity of 75% and a specificity of 70.6% . In 2013, Anderson et al. further compared the JDS with lapses of attention and KSS, and they obtained Areas Under the Curve (AUCs) of Receiver Operating Characteristics (ROC) curves of 0.74 for lapses and 0.80 for KSS. The AUC is indeed a marker of good performances of a model.

4.3 Material and methods

Our POG-based drowsiness characterization system is made of several processing blocks:

- for each image of the eye, the system extracts the positions of both eyelids and the position of the pupil via image processing algorithms;
- on a sliding time window of 1-minute in duration, the system computes several values of OPs indicative of drowsiness from the positions of the eyelids and of the pupil obtained at the preceding step;
- the system converts the values of OPs into a unique/single POG-based LoD.

Figure 4.1 shows the corresponding block diagram.

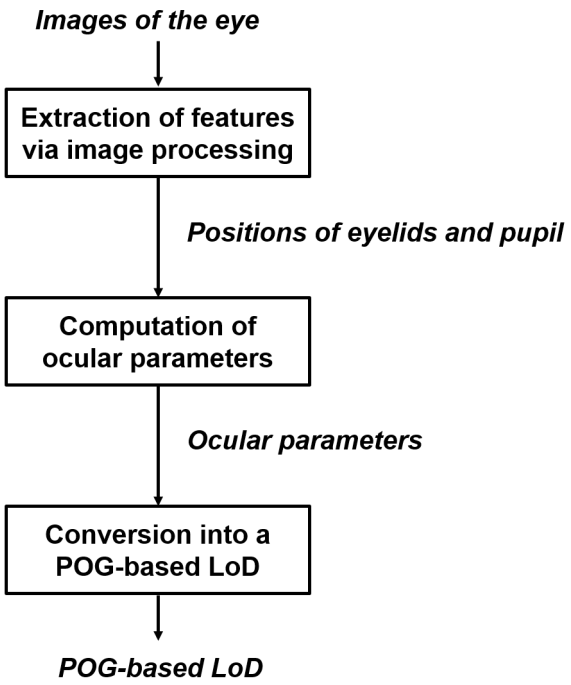

Figure 4.1: Block-diagram of our POG-based drowsiness characterization system.

4.3.1 Extraction of features via image processing and analysis

The images of the eye are acquired using a head-mounted POG-based drowsiness characterization system that we developed at University of Liège and that constitutes a prototype of the Drowsimeter R100 commercialized by Phasya company since 2015. This system is in the form of a pair of eyeglasses integrating

- an infrared (IR) illuminator (LED),
- a hot mirror that reflects IR light/radiation, and
- a small high speed camera sensitive in the IR.

The system provides IR images of the eye at a rate of 120 fps with a resolution of 320x240 pixels. One example of such images can be found in Figure 4.2.

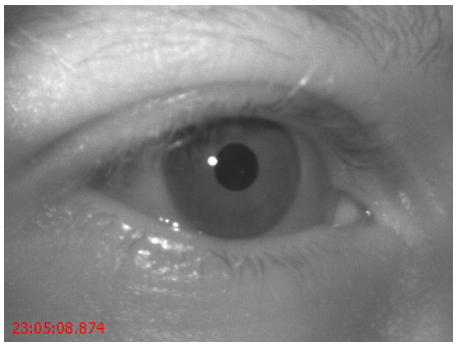

Figure 4.2: Example of eye image recorded by the prototype of the Drowsimeter R100.

Then our system extracts different characteristics/features from eye images to compute the values of OPs that will serve to determine the LoD of an individual. Therefore, we developed very specific and sophisticated methods of image processing to extract precisely the positions of both eyelids and the position of the pupil in each image. Figures 4.3 and 4.4 shows the result of the image processing algorithm.

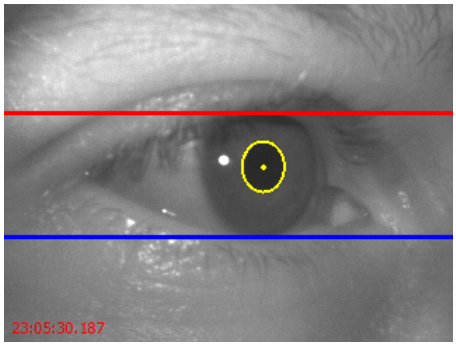

Figure 4.3: Example of the extraction of the positions of both eyelids and of the position of the pupil in an eye-opened image.

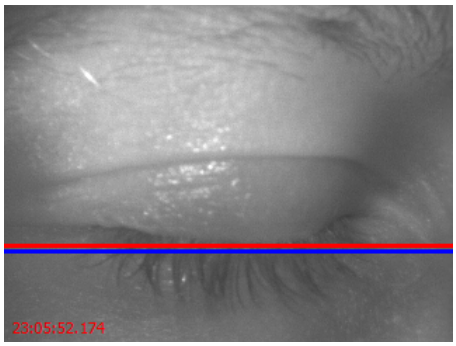

Figure 4.4: Example of the extraction of the positions of both eyelids in an eye-closed image.

In order to verify the quality of our image processing, our team manually annotated 86,630 images from 201 individuals and we compared this manual annotation with the results of the image processing algorithm. For the positions of the eyelids, we obtained an accuracy of 87%.

4.3.2 Computation of ocular parameters indicative of drowsiness

In this thesis, we mainly consider OPs that are related to the positions of the upper and lower eyelids and to the position of the pupil, obtained as explained in Section 4.3.1. Even though one could imagine defining some OPs for each individual image, here we only consider OPs that are each defined over a time window, with a duration denoted by W (in seconds). The window thus contains several consecutive POG images, from

which the eyelids and the pupil positions are extracted. For example, at 120 fps, we have about 7,200 images per 1-min window. In this case, each OP would capture some relevant characteristic of a succession of 7,200 upper eyelids positions, 7,200 lower eyelids positions, and 7,200 pupil positions. We emphasize that each OP produces one numerical value per window. The window can butt or overlap and, in this work, we use overlapping windows of 60 sec (1 min) with a step of 20 sec.

The positions of the upper and lower eyelids enable us to measure the distance between the eyelids, referred to here as the opening of the eye. From the opening of the eye, we can detect blinks, and thus compute a series of OPs related to the blinks and prolonged closures of the eye that are indicators of the state of alertness or drowsiness of an individual (as explained in Chapter 2).

The average distance between the eyelids varies greatly from one individual to another (depending mainly on the morphology). Therefore, in order to have a system that works on the majority of individuals, we developed an automatic eye opening baseline measurement for each individual. This baseline enables us to compute normalized OPs for each individual.

The position of the pupil in each image enables us to measure the displacements of the eyeball, which generally slow down when an individual becomes drowsy. From these displacements, we can compute a series of other OPs.

In addition, as mentioned above, all the OPs are computed for a time window. The duration of the window (W) is such that several "events" typically occur in a window (as long as the subject is not asleep), such as blinks. Most OPs are thus either a count of such events, the sum of their individual durations defined in some way and perhaps, expressed as a percentage, or some sample statistics, such as the sample mean of these individual durations. Some examples of OPs that we extract from images of the eye are:

- the mean duration of blinks;
- the PERCLOS 70 (which is, as a reminder, the proportion of time in a 60-s window that the eye is at least 70% closed);
- the percentage of microsleeps (with a microsleep being defined here as an eye closure of at least 0.5 s);
- the average displacement of the pupil.

We thus obtain the values of a set of several OPs for each 60-s window of interest.

4.3.3 Conversion into a POG-based level of drowssiness

From the set of OPs computed in Section 4.3.2., we derive a score using our proprietary conversion algorithm. The score represents the state of wakefulness/drowsiness

of the individual, and we call it a POG-based LoD. As a reminder, POG means photooculography, i.e. the use of images of the eye. This POG-based LoD, determined automatically by our system, is a numerical integer value between 0 (well awake) and 10 (very drowsy).

We developed our proprietary conversion algorithm using some data from Dataset A (other data, not described in this work, were also used). The Dataset A was presented in Chapter 3, Section 3.2.1. It gathers the data acquired during the PVTs. Objective measures were used as ground truth. We could have hypothesized that, during the first PVT (Test 1), the participants were well awake and, during the third PVT (Test 3) that they were very drowsy but we did not make such a hypothesis. Indeed, individuals do not all have the same sleep/wake cycle or sleep habits. Some individuals are more morning-type and others more evening-type. Some have good sleep quality or are more resistant to sleep deprivation than others. We therefore used objective data to have ground truth outputs to develop our proprietary algorithm that converts OPs into LoDs.

4.3.4 Methods for performance evaluation

We describe the methods that we used to evaluate the performances of our POG-based drowsiness characterization system (i.e. to validate it). Specifically, we want to show the relation between the POG-based LoD determined by our system and

- the effect of sleep deprivation;
- several references;
- the use of a single OP.

As a reminder, the references we use are (1) the LoD obtained by visually analyzing PSG signals, (2) the performance in the accomplishment of PVTs, (3) the performance in the accomplishment of a driving session in a driving simulator, (4) the self-estimated LoD using the KSS. These references were described in Chapter 3 Section 3.

We also determine the best threshold on our drowsiness scale to alert an individual that would represent a risk.

Effect of sleep deprivation on POG-based level of drowsiness

For testing the influence of sleep deprivation on our POG-based LoD, we consider the data made of POG-based LoDs from all PVTs and all participants (i.e. the data from the Dataset A described in Chapter 3). As we already did in Chapter 3 for the references, a first quantitative information is obtained by dividing the data into three groups according to the different sleep deprivation conditions (not-sleep-deprived (PVT 1), moderately-sleep-deprived (PVT 2), and sleep-deprived (PVT 3)), and by showing the evolution of the POG-based LoD as a function of time for the three PVT tests and for all participants. We also provide the sample mean and the sample standard deviation of POG-

based LoDs in each group.

We performed several analyses of variance to obtain better quantitative information. To do so, as the POG-based LoD is not normally distributed, we decided to use again the Kruskal-Wallis test in order to examine the influence of sleep deprivation on POG-based LoDs, and to determine whether there are differences between individuals. There are three kinds of analyses of variance which test the differences of POG-based LoDs between:

1. participants, regardless of their sleep deprivation condition;
2. participants, when they are not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived;
3. sleep deprivation conditions (for all participants).

We accept statistical significance for a p-value less than 0.01.

To visualize the results, we decided to use box plots to illustrate the dispersion of data within groups.

Relations between the POG-based levels of drowsiness and all references

We consider the data windows from all PVTs and all participants (data from Dataset A excluding the data used to develop our drowsiness characterization system), regardless of their sleep condition. In each window, we can compute the POG-based LoD and several references: the sample mean of RTs, the percentage of lapses, and the PSG-based LoD. This enables us to obtain a graph in the form of a cloud of points for each reference (i.e. gold standard) with respect to our POG-based LoD. One can also associate each value of a reference to a POG-based LoD and compute the sample mean and sample standard deviation of all reference values for each integer POG-based LoD. With this, we can obtain bar graphs representing the evolution of each reference with respect to our POG-based LoD. These results were presented in several papers and conferences (the list of publications is available at the end of the report). In this thesis, as we already did in previous sections, we use box plots to illustrate the relationship between our POG-based LoD and each reference.

We are also interested in the relation between our POG-based LoD and the driving performance. We thus also consider the windows of all driving tests for all participants (Dataset B), regardless of their sleep deprivation condition. The Dataset B was presented in Chapter 3, Section 3.2.2. It gathers the data acquired during the driving sessions. In each window, we compute our POG-based LoD and the SDLP, and we perform analyses similar to the ones we did for the other references (as explained above for Dataset A).

Regarding the relationship between our POG-based LoD and KSS (Dataset A), as opposed to the other references, we can only use one measure of POG-based LoD per test.

Decision	Actual situation	
	F	$\bar{F}$
PF (POG-based LoD ≥ T)	Correct decision: true positive (TP)	Incorrect decision: false positive (FP)
P$\bar{F}$ (POG-based LoD < T)	Incorrect decision: false negative (FN)	Correct decision: true negative (TN)

Figure 4.5: Illustration of the table of confusion.

Indeed, participants had to evaluate themselves using the KSS only once, just before starting each PVT. We therefore decided to compare this reference to the mean of the POG-based LoD calculated for the whole test.

POG-based level of drowsiness as predictor of failure

In order to determine a critical value of POG-based LoD for predicting imminent risk of failure (risk of doing lapses in PVTs), we use Receiver Operating Characteristic (ROC) curves.

We denote the failure event by F. A prediction of F is made by thresholding the POG-based LoD: for a given threshold T, the prediction of F is indicated by PF(T). The mathematical expressions are:

$$\begin{cases} PF(T) = 1 \text{ if POG-based LoD} \geq T \\ P\bar{F}(T) = 1 \text{ if POG-based LoD} < T \end{cases}$$

In the current context, the thresholds are the integers from 0 to 10 corresponding to the different values of POG-based LoD. For each threshold, we can draw a confusion table that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN), and we can then estimate the sensitivity (SN) and the specificity (SP). "Sensitivity" is also called "true positive rate" or "probability of correct detection". "Specificity" is also called "true negative rate". Table 4.5 gives an illustration of the table of confusion in the current context.

The estimates of SN and SP are given by:

$$\widehat{SN} = \frac{N_{TP}}{N_{TP}+N_{FN}}$$
$$\widehat{SP} = \frac{N_{TN}}{N_{TN}+N_{FP}},$$

where N_{TP}, N_{TN}, N_{FP}, and N_{FN} are the number of TPs, TNs, FPs, and FNs that are observed.

For each threshold T, we thus obtain a pair $(\widehat{SN}(T), \widehat{SP}(T))$. To illustrate all pairs (corresponding to all threshold values), we use a ROC curve. A ROC curve is indeed

a plot representing the sensitivity as a function of "1-specificity" ($1 - SP$) for different threshold values. "1-specificity" is also called "false positive rate" or "probability of false detection". For all tests, a point of the ROC curve thus represents the couple ($1 - \widehat{SP}(T)$, $\widehat{SN}(T)$) for the threshold T. To identify the best threshold to predict failure events, we need to take the best compromise between sensitivity and specificity for the context of interest.

The optimal point (i.e. the best threshold) on a ROC curve is located at the upper left corner because it corresponds to the maximum value of SN and the maximum value of SP. However, very often in practice, no point on the ROC curve is actually in the corner but some are close to it. In these cases, to find the best threshold (i.e. the best compromise between SN and SP), one uses the geometric distance between the point and the upper-left corner, and the best threshold is the one corresponding to the smallest distance.

Comparison with the use of single ocular parameters

Our POG-based LoD is established using a combination of OPs indicative of drowsiness. Based on the initial review of the literature, we have effectively hypothesized that the use of a combination of parameters is more robust and reliable than the use of a single parameter to determine the LoD of an individual. In order to verify this hypothesis, we now analyze the relationship between OPs (recognized in the literature as good indicators of drowsiness, namely the mean blink duration, the percentage of microsleeps, the PERCLOS, and the variation of the pupil diameter) extracted from images of the eye and:

- the presence of alpha rhythm in PSG signals;
- the presence of theta activity in PSG signals;
- the PSG-based LoD.

We then do the same exercise with the POG-based LoD instead of each OP taken separately and we compare all the results.

The comparison between, on one hand, the OPs and the POG-based LoD, and, on the other hand, the presence of alpha rhythm and of theta activity in the EEG signals is interesting because there is a progression of the LoD depending on whether we find alpha rhythm or theta activity in the EEG signals. Indeed, alpha rhythm refers to early drowsiness, a more relaxed state of the individual, whereas theta activity corresponds to a real decrease of consciousness. We therefore want to analyze the correlation between, on one hand, our OPs and our POG-based LoD, and, on the other hand, these EEG features in order to determine whether our OPs and our POG-based LoD are more indicative of mild or severe drowsiness.

4.4 Experimental results and evaluation of performance

We determine the relation between the POG-based LoD produced by our POG-based drowsiness characterization system and

- the effect of sleep deprivation;
- several references;
- the use of a single OP.

For each result, we used 1876 overlapping 1-minute windows of data from 24 participants for Dataset A and 4320 overlapping 1-minute windows of data from 12 participants for Dataset B. We faced losses of data mainly due to recording issues.

We also determine the best threshold on our drowsiness scale to alert an individual that would represent a risk.

4.4.1 Effect of sleep deprivation on the POG-based level of drowsiness

This section shows the effect of sleep deprivation on the POG-based LoD computed from the images of the eye acquired during the PVTs (Dataset A). We particularly analyze:

- the evolution of the POG-based LoD for the three tests (three sleep deprivation conditions);
- the variations between individuals.

Figure 4.6 shows the evolution of the POG-based LoD as a function of time for the three PVT tests. Each curve in the figure represents the mean value for all participants. For We can observe that, on average, the POG-based LoD increases with sleep deprivation. In the first PVT test, the LoD seems to remain quite constant for the duration of the test, on average, for all participants. During PVT 2, and especially during PVT 3, the LoD tends to increase slightly with time. Figure 4.7 also indicates an increase of the POG-based LoD as a function of sleep deprivation for all participants.

Moreover, if we compare the mean of all POG-based LoDs for all participants, we can indeed observe that it increases from 1.97 ± 1.08 in not-sleep-deprived condition (Test 1) to 2.75 ± 1.89 in moderately-sleep-deprived condition (Test 2) and, to 4.45 ± 2.60 in sleep-deprived condition (Test 3).

We decided to go further in this analysis by splitting the last sleep deprivation condition into “sleep-deprived and not lapsing” (representing 526 1-minute windows out of 646) and “sleep-deprived and lapsing” (representing 120 1-minute windows out of 646). "Lapsing" here refers to having lapses during at least 30 seconds out of the 1-minute window. The means and standard deviations of the POG-based LoDs for these

two other conditions are 3.70 ± 2.16 and 5.70 ± 2.74, respectively. This shows that there is a significant increase in the POG-based LoD with sleep deprivation, and even more during a decline of performance such as lapses.

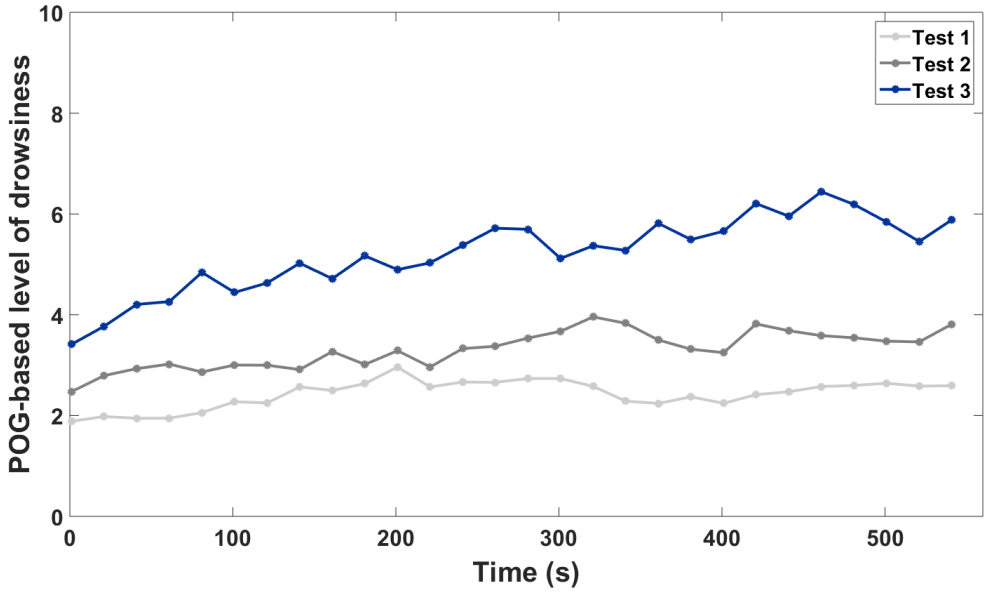

Figure 4.6: Evolution of the POG-based LoD as a function of time for PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively).

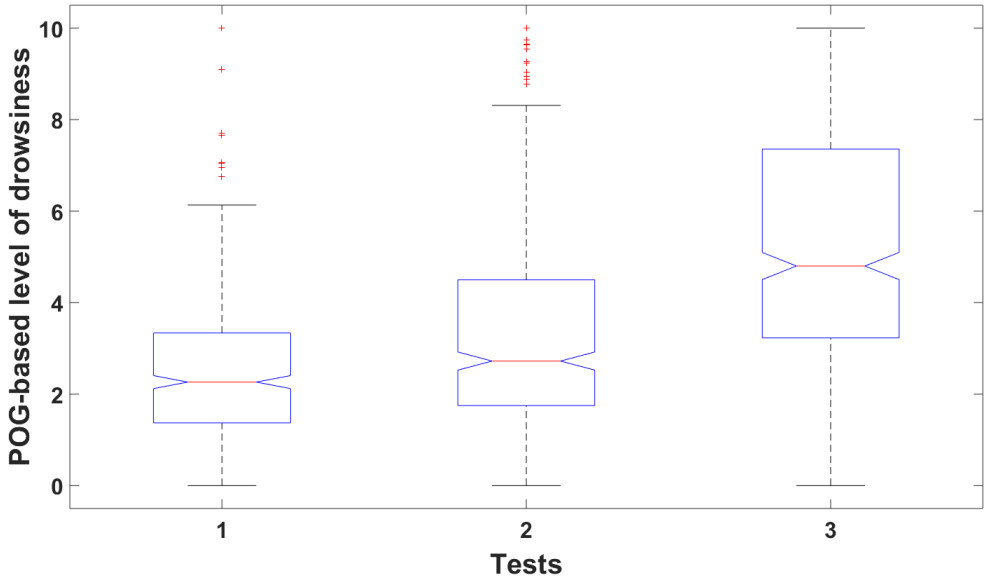

Figure 4.7: POG-based LoD as a function of PVT Tests 1, 2, and 3 (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived, respectively). The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

We also performed a Kruskal-Wallis analysis on POG-based LoD to distinguish differences between the participants. Even though the study involved 30 participants, these analyses were performed on data from 17 of them because we needed data from the three tests for each participant and we faced some losses of data mainly due to recording issues. The analyses led to the following results. There are significant differences of POG-based LoD between participants, regardless of their sleep deprivation condition ($\chi^2 = 568.29; p < 0.01$), and between participants when they are not-sleep-deprived ($\chi^2 = 306.82; p < 0.01$), moderately-sleep-deprived ($\chi^2 = 364.05; p < 0.01$), and sleep-deprived ($\chi^2 = 291.34; p < 0.01$). The boxplot in Figure 4.8 indicates the variability of the POG-based LoD between participants regardless of their sleep deprivation condition, i.e. for the three tests combined.

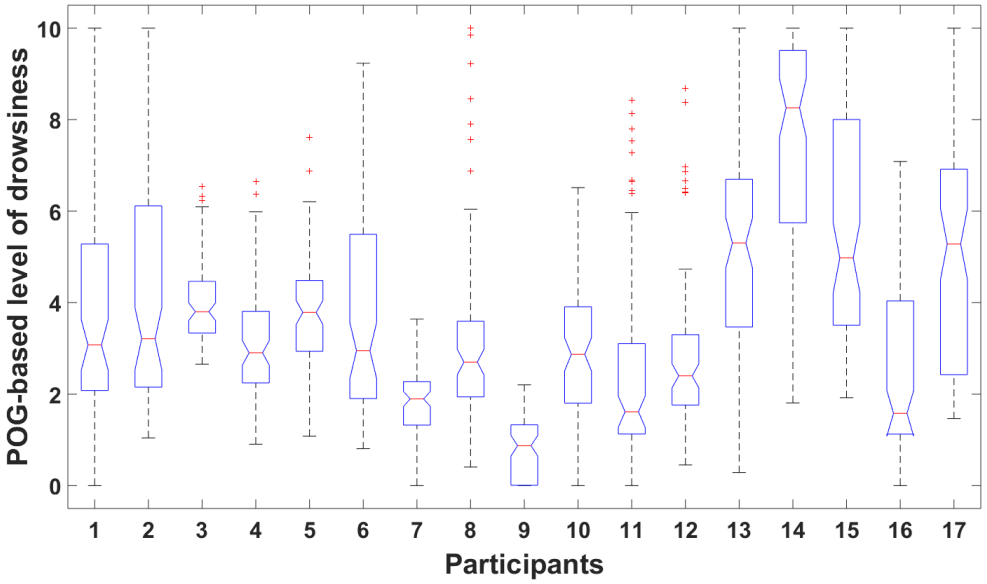

Figure 4.8: Variability of POG-based LoD between participants regardless of their sleep deprivation condition. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

4.4.2 Relations between the POG-based level of drowsiness and all references

For each 1-minute window of a PVT (Dataset A), we computed the POG-based LoD, the PSG-based LoD (reference), the mean RT (reference), and the percentage of lapses (reference). Each POG-based LoD value can thus be associated to a reference value and we can thus illustrate the distribution of the POG-based LoD in relation to each reference.

We analyze the relation between the POG-based LoD and the PSG-based LoD. Figure 4.9 shows the resulting boxplot. We can observe that when the reference PSG-based LoD increases, the POG-based LoD also increases. In order to go one step further in understanding the relationship between the POG-based LoD and the PSG-based LoD, we decided to make a graph showing the proportion of the PSG-based LoDs for each POG-based LoD value. Concretely, we divided the range of PSG-based drowsiness values into three zones representing three different levels of risk:

- "PSG-based LoD $\leq$ 3" constitutes the first zone, a zone of very low risk (equivalent to a waking state);
- "3 < PSG-based LoD $\leq$ 5" constitutes the second zone, a zone of medium risk (equivalent to a moderate-drowsiness state);

- "PSG-based LoD > 5" constitutes the third zone, a zone of high risk (equivalent to a high-drowsiness state).

We did not set the thresholds of 3 and 5 randomly. Indeed, the literature shows (1) that, a value at and above 30% on the KDS scale (the scale from which we obtained our PSG-based LoD), represents a greater risk of deviating when driving (hence a moderate-risk zone) [65], and (2) that a value above 50% on that scale is considered to be the initiation of sleep (hence a high-risk zone) [60, 123]. After establishing this division into three zones (refered here as three PSG-based zones), we associated each 1-minute POG-based LoD to a zone depending on the corresponding visual PSG-based LoD obtained for that 1-minute window. Figure 4.10 shows the resulting graph, i.e. the POG-based LoD as a function of the three PSG-based zones. We observe that, as our POG-based LoD increases, so does the percentage of high risk zone, and conversly.

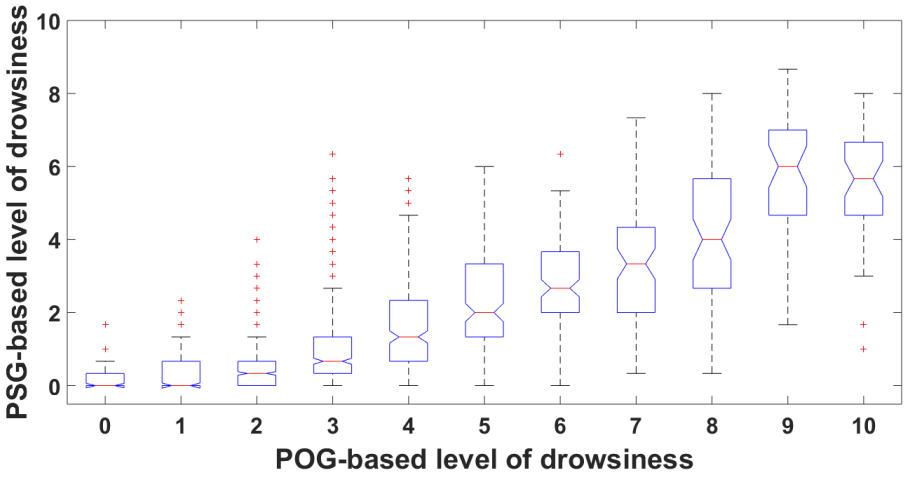

Figure 4.9: Relation between POG-based LoD and the corresponding PSG-based LoD. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

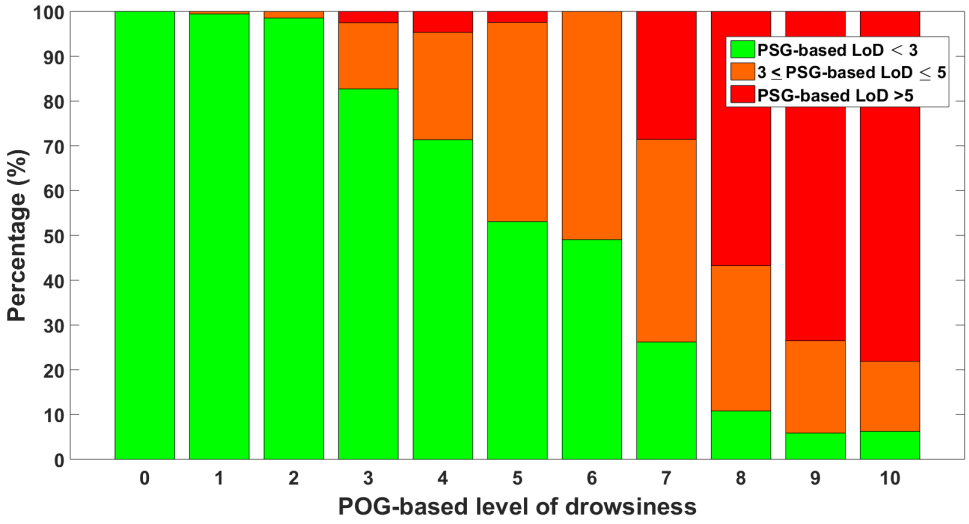

Figure 4.10: Relation between POG-based LoD and three PSG-based zones representing three different risk-levels (i.e. green for a low risk level, orange for a medium risk level, and red for a high risk level).

We analyze the relation between the POG-based LoD and the performance while executing the PVTs. Figure 4.11 shows the comparison with the RT, and Figure 4.12 shows the comparison with the percentage of lapses. We can again notice an increase in reaction time and in lapses with our POG-based LoD.

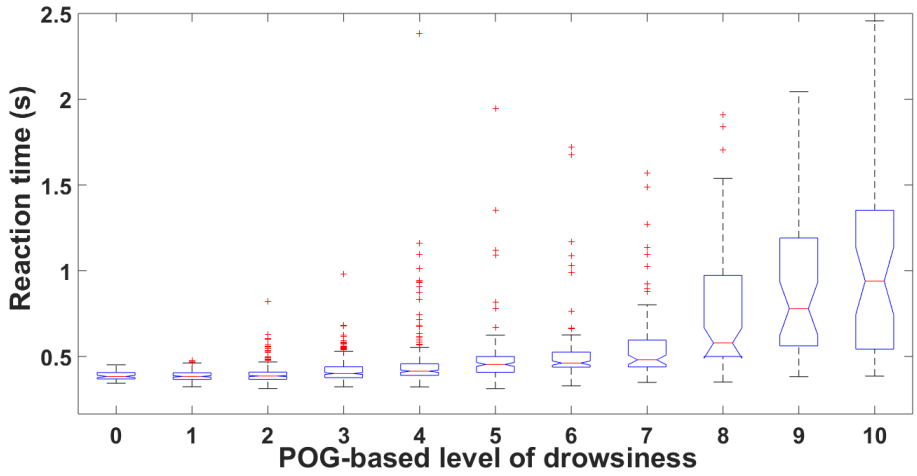

Figure 4.11: Relation between POG-based LoD and RTs. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

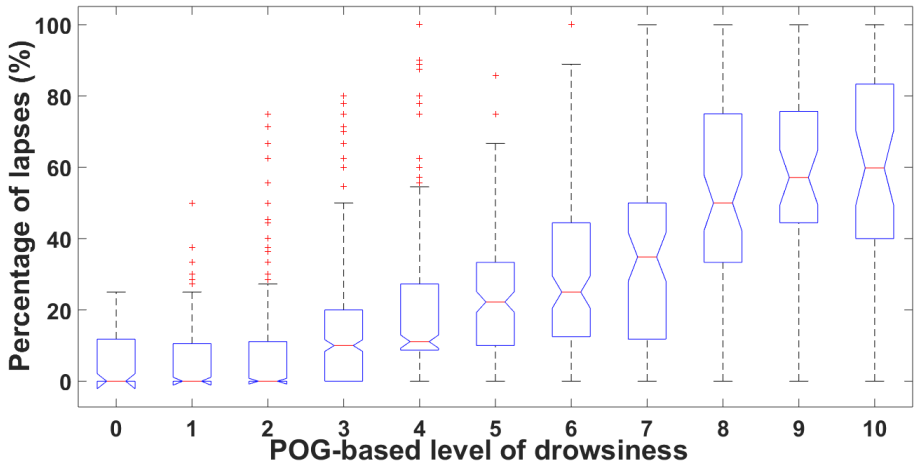

Figure 4.12: Relation between POG-based LoD and percentage of lapses. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

We also computed correlation coefficients between our POG-based LoD and the two references. The Pearson's correlation coefficient between POG-based LoDs and PSG-based LoDs is positive and high ($R = 0.78, p < 0.01$). The same result is obtained for the correlation coefficient between the POG-based LoD and the percentage of lapses ($R = 0.61, p < 0.01$). The Pearson's correlation coefficient between the POG-based LoD and the mean RT is positive and moderate ($R = 0.49, p < 0.01$).

We next consider the comparison of our POG-based LoD and the performance while driving in a simulator. For each 1-min window of the driving test (Dataset B), we computed the POG-based LoD and the SDLP. Each POG-based LoD value can thus be associated to a SDLP value. Figure 4.13 shows the resulting boxplot.

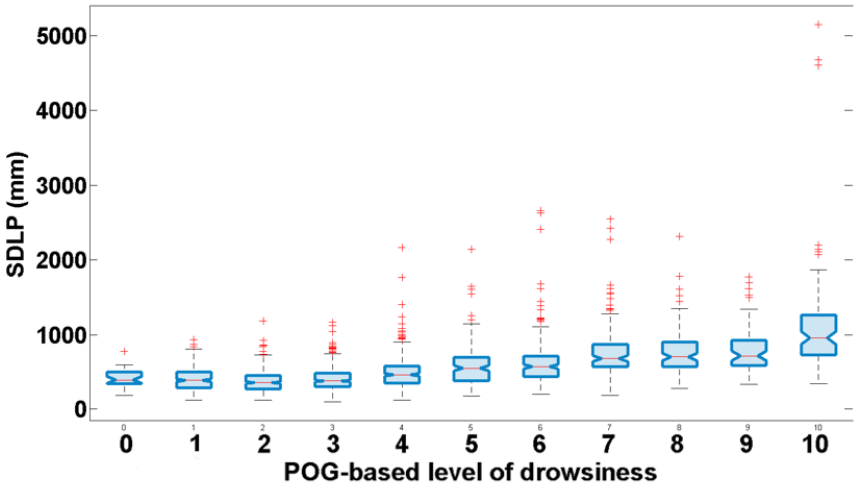

Figure 4.13: Relation between POG-based LoD and SDLP.

Finally, we demonstrate the relation between our POG-based LoD and the KSS. As already explained, we have only one value of KSS for each PVT. Therefore, in order to perform the comparison, we compute the mean POG-based LoD for each whole test. Figure 4.14 shows the result.

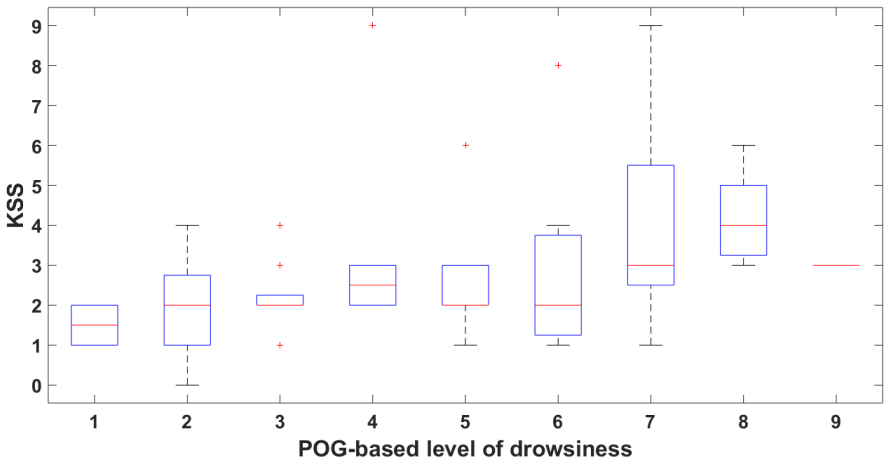

Figure 4.14: Relation between POG-based LoD and KSS.

4.4.3 POG-based level of drowsiness as predictor of lapses

In order to determine the best threshold among the POG-based LoDs to alert individuals before they become a risk, we use the concept of the ROC curve. Since we consider that lapsing during the PVT is synonymous with being a risk while performing critical tasks, like driving a vehicle, we find the best threshold among our POG-based LoDs that would

best predict lapses. To do this, we take integer thresholds from 0 to 10 (representing the different values of LoDs that we can obtain) and we compute values of sensitivity and specificity based on confusion tables for each threshold. Figure 4.15 shows the resulting ROC curve. In this figure, one should note that the threshold on the POG-based LoD increases from right to left. Indeed, a threshold of 10 implies that we would never predict lapses. The rate of true positives (TPs), or equivalently the sensitivity, would therefore be 0%. A threshold of 0 implies that we would always predict lapses. The rate of TPs, or equivalently the sensitivity, would then be 100%. The best threshold is the one with the highest probability of correct detection but with the lowest probability of false detection, i.e., the best compromise between sensitivity and specificity. A threshold of 5 on our POG-based scale of drowsiness seems to be the best to predict lapses. This threshold corresponds to a sensitivity of 72% and to a specificity of 80%. Figure 4.15 shows the ROC curve, and the best threshold is highlighted with a circle.

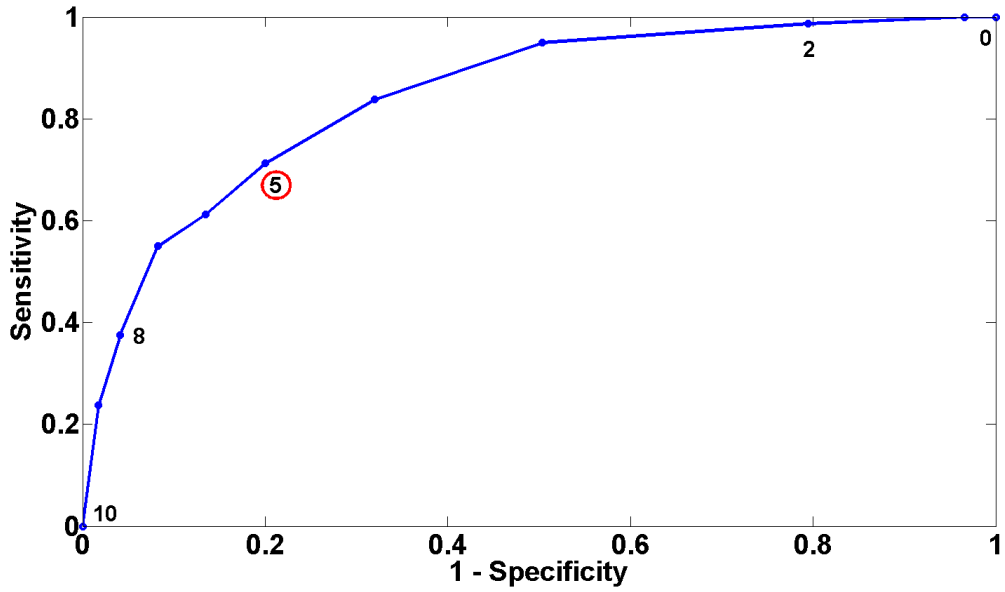

Figure 4.15: ROC curve describing the relation between POG-based LoD and lapses for the threshold values from 0 to 10.

4.4.4 Relations with OPs

For a given 1-minute window, we can compute a value for each OP, a POG-based LoD, the proportion of alpha rhythm, the proportion of theta activity, and the PSG-based LoD for this window. We can thus analyze the relations between the values of OPs and the POG-based LoD with these three references. We decided to examine four well-known OPs, i.e. the mean blink duration, the percentage of microsleeps, the PERCLOS 70, and the standard deviation of the pupil diameter. We thus computed correlation coefficients to analyze the relation between all data. Figure 4.16 gives the results.

Pearson's correlation coefficients	Alpha rhythm	Theta activity	PSG-based LoD
Mean blink duration	0.18	0.26	0.53
Percentage of microsleeps	0.19	0.32	0.52
PERCLOS 70	0.19	0.27	0.59
Standard deviation of the pupil diameter	0.27	0.12	0.25
POG-based LoD	0.36	0.33	0.78

Figure 4.16: Correlations between selected OPs and POG-based LoD and alpha rhythm, theta activity, and PSG-based LoD.

These results show that there is:

- a weak correlation between the OPs and the presence of alpha rhythm in the EEG signals,
- a better (except for the variation of the pupil diameter) but still weak correlation between the OPs and the theta activity in the EEG signals,
- a good correlation between OPs and the PSG-based LoD, except for the variation of the pupil diameter),
- a better correlation between the POG-based LoD and the presence of alpha rhythm and theta activity in EEG signals,
- a strong correlation between the POG-based LoD and the PSG-based LoD, as already covered in Section 4.4.2.

4.5 Discussion

We presented a new and innovative POG-based drowsiness characterization system that automatically and continuously determines an LoD. We showed in our experiments (Figure 4.7) that the POG-based LoD determined by our system increased significantly with sleep deprivation, on average, for all participants. This result is perfectly in line with previous studies that showed an increase in drowsiness with an increase in sleep deprivation [124]. However, the difference between these studies and ours is that our POG-based LoD is determined fully automatically via an objective drowsiness characterization system.

From Section 4.4.1, one can notice that the mean LoD in the third condition (sleep-deprived) remains quite low compared to the upper bound of our scale. Nevertheless, if one looks at the mean LoD in the "sleep-deprived and lapsing" condition, one can note that it is much higher than in the "sleep-deprived and not lapsing" condition. This means that our drowsiness characterization system reflects well performance decrements such

as lapses.

Regarding the time-on-task effect, our findings can be compared to results of earlier studies that reveal a stronger time-on-task effect under sleep deprivation condition than under normal condition for the PVT [125]. In Chapter 3, Figure 3.7 indeed shows a much higher increase of the RT with time for the third test and, in Figure 4.6, we can also observe an increase in the POG-based LoD with time for the third test, on average, for all participants.

Figure 4.8 shows great variability of POG-based LoD among individuals. Some individuals seem almost unaffected by sleep deprivation while others have a relatively high LoD for all three tests. This observation is not new. In 2005, Van Dongen et al, indeed confirmed that there is evidence that some aspects of sleep/wake-related variability (e.g. sleep duration, daytime sleepiness, vulnerability to the effects of sleep deprivation) involved traits specific to each individual [126]. In addition, we are dealing here with a young population, the rhythm of life of which is sometimes not very stable. Research shows that this segment of the population is one of the most at risk for road accidents due to drowsiness [14]. We also know that the chronotype, that gives rise to morning-type or evening-type individuals, generates differences in states of wakefulness and drowsiness depending on the time of the day from one individual to another [127]. Therefore, there is a real variability between individuals regarding sleep deprivation and drowsiness.

Looking again at Figure 4.7, one can notice that the variability between individuals is even stronger in the sleep-deprived condition than in the not-sleep-deprived condition. We can indeed observe that both the data sparseness and the standard deviation of the LoDs increase with sleep deprivation for all participants. In 2001, Doran et al. also reported this greater level of variability in drowsiness during sleep deprivation [125]. This may be explained by all the reasons given in the preceding paragraph.

In this thesis, we are not interested in analyzing intra-individual variability. Drowsiness can indeed be influenced by a lot of different factors, and controlling everything to put the individual in the same conditions to reproduce the results of a test is almost impossible. In general, it is true that an individual in good health and with a healthy lifestyle and a good sleep hygiene should have a similar behavior during two tests carried out at the same time on different days. However, if the individual has had fewer hours of sleep the night before, or a sleep of lesser quality, if he is more stressed or has a less active day, this is enough to vary the results and make them non-reproducible. The advantage of our system is precisely that it is not based on the habits of the individual, and that it does not take into account information about neither the profile of the individual using it, nor the hours of sleep, or the chronotype, etc. Our POG-based drowsiness characterization system is based solely on physiological parameters determined at the present moment and that reflect brain activity.

In Section 4.4.2, we compared our POG-based LoD with several references. We first discuss the comparison with the physiological reference, i.e. the PSG-based LoD. Figure

4.9 shows that, when the PSG-based LoD increases, the POG-based LoD also increases. Even more interesting and informative is the graph in Figure 4.10. It indeed associates each 1-minute POG-based LoD to a risk zone related to thresholds on the PSG-based LoD. We can thus notice that for low POG-based LoDs (between 0 and 4), we have a great majority of minutes with a PSG-based LoD below 3, which corresponds to very low risk and wakefulness. For POG-based LoDs between 5 and 7, the graph shows more variability between the zones but there is a higher percentage of minutes with a PSG-based LoD between 3 or 5, which can be associated with a medium risk zone. For POG-based LoDs above 8, the majority of minutes have a PSG-based LoD above 5, which is equivalent to being nearly asleep and thus being in a high risk zone.

The correlation coefficient obtained between POG-based LoDs and PSG-based LoDs further confirms this finding, i.e. there is a high correlation between POG and PSG-based LoDs. As polysomnography is the physiological reference, this means that our system correctly assesses the actual, physiological state of drowsiness of individuals.

Figures 4.11 and 4.12 show that the mean reaction time and the percentage of lapses also increase with our POG-based LoD. The computed correlation coefficients reveal that the POG-based LoD is more correlated with the percentage of lapses than with the mean RT. This indicates that our POG-based LoD is well correlated with performance decrements during the execution of a task. These results demonstrate the relevance of the LoD determined by our POG-based drowsiness characterization system.

Regarding driving performance, it is apparent from Figure 4.13, that there is also a good correlation between an increase in our POG-based LoD and an increase in lane deviations measured via the SDLP. This is very important because the purpose of developing a drowsiness characterization system like ours is obviously to avoid certain accidents related to drowsiness and especially in the automotive sector. On the commercial market, there are already some embedded systems that can detect deviations of trajectory but this is only the measure of the consequence of drowsiness or perhaps the consequence of another phenomenon such as distraction or difficulties related to climatic conditions. In addition, with the arrival of autonomous vehicles, these existing systems will become obsolete, and it will be necessary to focus on monitoring the driver to know if he is still fit to drive.

Going back to the results of our study, we know that the use of a driving simulator has, of course, a limitation: both the behavior of driver and his drowsiness pattern might be a lot different if the test was conducted on an actual road. However, a driving simulator offers the possibility of going further in terms of sleep deprivation and, therefore, in terms of drowsiness and lane deviation without any risk for the driver. Moreover, several studies that performed the comparison between driving in a simulator and driving in a real car on the road concluded that the relative validity of simulators is acceptable but simulators can cause higher drowsiness and effects of higher amplitude than real driving [128, 110].

Even if the KSS is a self-reported evaluation of sleepiness, its validity was confirmed by several studies and especially by Kaida et al in 2006 [129]. We thus wanted to verify that our POG-based LoD also made sense when compared to this scale. It must be said that the participants were involved in a study, and that they had thus no interest in underestimating or over-evaluating their LoD when using the KSS. They are not, e.g., on a road in the middle of the night wanting to go home at all costs, fighting against sleep. The results of the comparison show a positive correlation between KSS and our POG-based LoD. One should carefully note again here that we asked the participants to self-assess their LoD using KSS only once before the beginning of each test and that we compared this value with the average of the POG-based LoD calculated over the entire test. We should perhaps have asked for a self-evaluation at the end of the PVT, or taken only the first few minutes of the test for comparison.

In Section 4.4.3, we found that a threshold of 5 on our scale of drowsiness (from 0 to 10) would be the best to predict lapses. Of course, it should be clear that this "optimal" value of threshold is specific to the dataset used in this study. This threshold is the best compromise between sensitivity and specificity for this dataset. Depending on the application, one may want to be more "sensitive" than "specific", or conversely. In the case of drowsy driving for example, being less sensitive could lead to a severe accident. In the case of a dangerous task, it may be better to sound an alarm more often to maximize the detection of drowsiness-related decrements in performance, while accepting more false alarms. However, too many false alarms are not desirable either, as the operator may decide to ignore the alarms.

This threshold of 5 is very interesting because, if one goes back to the discussion concerning Figure 4.9, one can note that it is precisely the transition between the low-risk zone and the moderate-risk zone as defined by the thresholds on the PSG-based LoD. Indeed, 5 on our POG-based drowsiness scale corresponds to a much larger proportion of PSG-based LoDs between 3 and 5, and these PSG-based LoDs were associated, in Anund's thesis, with high risks of deviating from the trajectory in a driving simulator [123].

In Section 4.4.4, we compared the correlations between several OPs and our POG-based LoD with alpha rhythm, theta activity, and PSG-based LoD. From these results, we can draw several conclusions. Overall, it seems that our POG-based LoD (which, as a reminder, combines several OPs into a unique LoD) has a better correlation with all PSG-based references than any single OP. This means that a single OP is less accurate and reliable to analyze and determine changes in drowsiness than our POG-based LoD.

Looking in more detail at the results of the correlations with alpha rhythm and theta activity, one can notice that most OPs (except for the standard deviation of the pupil diameter) are better correlated with the presence of theta activity rather than with alpha rhythm. This means that each OP taken independently is indicative of a more advanced/severe drowsiness. The presence of theta in EEG signals refers indeed to a loss of consciousness. Therefore, if we decide to rely on only a single of these parameters

to give a feedback to the individual/operator on his state, this feedback may come too late. By contrast, our POG-based LoD shows a better correlation with alpha rhythm. Alpha rhythm as been shown (1) to detect early stages of drowsiness (notably in the driving context) and (2) to be associated with the deterioration of performance and RTs for visual stimuli (like the driving task) [61, 49, 14]. The conversion of several OPs into a single LoD would therefore help to better inform the individual on his state as this conversion enables to detect earlier the signs of drowsiness. These results are consistent with those of earlier studies [80, 82, 122]

From Figure 4.16, one can also confirm that we find the strongest correlations with the PSG-based LoD which also combines several parameters (i.e. alpha rhythm, theta activity, and slow eye movements).

Compared to some other existing drowsiness characterization systems that are based on the behavior of the individual or on the behavior of the process in which he is involved (e.g., driving a car) [130], and that can be influenced by several factors (e.g., climatic conditions, road conditions, and distraction), our POG-based drowsiness characterization system has the advantage of objectively characterizing the real physiological LoD of the individual and of being task independent. Our system is indeed based on a combination of several OPs to determine an LoD, and, specifically, it is not related to the task performed by the individual being monitored. It can therefore be used in many different applications. Moreover, compared to other technologies that use OPs [122], the use of images of the eye enables a more precise measurement of the LoD. Indeed, since the behavior of the eyes is a time-space phenomenon, the best way to characterize it, and to isolate specific eye movements, is to use images. Moreover, while POG-based technologies are non-invasive, the ones described in the literature remain fairly basic. The studies that we found are mainly based on the open or closed state of the eye to determine drowsiness, or on several OPs related to the movement of the eyelids but the resulting drowsiness state is not very precise and even, in some cases, just binary, e.g. alert vs drowsy. Our drowsiness characterization system computes a large set of different OPs and provides a precise and validated LoD.

The validation was performed in laboratory conditions. A further study should also evaluate this new, end-to-end POG-based drowsiness characterization system in real conditions, e.g., while driving a car on a road.

4.6 First adaptation to process face images instead of eye images

We describe a way to adapt (or "port") our validated POG-based characterization system detailed in the previous sections to use images of the face of the operator acquired remotely (e.g. from a dashboard) and at a low frame rate, rather than images of the eye acquired locally (from a head-mounted system in the form of a pair of eyeglasses) and at a high frame rate.

4.6.1 Method

We consider here that our head-mounted drowsiness characterization system (system A) is composed of two successive processing blocks. The first, called "eye image processing", takes as input eye images at a rate of 120 Hz, and outputs, for each image, the eyelids distance (i.e. the distance between the eyelids) with an amplitude ranging from 0 to 100 pixels. The second, called "LoD computation", first computes, for each (past) 1-minute window, several OPs, such as the mean blink duration and then determines an LoD.

Our remote system (system B) is composed of two successive processing blocks. The first, called "face image processing", takes as input face images at a rate of 30 Hz, and outputs, for each image, the eyelids distance with an amplitude ranging from 0 to 6 pixels. This system was developed by Quentin Massoz and will be presented in his thesis. In order to determine an LoD from the eyelids distance, the ideal would be, for the second block, to use the "LoD computation" block of system A. However, there are differences in the frame rate, amplitude resolution, and distribution of the values between the two systems.

We thus created a new system A' that takes as input eye images (still at 120 Hz) and reuses the "eye image processing" block of system A to get the eyelids distances (with amplitudes ranging from 0 to 100 pixels). Then, a new "adapter" block downsamples by four the acquisition rate of the images to reach a rate of 30 Hz, and non-linearly rescales the eyelids distance values via histogram matching. Then, the "LoD computation" block of system A is adapted to take as input the transformed eyelids distances. At the end, the adapted "LoD computation" block can be transferred/ported as such to system B.

Figure 4.17 illustrates this method.

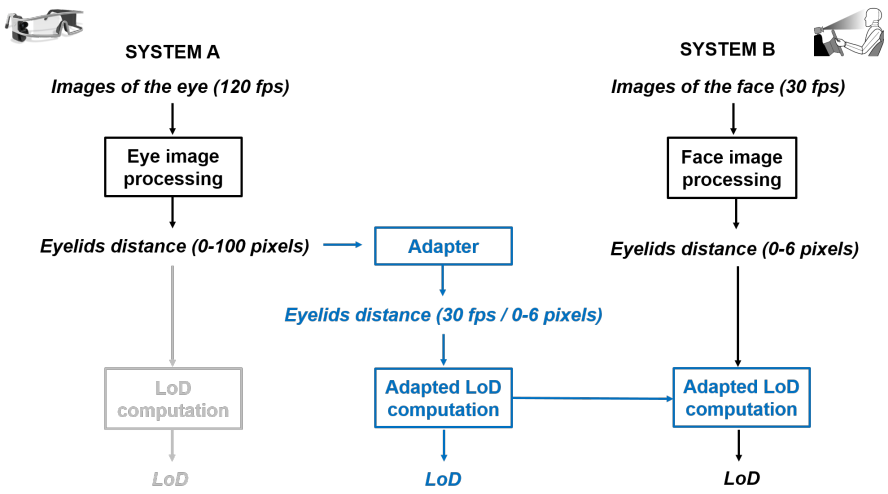

Figure 4.17: Block diagram of our method to adapt our head-mounted drowsiness characterization system to process face images instead of eye images.

Measures	LoD 1	LoD 2	LoD 3
Mean PERCLOS + SD	0.05 ± 0.03	0.12 ± 0.06	0.21 ± 0.11
Mean RT (s) + SD	0.386 ± 0.151	0.431 ± 0.184	0.518 ± 0.343
Mean percentage of lapses (%) + SD	9.30 ± 14.1	18.2 ± 21.1	31.9 ± 31.8

Figure 4.18: Mean values of three measures (PERCLOS, RT, percentage of lapses) as a function of the three LoDs.

4.6.2 Results

We conducted an experiment using the same protocol as before but in this case, the participants were seated in front of a remote system, consisting of a Microsoft Kinect v2 sensor that recorded face images, while performing three PVTs.

For all test performed with the head-mounted system, we applied system A' to all eye images and adapted the "LoD computation" block.

Subsequently, for each test performed with our remote system (system B), we analyzed the images and extracted the eyelids distance in each image. Then, for each 1-minute window, we applied the adapted "LoD computation" block to produce an LoD on a three levels scale (1 = awake, 2 = slightly drowsy, 3 = very drowsy). For each same 1-minute window, we also computed the PERCLOS, and we analyzed the PVT data to determine the mean reaction time (RT) and the percentage of lapses. This enabled us to associate each of these three measures to an LoD and to compute the mean PERCLOS, mean RT, and mean percentage of lapses for all participants for the three LoDs. Results of this analysis are presented in Figure 4.18.

To further verify the significance of the LoD determined by system B, we also analyzed the distribution of each measure (PERCLOS, RT, and percentage of lapses) as a function of LoD, as illustrated in Figures 4.19, 4.20, and 4.21 respectively.

Figure 4.19: PERCLOS as a function of LoD determined by system B. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

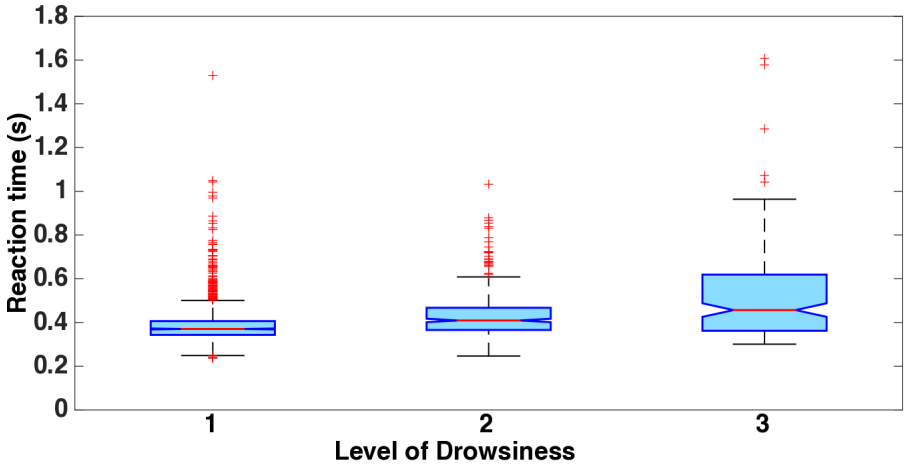

Figure 4.20: Reaction time as a function of LoD determined by system B. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

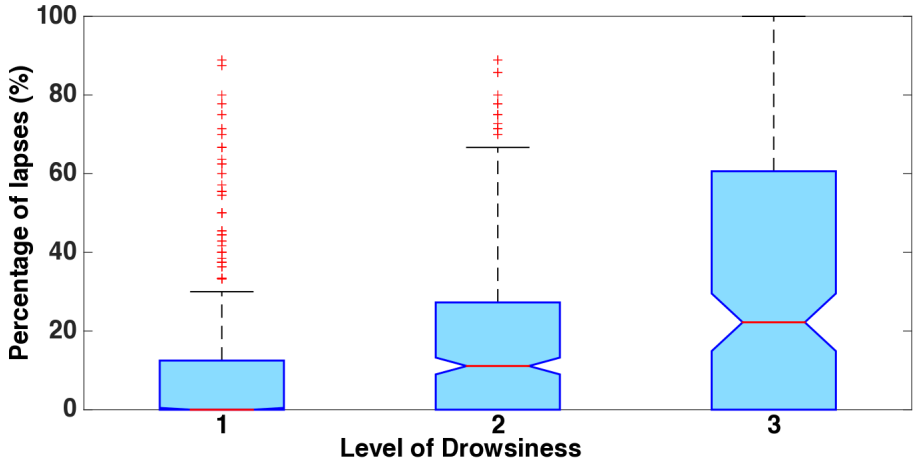

Figure 4.21: Percentage of lapses as a function of LoD determined by system B. The red line inside a box represents the median value. The lower and upper borders of a box show the lower and upper quartiles, respectively. The extended vertical lines indicate the variability outside the quartiles, and the red crosses represent outliers.

4.6.3 Discussion

The three figures indicate that, for all participants, the median and the variance of the PERCLOS, RT, and percentage of lapses increase with the LoD determined by system B.

Moreover, Table 4.18 shows a significant increase for all the mean values with respect to the LoD. This means that the LoD produced by system B is consistent and reflects well performance decrements linked to drowsiness.

To determine an LoD from system B, one could also imagine developing a new "LoD computation" block directly from face images. However, to do so, one would need a lot of time and the use of several references, which we do not have at the moment, to build the new block. In addition, the "LoD computation" block of system A has been validated with both physiological and performance references. Thus, as a first step, it was reasonable and meaningful to port this block from system A to system B.

4.6.4 Summary

This study demonstrated the validity of using the "LoD computation" part of our drowsiness characterization system developed for images of the eye acquired by a head-mounted camera to analyze images of the face acquired by a remote camera. Both modalities do have their specific advantages and disadvantages, and one will be preferred over the other depending on the situation or the environment in which one wants to monitor an individual's LoD. Nevertheless, we showed that our drowsiness characterization system can be adapted to be used in both cases. This research opens up the possibility of

monitoring a driver's LoD from dashboard mounted camera(s).

4.7 Summary of chapter

The POG-based drowsiness characterization system that we developed is a promising tool to characterize drowsiness, monitor its level, and determine the times when it reaches a dangerous level. We have indeed demonstrated that the POG-based LoD determined by our system is well correlated with the KDS-based LoD and with impairments of performance while performing a PVT. We have also determined that the best drowsiness level threshold to predict lapses is 5 on the scale from 0 to 10. Our system also has the advantage of being non-invasive, usable in any condition, and of requiring no intervention from the individual/user. It thus has significant potential for reliably quantifying the LoD of individuals accomplishing a task, and, ultimately, for preventing drowsiness-related accidents.

Chapter 5

Development and test of an innovative PSG-based drowsiness characterization system

This chapter describes the innovative, automatic PSG-based drowsiness characterization system that we developed. The approach used for the development of this system consists in extracting automatically from the EEG and EOG signals features indicative of drowsiness (alpha rhythm, theta activity, and slow eye movements) and producing automatically from them an LoD inspired from the Karolinska Drowsiness Scale (KDS). Section 5.1 gives an introduction on the choices we made. Section 5.2 describes the methods that we implemented in order to develop the PSG-based drowsiness characterization system. Section 5.3 shows the results produced automatically by our system and compares them to the results produced visually by experts. Section 5.4 discusses the results.

5.1 Introduction

In previous chapters, we have seen that, in order to improve health and safety, there is an obvious need for drowsiness characterization systems that are suitable for use in an operational environment. To develop such systems, there are several methods, e.g. via ocular parameters (OPs), but for most of these methods, the question of their validation unavoidably arises. We indeed need to know whether these systems are reliable and robust enough to determine the state of drowsiness of an individual and to be used in real life. In order to validate these systems, i.e. to evaluate their performance, one needs a reference system that would be intended for use as a gold standard, i.e. a point of comparison, for the off-line validation of other drowsiness characterization systems. However, such a reference system does not exist yet. The aim of this chapter is therefore to develop a reference drowsiness characterization system that is automatic and usable in any context for the validation of other systems.
In order to develop this reference drowsiness characterization system, we must first determine which reference method to use. In previous chapters, we already selected

several reference methods to validate our POG-based drowsiness characterization system:

- the visual analysis of PSG signals,
- the analysis of performance while performing PVTs,
- the analysis of performance while performing Driving Sessions in a driving simulator,
- the KSS.

The visual analysis of PSG signals is objective, independent of the task performed by the individual, and can thus be used in any context. Moreover, there is a universal agreement that PSG is the gold standard for quantifying sleep, and many experts tend to agree that PSG should also be the yardstick for quantifying drowsiness.

The analysis of performance depends on the task performed by the operator and cannot therefore be used as a single reference for validating every drowsiness characterization systems.

The KSS is very easy to use but is subjective.

PSG thus seems to be the most fitting approach. However, the analysis of EEG and EOG signals is generally performed manually, i.e. by visual inspection. This operation is very time consuming and the results may differ from one expert to the other. The general agreement between experts (sleep specialists) is indeed around 85%.

We thus decided to develop a novel, end-to-end, automatic PSG-based drowsiness characterization system that takes raw EEG and EOG signals and produces a level of drowsiness (LoD). Because of the emphasis on validation, our PSG-based system does not need to operate in real-time and, it can take advantage of sophisticated signal processing techniques.

Our system consists of different major processing steps.

- We perform a preprocessing step that consists in band-pass filtering the signals to get rid of the frequencies that are not useful in determining an LoD.

- We extract a number of drowsiness-indicative parameters/features from the EEG and EOG signals. The parameters/features of interest are:
  - the presence of alpha rhythm in the EEG;
  - the presence of theta activity in the EEG;
  - the presence of slow eye movements in the EOG.

There are several methods in the literature for this purpose, and we implemented some of them but these did not provide sufficiently reliable results. This probably explains why these methods do not seem to be used in practice. We thus considered a new, innovative method based on the Hilbert Vibration Decomposition. This method was initially created to study vibrations in mechanical systems but, since it is well adapted to analyze nonlinear and non-stationary signals, as is the case for EEG and EOG signals, we decided to implement it and to use it to extract the parameters of interest from these signals. To our knowledge, this method has never been applied before to EEG and EOG signals to detect drowsiness.

- We convert the features extracted from the EEG and EOG signals into an LoD using machine learning (and, specifically, a classification-tree technique) and the Karolinska Drowsiness Scale (KDS). As a reminder, the KDS scoring method consists in dividing the data in successive windows of 20 seconds and in visually determining a KDS score. Each KDS score is based, as explained in previous chapters, on the subjective assessment of the presence or absence of alpha rhythm and/or theta activity in the EEG, and/or SEMs in the EOG.

5.2 State of the art

The standard method to characterize drowsiness using EEG and EOG signals is to visually (i.e. manually) extract signs of drowsiness from these signals. This method is very time consuming, and the results of the characterization of drowsiness may also vary from one expert to another [131].

Therefore, researchers have developped several automatic methods for facilitating the analysis of EEG and EOG signals. However, before proceeding to the analysis of existing techniques, we recall in a few words the signs of drowsiness that we have to extract from the EEG and EOG signals in order to determine the state of an individual. These signs have indeed already been described in Chapter 2, but since we use them extensively in this chapter, a brief review seems fitting.

Many studies agree that drowsiness can be described by both brain activity and ocular movements [132, 60, 71]. Regarding brain activity, various power measures in various frequency bands as well as combinations of these measures are indeed affected by drowsiness. Among the frequency bands, researchers appear to agree that the most prominent ones to detect drowsiness are the alpha and theta bands [48, 60, 14]. Otmani et al have indeed shown that sleep deprivation and driving time have an effect on alpha and theta activities [14]. Some studies were also performed in driving simulators, and their results are similar: alpha and theta activities, slow eye movements, or SDLP increase before driving off the road [65, 64, 61]. Sandberg and his colleagues also investigated the effect of drowsiness during real driving and they confirmed that alpha and theta activities, and blink duration increase with drowsiness [133].

It is not surprising to see that theta activity is recognized as being a sign of drowsiness because, when we look at the sleep stages defined by Rechtschaffen and Kales (and subsequently redefined by the AASM in 2007), Stage 1 sleep (also known as the drowsiness stage) is characterized by a strong presence of theta activity (i.e. an increase of the spectral power in the theta frequency band) [56].

Regarding the alpha activity in the sleep stages, it is rather considered as characteristic of the wakefulness stage in relaxed condition with closed eyes. However, as already discussed in Chapter 3, the definition of Stage 1 sleep is for an individual lying in a bed with eyes closed and not an individual doing any other particular task with eyes open. It was thus shown that, for people with eyes open, an increase of the spectral power in the alpha frequency band is also observed when drowsiness occurs [58, 59]. Moreover, in 1999, Klimesch demonstrated that EEG alpha and theta oscillations reflect cognitive and memory performance [134]. Sleep experts still seem to agree that alpha activity is more a precursory sign of drowsiness, and that theta activity is more a sign of advanced drowsiness. There would thus be a progression in the state of drowsiness, starting with the arrival of alpha and continuing with the arrival of theta. It is thus interesting to extract both parameters (i.e. alpha activity and theta activity) from the EEG signals in order to determine whether an individual is still able to perform his task. The ability of an individual to be vigilant, perform well a task, and make good decisions is indeed influenced by his state of alertness/drowsiness.

From the above discussion, it seems clear that drowsiness is associated with increased alpha and theta activities in the EEG as well as with slow eye movements in the EOG. In order to convert these features into an LoD, several research groups proposed different methods. Among these, we selected the one developed by Gillberg and Åkerstedt in 1996 [60], i.e. the KDS. As a reminder, this method consists in dividing the 20 sec-window of analysis into 10 sub-windows of 2 seconds each. In each sub-window, if one detects the presence of at least one sign of drowsiness (alpha activity and/or theta activity and/or slow eye movements), the score for this sub-window is 10. The total score, i.e. the LoD, for the window of 20 seconds is therefore a value between 0 and 100. In this thesis, as we already did in the previous chapter, we increment the score by 1 and not by 10 if a sign of drowsiness is detected in a 2-second sub-window. The final score for a window of 20 seconds is then between 0 and 10.

In the literature, one finds different EEG signal processing techniques to determine the power in the alpha and theta frequency bands from EEG signals. Most of these use power spectrum analysis tools, like the Fast Fourier Transform (FFT), combined with principal component analysis (PCA), and/or machine learning techniques. In 1997, Jung et al. already proposed a method involving power spectrum estimation, PCA, and Artificial Neural Network (ANN) to detect alertness [135]. They showed that it was possible to accurately estimate an operator's LoD from two channels of the EEG. In 2003, Lal et al. used FFT to detect different phases of drowsiness but further research was needed to evaluate the software they had developed [136]. In 2005, Liang et al. analyzed the EEG power spectrum combined with a correlation analysis, PCA, and linear regression

models to assess the LoD of a driver in a virtual-reality-based driving simulator [137]. They showed that it was feasible to estimate driving errors based on the spectral analysis of EEG signals. More recently, Jap et al. came up with the analysis of four EEG frequency activities ratios for detecting drowsiness in operators [138]. They found that one ratio used in combination with a measure of decreased beta activity and increased theta activity is a good indicator of drowsiness. In 2010, Correa also introduced wavelet decomposition combined with spectral analysis of EEG signals to detect 2 stages (alert versus drowsy) [139]. In 2014, the same group also added the use of neural networks to discriminate between the alert and drowsy states, and they reached a correct detection rate of 87.4% for "alert" and of 83.6% for "drowsy" [140]. In 2013, Yu et al. also reached a very good accuracy for determining awake or drowsy states using FFT and a Support Vector Machine (SVM) [141]. They indeed achieved an accuracy and a precision of 98.01% and 97.91%, respectively, but using recordings of normal nights of sleep and not recordings of people performing a particular task. Among other techniques, we can also cite entropy measurements [62, 142] or fuzzy logic [143].

Some of these methods/systems provide interesting results, but we judge the results insufficient for these techniques to become a standardized reference for the validation of other drowsiness characterization systems. We thus felt that we had no choice but to develop a new, innovative system for characterizing drowsiness. We call it an automatic, PSG-based drowsiness characterization system. We add the qualifier automatic to distinguish this automatic system from the visual/manual approach.

5.3 Material and methods

Our automatic PSG-based drowsiness characterization system comprises several processing blocks:

- a preprocessing on the EEG and EOG signals, which consists mainly in bandpass filtering the signals between 1 and 25 Hz;
- the extraction of important features/attributes for each sign of drowsiness (alpha rhythm, theta activity, and slow eye movements) using the Hilbert Vibration Decomposition technique. This is done on a sliding time window;
- the conversion of the features extracted at the previous step into an automatic PSG-based LoD using machine learning and the KDS.

Figure 5.1 shows the high level block diagram of our automatic PSG-based drowsiness characterization system. We add the qualifier "automatic" to the PSG-based LoD for the same reason as we added it to "PSG-based drowsiness characterization system". We drop the qualifier when the context indicates clearly that the system is automatic, as opposed to visual/manual.

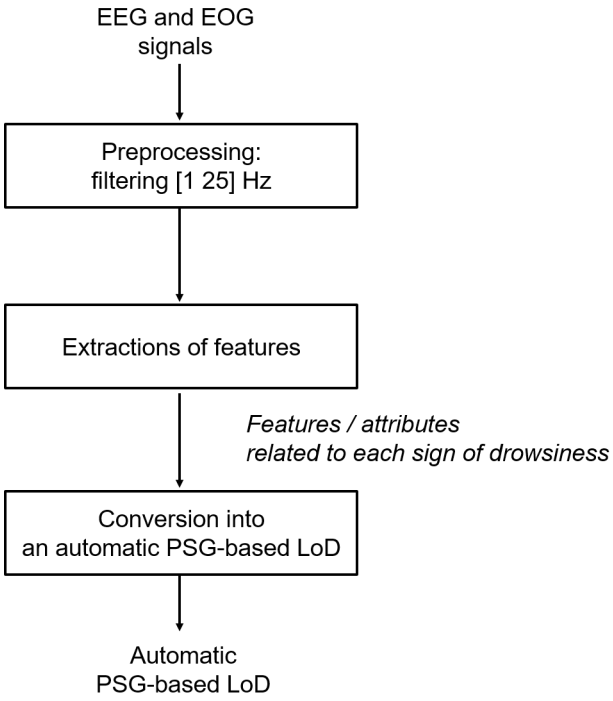

Figure 5.1: High level block-diagram of our automatic PSG-based drowsiness characterization system.

We now detail each of the processing blocks.

5.3.1 Preprocessing

At the beginning of our doctoral thesis, we decided to work on correcting ocular artifacts (OAs) in the EEG recordings and to develop corresponding algorithms. EEG signals are indeed often contaminated by OAs and especially in studies like ours where the participants are awake and perform a visual task. Therefore, in order to properly interpret and analyze the EEG signals, it is desirable to detect the artifacts, remove them or, better, to correct them. We have thus identified several methods for dealing with OAs, and we also created new ones. We published in "Signal Processing", in 2014, a paper giving a detailed account of all the methods implemented (the reference can be found in the publications chapter at the end of this report).

However, throughout this thesis, when processing EEG signals by computer, we decided not to add a step for processing OAs because our methods to analyze these signals and determine an LoD are quite robust to artifacts and, in addition, it would have removed a significant amount of our data, which is already limited. Another argument is that an OA can sometimes be seen as a sign of awakening as the individual blinks.

Therefore, the EEG and EOG signals are simply filtered using a Butterworth bandpass filter with cutoff frequencies of 1 and 25 Hz.

5.3.2 Extraction of features via signal processing and analysis

The features that we need to extract from EEG signals are based on the presence or absence of alpha and theta frequency bands. The approach that comes directly to mind is to use the Fourier Transform (FT), which provides the frequency domain description of a (time-domain) signal. This technique enables one to provide a good description of stationary signals but has limitations for non-stationary signals such as EEG and EOG. Stationarity is indeed a hypothesis underlying in many time-frequency signal analysis methods, and one should keep this in mind. We therefore considered using the Short Time Fourier Transform (STFT), which provides an analysis of the signals on sliding windows where the signals are assumed to be stationary. This is the simplest method for frequency analysis of non-stationary signals. However, the choice of the length of the window imposes a compromise between frequency and time resolution. If one wants a high frequency resolution, one must choose a large window, which will decrease the temporal resolution, and vice versa. This is an important issue for our application. Indeed, the goal of our system is to detect the first signs of drowsiness in order to prevent performance decrements. The first signs of drowsiness are often quite short in time (e.g. small bursts of alpha rhythm) and they are detected via small variations in frequency. We thus need to be accurate in time and in frequency. In addition, the choice of the type of window (i.e. rectangular, hamming, among others) used will influence the results, which is not ideal.

After considering the STFT, we considered several different methods through the thesis work of several Master Students.

With the help of Daniela Mazzeo, we developed a system mainly using spectral analysis and neural networks to determine drowsiness. The overall method consists in extracting several features from the EEG signals using non-parametric spectral analysis methods (such as the periodogram or Bartlett-welch) and parametric spectral analysis methods (such as the Covariance or the Burg). These features were then used as inputs of a neural network in order to output 3 stages of drowsiness. The results were interesting but not conclusive because the dataset used at the time of this Master thesis was too small for these kind of methods.

With the help of Quentin Massoz, we developed a system mainly using the wavelet packet decomposition technique. This technique enables the analysis of a signal using a basic function called “mother wavelet” that is properly scaled in order to detect a particular feature in the signal. The goal of this work was to use the wavelet packet decomposition to estimate the energies of the different activities in the EEG (without using machine learning techniques, to avoid the need for training) and the KDS scale to derive a drowsiness score per minute. The different methods developed showed very promising results. However, the wavelet transformation also implies a compromise between frequency resolution and temporal resolution. We indeed have a greater temporal resolution during the analysis of high frequency components, and a lower one during the analysis of low frequency components. In addition, the choice of the mother wavelet at

the beginning can be crucial.

We then decided to explore in depth the concept of instantaneous frequency, which seemed to be the ideal concept to use in our framework of EEG signals. This concept is intimately linked to the Hilbert Transform (HT).

The HT actually enables one to transform a real valued signal into a complex signal by generating a quadrature phase component to serve as an imaginary part. The signal then becomes analytic, $z(t)$, and it has no negative frequency components. Mathematically, the HT of a real signal $x(t)$ is obtained by the convolution between $x(t)$ and a function $h(t) = \frac{1}{\pi * t}$. The advantage of the analytic signal is that it eases the task of computing the instantaneous characteristics (frequency and amplitude) of the underlying real signal [144].

Indeed, in continuous time, an analytic signal takes the form:

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)},$$

where $y(t)$ is the HT of $x(t)$ defined by:

$$h\{x(t)\} = y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(s)}{t - s} \, \mathrm{d}s,$$

where P is the principal Cauchy value.

The instantaneous amplitude is given by

$$a(t) = \sqrt[2]{X^2(t) + Y^2(t)} = |Z(t)|.$$

The instantaneous phase is given by:

$$\theta(t) = \arctan(\frac{Y(t)}{X(t)}).$$

The instantaneous frequency is then obtained as the derivative of $\theta(t)$, strictly speaking,

$$\omega(t) = \frac{d\theta(t)}{dt}.$$

However, the computation of the instantaneous frequency can be done only if the signal is mono-component, i.e. if it has only a single frequency at each moment. For multicomponent signals, such as EEG signals, it is therefore necessary to carry out a preliminary decomposition of the signal.

To carry out this decomposition, we identified two methods, i.e. the Hilbert Huang Transform (HHT) and Hilbert Vibration Decomposition (HVD). We investigated the use of HHT and HVD through the Master thesis work of Vincent Bosch.

The HHT involves a decomposition in empirical modes (which are mono-components), also called Intrinsic Mode Functions (IMFs), and the application of the HT on each resulting mono-component. This method was developed by Huang et al. in 1998 to process non-linear and non-stationary signals [145]. It consists in adaptively and empirically decomposing the signal into a sum of oscillating components, each at a single frequency. The decomposition identifies the successive, discrete components from the one with the highest (instantaneous) frequency to the one with the lowest. The instantaneous frequency and amplitude of each of these components are then computed using the HT. In 2007, Sharabaty already used this method for the analysis of EEG signals to characterize drowsiness [146].

In 2006, Feldman developed another method that is conceptually similar to the HHT: the HVD [147]. This method was designed for the analysis of vibratory systems in mechanics and, to our knowledge, it does not seem to have been applied to EEG signals for drowsiness analysis. Furthermore, it is only very recently (2018) that the HVD was applied to EEG signals, for the detection of dysfunctioning in people with epilepsy. In 2008, Feldman made a comparison between HHT and HVD [148]. HVD showed better performance, notably a higher frequency resolution and a better decomposition into mono-component. In addition, the HVD method suits more to our application than the HHT. The reasons are found below.

Therefore, we only further consider the HVD. Its description follows.

The HVD method is based directly on the analytic representation of the overall signal in the complex plane. The key hypothesis made by Feldman with the proposition of the HVD, is that the evolution of the phase of the overall signal is driven by the dominant component of the signal and that the other components will oscillate around it [149]. Figure 5.2 illustrates this concept. The method therefore consists, as already alluded to above, in extracting the dominant component, and in repeating this operation iteratively until having found all the mono-components. This method is very interesting because it is in fact very similar to what an expert would do to detect the features of interest in EEG signals.

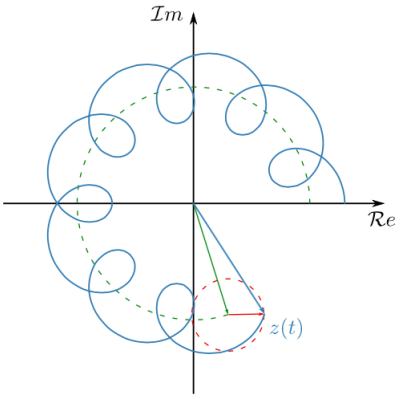

Figure 5.2: Representation of an analytic two-components signal with constant amplitude and frequency in the complex plane [3].

Specifically, the HVD technique implies the application of the HT to the overall (real) signal to obtain a (complex) analytic representation thereof. From this representation, one can compute the instantaneous frequency of the overall signal. If one takes the particular case of a signal containing only two harmonics, Feldman proved in his book that the instantaneous frequency, computed for that signal, contains two parts, i.e. a slow-varying frequency representing the first component and a rapidly varying frequency representing the other components [4]. If one then applies an integration operation, the rapidly varying frequency part disappears. For a signal with more components, the expression for the instantaneous frequency is more complicated but, using a low pass filter, one obtains the instantaneous frequency of the dominant component. Then, to obtain the instantaneous amplitude (the enveloppe of the dominant component), a synchronous demodulation is used. This technique consists in multiplying the overall signal by two reference signals exactly 90 deg out of phase with one another and by lowpass filtering each resulting signal and taking the square root of the sum of their squares. The dominant component can thus be reconstructed and subtracted from the overall signal (this is the sifting process) to repeat the operation and to find all the signal mono-components.

The HVD method therefore corresponds to a decomposition into a sum of components with slow-varying instantaneous amplitudes and frequencies so that

$$x(t) = \sum_{l} A_l(t)\, cos\,(\int \omega_l(t)dt),$$

where A_l and $\omega_l(t)$ are the instantaneous amplitude and frequency of the l^{th} component respectively.

Figure 5.3 shows the block diagram summaryzing the global HVD method.

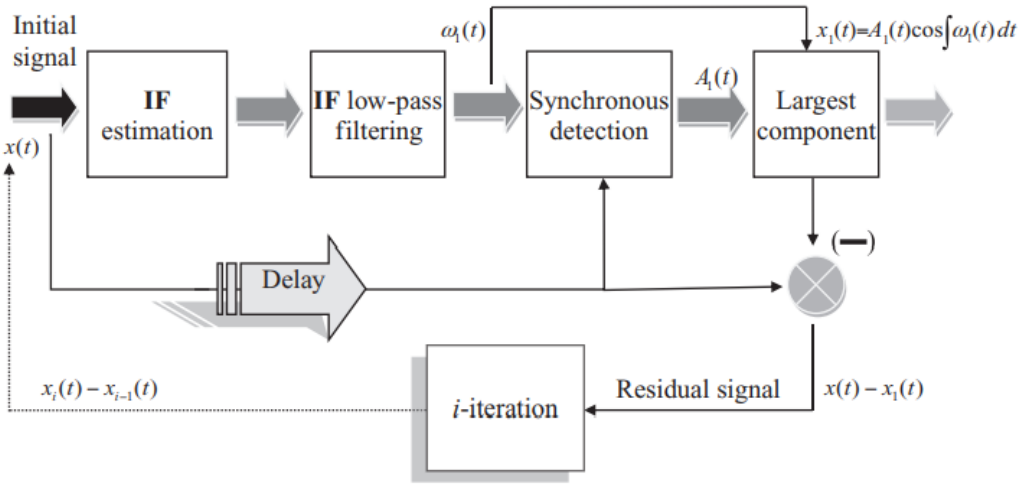

Figure 5.3: Block diagram of the HVD global method [4].

We now move on to the application of this method to our EEG signals.

As a reminder, the features of interest for the analysis of drowsiness using the KDS scale in EEG signals are:

- the presence of alpha rhythm in EEG signals. Alpha rhythm is more present in the occipital region of the brain. We therefore analyze the Pz channel of the EEG, which is the closest to the occipital region among the channels we use.
- the presence of theta activity in the EEG signals. Theta is rather present in the frontal region of the brain. We therefore analyze the Fz channel of the EEG.

We start with the analysis of an EEG signal using the HVD method. As we have just explained, the HVD method enables one to extract the dominant component of the signal, which is the component with the largest energy. Since the standard visual analysis of EEG signals implies to also extract the dominant frequency band present in the signals, we decided to limit ourselves to the analysis of this single dominant component.

For each 20-second window of the EEG signal, we therefore extract the dominant component via the method of Feldman, described above, and we thus obtain the instantaneous frequency and the instantaneous amplitude for the dominant component. Figure 5.4 shows an example of a result. Then, for each 2-second sub-window of the 20-second window, we compute several attributes that are used as inputs to an automatic learning system, the purpose of which is to determine whether the 2-second window can be labeled as containing a sign of drowsiness or not.

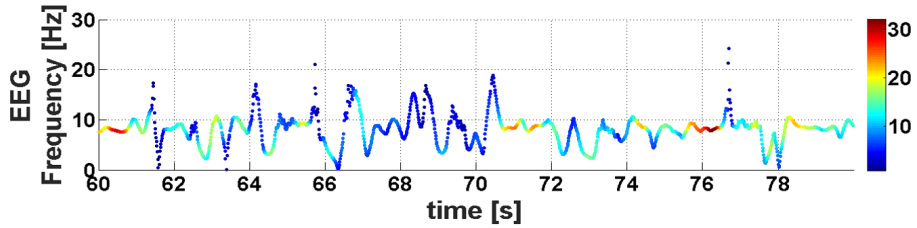

Figure 5.4: Example of a dominant component extracted from an EEG signal using the HVD method.

The extracted attributes for each 2-second sub-window and for each sign of drowsiness (i.e. alpha rhythm or theta activity) are as follows:

- the number of samples of the dominant component that are in the frequency band of interest;
- the average amplitude of samples of the dominant component that are in the frequency band of interest;
- the temporal variation of samples of the dominant component that are in the frequency band of interest;
- the variance of the frequency of samples of the dominant component that are in the frequency band of interest;
- the variance of the frequency in the 2-second sub-window;
- the average frequency in the 2-second sub-window.

The above list contains attributes that were not present in the initial work performed by Vincent Bosch in his Master thesis.

We continue with the analysis of the EOG signal to detect slow eye movements. Here we combine two methods, i.e. a method that detects fast blinking of the eye in the EOG via its derivative, and the HVD method that detects slow eye movement.

As a reminder, the feature of interest for the analysis of drowsiness using the KDS scale in the EOG signal is:

- the presence of slow eye movements in the vertical EOG signal.

The derivative of the EOG signal has already been used extensively in the literature for the analysis of the EOG (and particularly for the detection of blinks) [146]. The equation used here to compute/estimate the derivative is the first difference

$$dx[n] = \frac{x[n] - x[n-1]}{T}.$$

The derivative provides a signal representing the velocity of eye movements. In this derivative-signal, the blinks are characterized by abrupt rising edges followed by falling edges. Figure 5.5 illustrates the EOG signal and its derivative computed according to the above formula. Therefore, by using a method of peaks detection (extrema) in the signal and applying relevant (typically fixed) thresholds, it is possible to detect the successive closing and opening phases of the eye during a blink and thus to calculate the duration of a blink and the speed of closing and opening of the eyelids of this blink. However, this method does not work well for the detection of slow eye movements because the peaks in the derived signal are much smaller, sometimes almost non-existent, so that they cannot easily and reliably be detected.

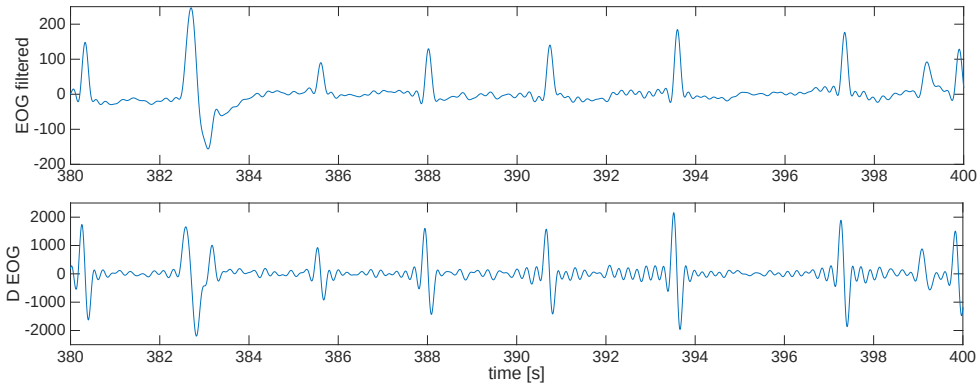

Figure 5.5: Vertical EOG signal (above) and its derivative (below).

The extracted attribute for each 2-second sub-window is:

- the number of fast blinks.

Regarding the analysis of slow eye movements, we decided to use again the HVD method. Slow eye movements are indeed characterized by a very low frequency and a high amplitude, which are characteristics of the dominant component that can be extracted with the HVD method.

The extracted attributes for each 2-second sub-window are as follows:

- the number of samples of the dominant component with a sufficiently low frequency and a sufficiently high amplitude;
- the average amplitude of samples of the dominant component that are in the frequency band of interest;
- the temporal variation of the samples of the dominant component that are in the frequency band of interest;
- the average frequency in the 2-second sub-window.

5.3.3 Conversion into an automatic PSG-based LoD

Once all the attributes are computed, it is necessary to decide whether each 2-second sub window can be labeled as containing a sign of drowsiness or not. To do this, we decided to create detectors for each sign of drowsiness (i.e. an alpha detector, a theta detector, and a SEMs detector) using a machine learning technique called decision tree or classification tree. This is a supervised machine learning method. We use a set of data (or attributes) for which we know the target value in order to build/train a tree (a model). Then, we extrapolate the results to a set of test data that were not used to build the model. The tree is composed of branches, nodes, and leaves. At each node of a tree, there is a test on the input variable that best separates the training set. The selection takes place according to a very particular criterion. We used the index of Gini diversity. This index measures the probability that a variable be selected but misclassified. A branch then contains the result of the test and the leaf at the end of the tree contains the target value, 0 or 1 in the present case.

In practice, one hopes that the training set will at best represent a broad population and that the tree will be able to generalize the decision to new datasets. Therefore, we must build the simplest possible tree that can properly segment and separate the training set without being too specific to this dataset so that it can be extrapolated to new, previously unseen data.

Decision trees have the advantage of being simple to interpret and were evaluated as a promising technique for brain computer interface (BCI) in a study comparing several classification techniques for EEG-based BCIs [150].

For the alpha detector, we therefore trained a decision tree model taking as inputs the attributes related to the detection of the alpha rhythm and using the visual scoring of the signals as the reference. We did the same for the theta detector and the SEMs detector with their respective attributes. The learning was done with a training set containing a fair distribution of the presence or absence of signs of drowsiness.

Each detector gives a binary output, i.e. either 0, which means there is no sign of drowsiness in the 2-second sub-window, or 1, which means there is at least one sign of drowsiness in the 2-second sub-window. We can then easily determine a KDS score per window of 20 seconds by adding all the scores obtained for each 2-second sub-window, knowing that, even if several signs of drowsiness are found for the same 2-second sub-window, the overall score will not be increased, i.e. it will stay at 1. The final result for a 20-second window is therefore a numerical value between 0 and 10.

Figure 5.6 shows the complete and detailed block diagram of the whole automatic PSG-based drowsiness characterization system.

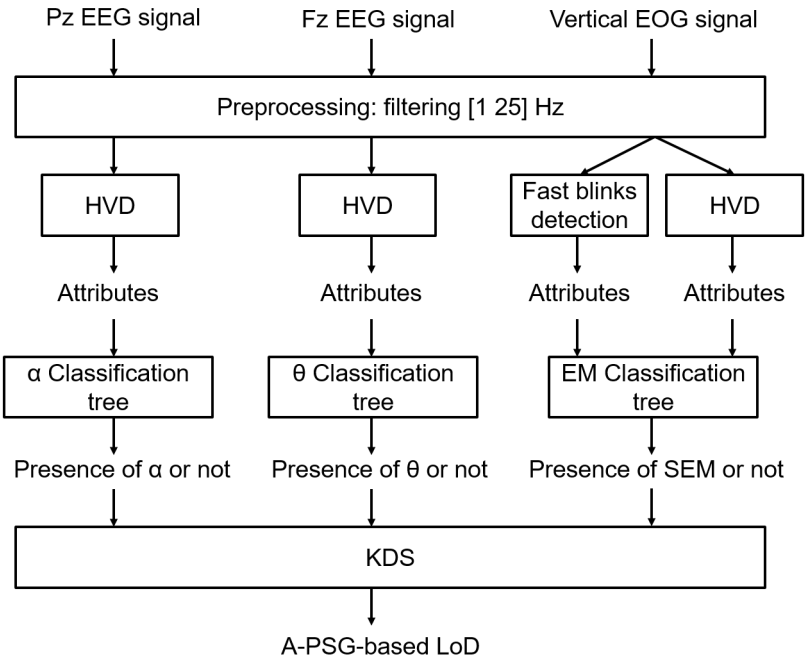

Figure 5.6: Detailed block diagram of our automatic PSG-based drowsiness characterization system.

5.3.4 Performance evaluation methods

The main goal of the system developed in this chapter is to automatically and objectively provide an LoD from EEG and EOG signals using the KDS scale. We therefore analyze the results provided by this automatic system by comparing them with the results obtained via our gold standard, which is the visual analysis of EEG and EOG signals.

However, before presenting these results, we begin by independently evaluating each detector of drowsiness signs. Indeed, as explained above, the global method developed in this thesis uses the HVD to analyze the EEG and EOG signals. Then, specific attributes are extracted for each sign of drowsiness and a model/tree is built for each such sign of drowsiness to determine whether or not the sign is present within a given time window. The result of this determination is then used to provide an overall drowsiness score for the considered time window.

Evalution of the performance of each detector of a sign of drowsiness

In order to evaluate the reliability and accuracy of our decision trees for new data, we use a 10-fold cross-validation. Cross-validation enables one to estimate the model performance on available data that is not used for training. It also ensures that our model does not overfit. This technique is particularly useful for small datasets.

Concretely, cross-validation consists in randomly dividing the training set into ten equal

subsets. Nine of them are then used to learn/train a new tree and the tenth one is used to test. We iterates the operation so that each subset has been used once for the test. The performance of each tree is examined to determine the output on a set that has not been used for training. A global measure of the performance of our model is obtained by averaging the performance of each tree built during the cross-validation.

This method gives a good estimate of the performance and reliability of the final model built with all the data.

Relation between the automatic PSG-based LoD and several references

After evaluating the performance of each detector of a sign of drowsiness, we assess the LoD determined by our automatic PSG-based characterization system. To do this, we compare (i.e. correlate) it to several references:

- the visual/manual PSG-based LoD,
- the percentage of lapses.

To illustrate the relations between them, we use scatter plots.

5.4 Experimental results and performance evaluation

For all results, we consider the data windows from all PVTs and all participants (Dataset A), regardless of their sleep deprivation condition.

For each result, we used 2128 overlapping 1-minute windows of data from 26 participants for Dataset A.

5.4.1 Evaluation of the performance of each detector of a sign of drowsiness

Figure 5.7 shows the performance of each detector in terms of sensitivity and specificity. The best results are for the alpha detector, which is not surprising. Indeed, the alpha bursts in the EEG signals are generally well recognizable as they look like pure sine waves oscillating at a frequency in the alpha band and they are almost not tainted by noise or other frequencies. It is therefore easier for a scorer to detect them, and this is also the case for an algorithm. The weakest results concern the detector of theta activity. This activity is indeed the most difficult pattern to recognize in the signals as, in contrast with alpha rhythm, the wave is more contaminated by other frequencies and it is also more likely to be confused with artifacts. Moreover, in our dataset, the theta activity was not much present, which also reduces the size of the training set.

Detectors	Sensitivity	Specificity
Alpha detector	0.8960	0.8663
Theta detector	0.7200	0.7625
SEMs detector	0.8805	0.8441

Figure 5.7: Results of cross-validation of each detector of a sign of drowsiness.

5.4.2 Relation between the automatic PSG-based LoD and several references

First, we compare our automatic PSG-based LoD to the manual PSG-based LoD. One can notice that our automatic PSG-based LoD slightly overestimates drowsiness as compared to the manual PSG-based LoD. In fact, to distinguish between awake and drowsy, the KDS scoring method, on which we based our manual PSG-based LoD, states that we have to set the threshold at 50 on a scale from 0 to 100. However, if one compares the automatic PSG-based LoD with the manual PSG-based LoD for different thresholds via a ROC curve, one realizes that one should rather put the threshold at 70 on the automatic PSG-based LoD scale to make the distinction between awake and drowsy. Figure 5.9 shows the resulting ROC curve.

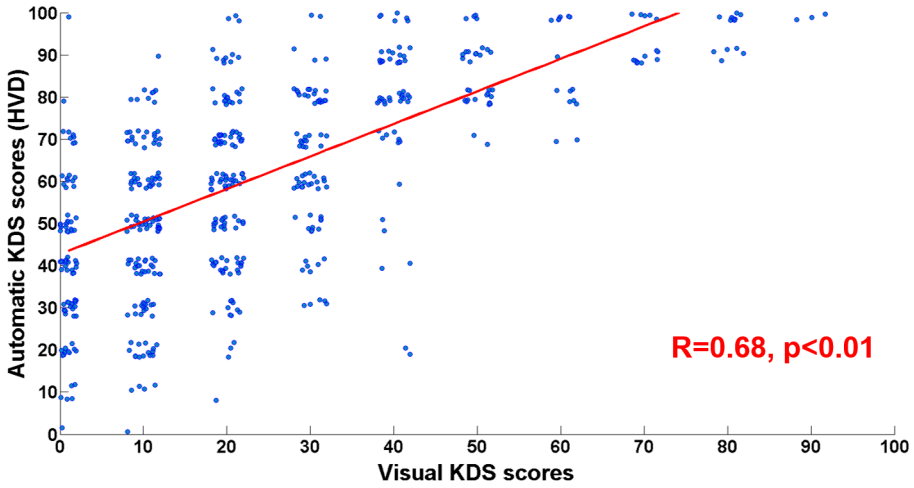

Figure 5.8: Correlation between the automatic PSG-based LoD and the visual PSG-based LoD.

In order to further compare the automatic PSG-based LoD and the manual PSG-based LoD, we also show their correlation using a scatter plot. Figure 5.8 presents the corresponding graph.

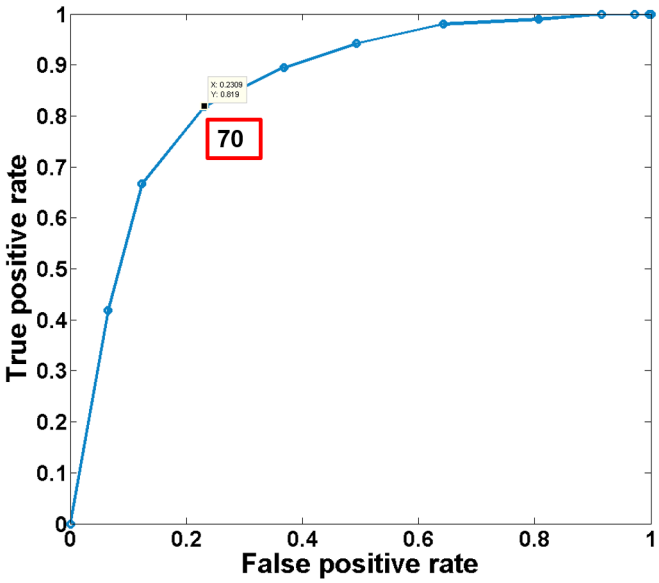

Figure 5.9: ROC curve describing the relation between the automatic PSG-based LoD and the visual PSG-based LoD for several threshold values from 0 (at top right) to 100 (at bottom-left).

Second, we continue with the relation between the automatic PSG-based LoD and the performance while executing the PVTs. Figure 5.10 shows the comparison with the percentage of lapses. One can notice an increase in the percentage of lapses with the automatic PSG-based LoD produced by the corresponding system. A correlation coefficient is also computed between the automatic PSG-based LoD and the percentage of lapses. The value is 0.56, which means that there is a positive and high correlation between the two measures.

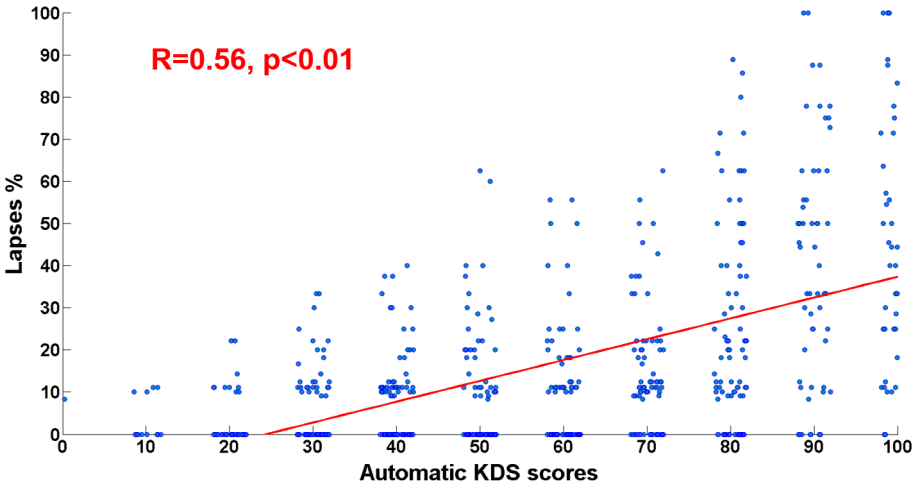

Figure 5.10: Relation between the automatic PSG-based LoD and the percentage of lapses during the PVT tests.

5.5 Discussion

We presented a new, innovative method for the automatic analysis of PSG signals to determine an individual's state of drowsiness. This method enables a very intuitive follow-up of each step of analysis and seems to correspond closely to what an expert does during a visual analysis. It is obviously difficult to imitate perfectly what a human would do, even a human could not do exactly the same analysis twice if he was asked. There is a subjective character to the interpretation of the signals, and the analysis will depend strongly on the configuration in which the signals are visualized (the scale used, etc.) as well as the prior knowledge of the data set and the protocol. The idea here was to get as close as possible to the visual analysis (which is still considered today as the gold standard) but trying to be as general as possible at the same time. Moreover, the fact that the method is very intuitive and rather open ("white box" as opposed to "black box") is a real advantage when we discuss our method with medical experts or when we want to demonstrate the feasibility of a method. Indeed, understanding each step of the system/method and having a visibility on the intermediate steps, can certainly reassure and promote the use of the system.

The KDS scoring method that we decided to use is not standardized or used by everyone. However, we chose it because it gathers the criteria most often encountered in the literature to determine drowsiness while performing a task. The criteria of the KDS were interesting for us in order to develop an automatic system but they were simple and finally very subject to interpretation and hesitation. We have therefore established stricter criteria/rules (as detailed in Chapter 3) to ensure a rigorous and reliable scoring and thus to maximize the chances of reproducibility of results. This would also enable us to ensure a good conversion into an automatic method. Rules that are too broad and

that leave too much room for interpretation cannot really be translated into algorithms. This is the case of the AASM scoring rules. We indeed tried to translate them into a pseudo code in order to automate scoring but this was not possible. Rules were established for visual/manual scoring of signals by humans, and they are not strict enough to cover all possible cases. We have therefore succeeded in correctly characterizing drowsiness via stricter criteria/rules, as compared to the KDS, and we have succeeded in translating these rules into a promising automatic method.

However, one must keep in mind that the data used here were acquired during PVT tests that are relatively short in time and that were not contaminated by too much noise. Indeed, in other more difficult contexts/circumstances, some patterns in the EEG signals are very difficult to identify even for an experienced scorer. For example, the individual may move or speak during a data recording, which has the effect of creating an artifact in the EEG signals that could be mistaken for a sign of drowsiness. It is thus sometimes necessary to visualize a video of the behavior of the individual in parallel with the scoring to identify specific patterns correctly. This being said, we could very well imagine a combination of an automatic EEG signals analysis with an automatic video analysis to fully characterize the state of an individual.

Regarding our models (trees), the results could certainly be improved if we had a larger and more complete database including a greater variability in the population (such as different age groups) and more data for each individual. Furthermore, in our data sample, we noticed that we had much less 2-second sub-windows labeled as containing theta activity as compared to SEMs and the presence of alpha rhythm. This can be explained by the fact that theta activity is a sign of more advanced drowsiness than alpha rhythm. The task lasting only 10 minutes, even with relatively high sleep deprivation, not all participants entered a phase of advanced drowsiness. This difference is important for two reasons. The first technical reason is that we therefore have fewer examples of theta activity in our database and so our model was trained with little data and could maybe detect theta activity as not as accurately as alpha rhythm. The second reason is that if theta activity is a sign of more advanced drowsiness as compared to alpha rhythm, perhaps we should put more weight into our drowsiness score when we detect theta activity versus alpha rhythm. The KDS scoring method does not do this, indeed the weight is the same if we detect slow eye movement, alpha rhythm or theta activity. This could be an improvement for future developments. Nevertheless, as we want to prevent drowsiness related-accidents, and thus to detect early signs of drowsiness, it is very important to detect alpha rhythm.

Notwithstanding the above remarks, the three detectors developed showed very good results. This is very interesting because they could probably be used in contexts and applications other than drowsiness, and because other detectors could also be created to detect other patterns of interest, such as the spindles for the detection of sleep stage 2.

The results regarding the comparison with the PVT parameters are also interesting and conclusive. Indeed, the percentage of lapses is objective data and completely indepen-

dent of the training of our model.

5.6 Summary of chapter

The automatic PSG-based drowsiness characterization system that we developed showed good performance when compared to the visual scoring of the signals and to the percentage of lapses. This system thus has the potential to become a reference for drowsiness quantification, thereby reducing the need for experts to always establish visually/manually the LoD in new conditions, and, also saving a lot of time. However, in addition to its role as a reference tool for validating other drowsiness characterization systems, this automatic PSG-based drowsiness characterization system could also be used as a diagnostic tool for people with excessive daytime sleepiness (EDS), which may be due to sleep disorders.

Each detector of a sign of drowsiness also showed very good performance. This is very interesting because these detectors could be used to detect alpha rhythm, theta activity, and eye movements in contexts other than drowsiness.

Chapter 6

Conclusion

The main objective of this present thesis was to develop an automatic, objective, and real-time drowsiness characterization system that is usable in operational environments. In today's society, drowsiness still has a significant impact on the number of accidents in transportation and industry. This is an important health and safety issue that is not about to disappear. It is therefore necessary to develop solutions to avoid the sometimes dramatic consequences of drowsiness.

Drowsiness remains a very complex phenomenon and all the experts do not yet agree either on its definition nor on which method is best to characterize it. Therefore, we proposed in Chapter 2 a non-exhaustive classification of several methods for characterizing drowsiness. We provided an analysis of the advantages and disadvantages of each method in order to identify the most adequate method to develop a drowsiness characterization system that can be used in operational environments. The method that was selected is the analysis of images of the eye to extract ocular parameters indicative of drowsiness and to determine a level of drowsiness (LoD). The advantage of this method is that it enables an objective, automatic, real-time, non-invasive, and physiologically-based (and therefore independent of the task performed by the individual) characterization of drowsiness.

We presented the development of this system in Chapter 4. We initially focused on a system worn by the individual (in the form of a pair of eyeglasses) that takes images of the eye at a high frame rate and at a high resolution. We did not deal with the development of image analysis algorithms in this doctoral thesis, but more with the computation of ocular parameters indicative of drowsiness from the data extracted from the images (eyelids and pupil positions at each image) and their conversion into a LoD on a numerical scale from 0 to 10. This is the first, multi-parameters, end-to-end system allowing such a fine analysis of drowsiness from images of the eye. Indeed, the other systems developed in the literature and, which are similar to ours, generally use a single ocular parameter, the PERCLOS. As a reminder, the PERCLOS computes the percentage of time that the eye remains closed more than 70%. As discussed in Chapters 2 and 4, the PERCLOS is indeed a relevant ocular parameter indicator of drowsiness but it does not allow a very fine and very preventive analysis of drowsiness. In addition, there are

many other ocular parameters that are relevant for the characterization of drowsiness and combining them all together enables to detect more precursory signs of drowsiness. We showed that the overall LoD that we provide with our POG-based drowsiness characterization system is more relevant than the PERCLOS alone. The purpose of our system is to detect early signs of drowsiness to warn the individual before he is no longer able to react to a risky situation.

We developed very modular algorithms so as to be able to exploit and adapt them to the case of images acquired by a remote camera. Indeed, at the end of Chapter 4, we showed a first adaptation of our POG-based drowsiness characterization system to process face images instead of eye images. The face images were acquired by a remote camera at a low frame rate and a low resolution. This is important because, the automotive sector, for example, is more interested in remote drowsiness characterization systems that can be integrated into the vehicle cabin (i.e. the dashboard) and that are thus not in contact with the driver.

Through the analysis of the various drowsiness characterization methods detailed in Chapter 2, we also selected different methods to validate the POG-based drowsiness characterization system we developed.

Another objective of this thesis was indeed to show that our POG-based drowsiness characterization system was relevant in comparison with several references (i.e. gold standards). As there is no universal agreement/consensus on which reference to use, we decided to select several of them. Some experts do believe that the reference must be related to the analysis of polysomnographic (PSG) signals because (1) drowsiness is fundamentaly a physiological phenomenon and, (2) physiology-based references hold true in any situation and context. However, other experts consider that a drowsiness characterization system must be designed to avoid accidents and therefore must be evaluated on the basis of its ability to detect performance decrements and thus prevent accidents. If we refer to the case of driving as an example, the system will be successful if it warns a driver before he crosses the lines of his driving lane. Finally, a third group of experts in the literature uses the Karolinska Sleepiness Scale (KSS) as a reference.

We therefore used the following references to validate our POG-based drowsiness characterization system:

- the visual analysis of PSG signals using the Karolinska Drowsiness Scale (KDS);
- the analysis of reaction times and percentage of lapses when performing Psychomotor Vigilance Tests (PVTs);
- the analysis of the standard deviation of the lateral position (SDLP) of a vehicle on the road when performing a monotonous driving task in a professional simulator;
- the analysis of the KSS.

In order to proceed with the development of our POG-based drowsiness characterization system and with its validation, we laid out significant groundwork for setting up a data acquisition. This has been detailed in Chapter 3 of this doctoral thesis. It took us several iterations to obtain the test protocol that we finally used.

Thanks to all the data acquired, we were able to show that the LoD determined by our POG-based drowsiness characterization system correlated well with (1) the LoD determined visually from the analysis of the PSG signals, (2) the performance decrements observed during the execution of the different tasks and, (3) with the KSS.

Our POG-based drowsiness characterization system has a lot of potential in different business sectors. Indeed, drowsiness affects the transport sector (automotive, professional, air, railway, etc.), but also the industry sector (control and surveillance rooms). In particular, a spin-off named Phasya has been founded in order to exploit the POG-based drowsiness monitoring system that we developed and to put it on the market.

We also concentrated our efforts in this doctoral thesis on the analysis of the PSG signals. It is indeed the physiological reference used for the automatic analysis of sleep and, by extension, for that of drowsiness. The standard method used in all sleep clinics is the visual analysis of PSG signals, which is time consuming. Therefore, we decided to develop an automatic reference PSG-based drowsiness characterization system. We presented the development of this system in Chapter 5. In this Chapter, we introduced the innovative methods of signal processing and machine learning that we used, i.e. the Hilbert Vibration Decomposition technique and the classification-tree technique. We particularly detailed how we developed the complete automatic PSG-based drowsiness characterization system by combining several detectors of a sign of drowsiness (i.e. an alpha rhythm detector, a theta activity detector, and a slow eye movement detector). We demonstrated that each detector of a sign of drowsiness was relevant in comparison with the visual analysis of PSG signals and that the overall LoD produced by the complete PSG-based system correlated well with the references.

The primary goal of our PSG-based drowsiness characterization system is to be easily used as a universal reference to validate other drowsiness characterization systems. However, with the advancement of technology, one could probably imagine in the future a way to recover brain activity in a more simple and less invasive way than with electrodes and thus be able to directly use this system in operational environments for characterization of drowsiness. In addition, our system could certainly also be used for applications other than the characterization of drowsiness, which could save time to doctors and experts for the diagnosis of certain pathologies and for the scoring of nights of sleep.

The work done in this doctoral thesis was multidisciplinary. We gathered several skills, expertise, and knowledge to develop a relevant solution for the characterization of drowsiness in operational settings. We indeed worked with:

- engineers and computer scientists for the development of image processing algo-

rithms, and for the development of the infrastructure for the acquisition of data;

- physicians and medical doctors to properly characterize the phenomenon of drowsiness, to develop our methods for the visual analysis of PSG signals, to elaborate our data acquisition protocol, and to analyze PSG signals;
- psychologists specialized in human factors to elaborate our data acquisition protocol, and discuss the analysis of the performances of individuals in the execution of a task.

In order to continue the developments of this doctoral thesis, it would be interesting to translate what we did for drowsiness into the characterization of other cognitive/physiological states that influence an individual's ability to perform a task. Indeed, the global attention level (i.e. vigilance level) of an individual directly impacts/alters his performance while executing a task. The attention level is heavily influenced by drowsiness but other cognitive/physiological states also have an influence (e.g. stress). The idea would be to have a global system of characterization of the attention of an individual that would be composed of several detectors of different cognitive/physiological states. The advantage of independent detectors is that if we know the origin of the loss of attention, we can take appropriate action. For example, if one detects that an individual is drowsy, the individual will be told to stop the car and have a nap. By contrast, if one detects that an individual is stressed, one will try to relax him in another way (e.g. diffusion of relaxing scents, massage in the seat of the car).

Publications

In each of the three categories, the publications are listed in reverse chronological order, i.e. starting with the latest ones.

Papers in peer reviewed academic journals:

- C. François, T. Hoyoux, T. Langohr, J. Wertz, J Verly. Tests of a new drowsiness characterization and monitoring system based on ocular parameters. *International Journal of Environmental Research and Public Health*, 13(2):174, 2016.
- M. Kirkove, C. François, J. Verly. Comparative evaluation of existing and new methods for correcting ocular artifacts in electroencephalographic recordings. *Signal Processing*, 98(C):102-120, 2014.

Scientific congresses and symposia, on invitation:

- C. François, T. Hoyoux, T. Langohr, J. Wertz, J.G. Verly. Real-time, automatic, and objective monitoring of drowsiness using images of the eye. *Seminar presented at the Washington State University, Sleep and Performance Research Center*, Spokane, USA, 24 July 2015.
- J. Wertz, C. François, J. Verly. Drowsiness monitoring for road safety. *Research presented at Plateforme de recherche en matière de sécurite routière*, Brussels, Belgium, 13 December 2013.

Scientific congresses and symposia, on a personal proposal:

- C. François, J. Wertz, J. Verly. Relationship between brain activity and ocular movements during wakefulness and drowsiness. *Poster presented at Joint Congress of Association of Sleep Medicine and World Sleep Federation (Worldsleep 2017)*, Prague, Czech Republic, October 2017.
- C. François, Q. Massoz, T. Hoyoux, J. Wertz, J. Verly. First adaptation of a validated drowsiness monitoring system to process face images instead of eye images. *Paper presented at 10th International Conference on Managing Fatigue*, San Diego, USA, 20 March 2017.
- C. François, J. Wertz, T. Hoyoux, T. Langohr, J. Verly. Objective drowsiness monitoring for assessing the ability of an operator to perform a task. *Paper presented at the BASS Autumn meeting*, Brussels, Belgium, 9 December 2016.

- C. François, T. Hoyoux, T. Langohr, J. Wertz, J. Verly. Objective drowsiness monitoring to assess fitness for duty. *Poster presented at 23rd Congress of the European Sleep Research Society*, Bologna, Italy, 16 September 2016.

- Q. Massoz, T. Langohr, C. François, J. Verly. The ULg Multimodality Database (called DROZY) and examples of use. *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, USA, March 2016.

- C. François, V. Bosch, Q. Massoz, B. Fortemps de Loneux, R. Poirrier, J. Verly. Development of an automated reference approach for quantifying drowsiness using polysomnographic signals. *Paper presented at SomnoSafe 2016, International Symposium on Somnolence and Safety*, Brussels, Belgium, 22 February 2016.

- C. François, V. Bosch, Q. Massoz, B. Fortemps de Loneux, R. Poirrier, J. Verly. Development and validation of an automatic reference polysomnographic system for quantifying drowsiness. *Poster presented at The 7th World Congress of the World Sleep Federation (Wordsleep 2015)*, Istanbul, Turkey, 3 November 2015.

- C. François, T. Hoyoux, T. Langohr, J. Wertz, J. Verly. Test of a new drowsiness monitoring system based on ocular parameters. *Paper presented at 6th International Conference on Applied Human Factors and Ergonomics*, Las Vegas, USA, 30 July 2015.

- P. Berastegui, C. Piette, C. François, T. Langohr, A. Blavier, J. Wertz, J. Verly, A.-S Nyssen. How novice and expert drivers adjust their driving behavior when they feel drowsy? *Poster presented at 6th International Conference on Applied Human Factors and Ergonomics*, Las Vegas, USA, 29 July 2015.

- C. François, T. Langohr, T. Hoyoux, J. Wertz, J. Verly. Test of a new drowsiness monitoring system based on images of the eye. *Paper presented at 9th International Conference on Managing Fatigue*, Perth, Australia, 24 March 2015.

- C. François, J. Marchat, P. Latour, J. Wertz, R. Poirrier, J. Verly. Comparison of the levels of drowsiness obtained via a new photooculography-based drowsiness scale and via a simple variation of the Karolinska Drowsiness Scale (KDS). *Poster presented at 22nd Congress of the European Sleep Research Society*, Tallinn, Estonia, 17 September 2014.

- C. François, J. Wertz, M. Kirkove, J. Verly. Evaluation of the performance of an experimental somnolence quantification system in terms of reaction times and lapses. *Paper presented at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society and published in its proceedings*, Chicago, USA, August 2014.

- J. Wertz, C. François, J. Verly. Validation of a new automatic drowsiness quantification system for drivers. *Poster presented at 5th International Conference on Applied Human Factors and Ergonomics (AHFE)*, Krakow, Poland, July 2014.

- C. François, J. Wertz, J. Verly. Prevalidation of a new drowsiness quantification system based on ocular parameters using polysomnography and performance measures. *Poster presented at Biomedica 2014*, Maastricht, The Netherlands, 17 June 2014.

- A. Blavier, C. Montagnino, J. Wertz, C. François, J. Verly, A.-S Nyssen. Detection of eye movements specific to drowsiness and their relation with subjective assessment: A cognitive ergonomic approach. *Poster presented at Annual Meeting of the Belgian Psychological Society*, Leuven, Belgium, 27 May 2014.

- A. Blavier, C. Montagnino, J. Wertz, C. François, J. Verly, A.-S Nyssen. Detection of eye movements specific to drowsiness and their relation with subjective assessment of sleepiness in PVT: A cognitive ergonomic approach. *Paper presented at SomnoSafe 2014, International Symposium on Somnolence and Safety*, Brussels, Belgium, 24 February 2014.

- C. François, J. Wertz, D. Grogna, T. Hoyoux, T. Langohr, P. Latour, M. Kirkove, J. Verly. Preliminary evaluation of an experimental somnolence quantification system based on images of the eye. *Paper presented at SomnoSafe 2014, International Symposium on Somnolence and Safety*, Brussels, Belgium, 24 February 2014.

- S. Kermi, M. Kirkove, C. François, J. Verly. Study of supervised machine learning techniques to classify subjects as alert, drowsy, or other based on ocular parameters extracted from images of the eye. *Paper presented at SomnoSafe 2014, International Symposium on Somnolence and Safety*, Brussels, Belgium, 24 February 2014.

- C. François, M. Kirkove, J. Wertz, T. Langohr, R. Poirrier, J. Verly. Development of an automatic reference system for directly quantifying drowsiness from a few polysomnographic signals. *Poster presented at Sleep Health and Safety 2013*, Washington DC, USA, 1 March 2013.

- M. Kirkove, C. François, A. Libotte, J. Verly. Performance evaluation of methods for correcting ocular artifacts in electroencephalographic (EEG) recordings. *Paper presented at 6th International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS 2013)*, Barcelona, Spain, 12 February 2013.

- C. François, J. Wertz, T. Langohr, A. Blavier, R. Poirrier, D. Grogna, J. Verly. Experimental design for a study of drowsiness using a driving simulator. *Poster presented at Sleep Well - Live Well*, Brussels, Belgium, 5 October 2012.

- C. François, J. Wertz, J. Verly. Development of an automatic drowsiness monitoring system based on EEG and EOG signals, and preliminary tests in a professional driving simulator. *Poster presented at Biomedica 2012*, Liège, Belgium, 18 April 2012.

- J. Wertz, C. François, J. Verly. Drowsiness monitoring: a matter of life and death! *Paper presented at Biomedica 2012*, Liège, Belgium, 19 April 2012.

Bibliography

[1] G. S. Richardson. The human circadian system in normal and disordered sleep. *Journal of Clinical Psychiatry*, 66(Suppl 9):3–9, 2005.

[2] T. Hori, M. Hayashi, and T. Morikawa. Topographical EEG changes and the hypnagogic experience. In *Sleep onset: Normal and abnormal processes*, pages 237–253. 1994.

[3] M. Bertha and J.-C. Golinval. Experimental modal analysis of a beam travelled by a moving mass using Hilbert Vibration Decomposition. In *9th International Conference on Structural Dynamics, EURODYN*, number July, pages 2789–2796, 2014.

[4] M. Feldman. *Hilbert Transform applications in mechanical vibration*. 2011.

[5] M. J. Thorpy and M. Billiard. *Sleepiness: Causes, consequences and treatment*. Cambridge University Press, 2011.

[6] N. Goel, H. Rao, J. Durmer, and D. F. Dinges. Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, 29(4):320–339, 2009.

[7] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. The impact of driver inattention on near crash/crash risk: an analysis using the 100-car naturalistic driving study data. Technical report, 2006.

[8] National Highway Traffic Safety Administration (NHTSA). NHTSA Drowsy Driving Research and Program Plan. Technical report, 2016.

[9] Association des sociétés françaises d'autoroutes (ASFA). Chiffres clés 2017. Technical report, 2017.

[10] Association des sociétés françaises d'autoroutes (ASFA). Somnolence au volant ... pausez-vous! Technical report, 2008.

[11] Fondation d'entreprise Vinci autoroutes pour une conduite responsable. De nombreux conducteurs de poids lourds manquent de sommeil. Technical report, 2014.

[12] H. Häkkänen, H. Summala, M. Partinen, M. Tiihonen, and J. Silvo. Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep*, 22(6):798–802, 1999.

[13] J. C. Stutts, J. W. Wilkins, O. J. Scott, and B. V. Vaughn. Driver risk factors for sleep-related crashes. *Accident Analysis and Prevention*, 35(3):321–331, 2003.

[14] S. Otmani, T. Pebayle, J. Roge, and A. Muzet. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior*, 84(5):715–24, 2005.

[15] National Transportation Safety Board (NTSB). Most wanted list of transportation safety improvements. Technical report, 2017.

[16] L. Bérubé. *Terminologie de neuropsychologie et de neurologie du comportement*. 1991.

[17] R. D. Ogilvie. The process of falling asleep. *Sleep Medicine Reviews*, 5(3):247–270, 2001.

[18] R. Chervin. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea. *Clinical Investigations: Sleep and Breathing*, 118(2):372–379, 2000.

[19] A. Négroni. Autoroutes: la somnolence première cause d'accident, 2008.

[20] J. Dorrian, N.L. Rogers, and D.F. Dinges. Psychomotor vigilance performance : Neurocognitive assay sensitive to sleep loss. *Sleep (Rochester)*, 2005.

[21] M. M. Ohayon. From wakefulness to excessive sleepiness: what we know and still need to know. *Sleep Medicine*, 12(2):129–141, 2008.

[22] D.-J. Dijk and C. A. Czeisler. Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *The Journal of Neuroscience*, 15(5):3526–3538, 1995.

[23] A. A. Borbély. A two process model of sleep regulation. *Human Neurobiology*, 1(3):195–204, 1982.

[24] C. B. Saper, T. E. Scammell, and J. Lu. Hypothalamic regulation of sleep and circadian rhythms. *Nature*, 437:1257–1263, 2005.

[25] R. A. España and T. E. Scammell. Sleep neurobiology from a clinical perspective. *Sleep*, 34(7), 2011.

[26] S. P. Fisher, R. G. Foster, and S. N. Peirson. The circadian control of sleep. In A. Kramer Merrow and M., editors, *Circadian Clocks*, pages 157–183. 2013.

[27] C. Buffet. Aux limites du corps, Michel Siffre et son horloge de chair, 2004.

[28] D.-J. Dijk and C. Cajochen. Melatonin and the circadian regulation of sleep initiation, consolidation, structure, and the sleep EEG. *Journal of Biological Rhythms*, 12(6):627–635, 1997.

[29] A. A. Borbély and P. Achermann. Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, 14(6):557–568, 1999.

[30] G. Hofer-Tinguely, P. Achermann, H. P. Landolt, S. J. Regel, J. V. Rétey, R. Dürr, A. A. Borbély, and J. M. Gottselig. Sleep inertia: Performance changes after sleep, rest, and active waking. *Cognitive Brain Research*, 22(3):323–331, 2005.

[31] P. Tassi and A. Muzet. Sleep inertia. *Sleep Medicine Reviews*, 4(4):341–353, 2000.

[32] M. F. Vecchierini and D. Léger. La somnolence diurne excessive et les hypersomnies centrales primaires: données épidémiologiques. *Médecine du Sommeil*, 7(4):129–138, 2010.

[33] C. Guilleminault and S.N. Brooks. Excessive daytime sleepiness: a challenge for the practising neurologist. *Brain: a Journal of Neurology*, 124(Pt 8):1482–91, 2001.

[34] N. B. Powell, R. W. Riley, K. B. Schechtman, M. B. Blumen, D. F. Dinges, and C. Guilleminault. A comparative model: reaction time performance in sleep-disordered breathing versus alcohol-impaired controls. *Laryngoscope*, 109(10):1648–1654, 1999.

[35] M. W. Johns. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep*, 14(6):540–545, 1991.

[36] M. W. Johns. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, 15(4):376–381, 1992.

[37] M. W. Johns. A new perspective on sleepiness. *Sleep and Biological Rhythms*, 8(3):170–179, 2010.

[38] M.A. Carskadon and W.C. Dement. Sleep tendency: an objective measure of sleep loss. *Sleep Research*, 6(200), 1977.

[39] M.A. Carskadon, W.C. Dement, M.M. Mitler, T. Roth, P.R. Westbrook, and S. Keenan. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep*, 9(4):519–524, 1986.

[40] D. L. Arand, M.H. Bonnet, M. Hurwitz, M. Mitler, R. Rosa, and R.B. Sangal. The clinical use of the MSLT and MWT. *Sleep*, 28(1):123–144, 2005.

[41] M. A. Carskadon. Evaluation of excessive daytime sleepiness. *Clinical Neurophysiology*, 23(1):91–100, 1993.

[42] M. J. Thorpy, P. Westbrook, R. Ferber, P. Fredrickson, M. Mahowald, F. Perez-Guerra, M. Reite, and P. Smith. The clinical use of the Multiple Sleep Latency Test. *Journal of Sleep Research and Sleep Medicine*, 15(3):268–76, 1992.

[43] O. Andlauer, H. Moore, L. Jouhier, C. Drake, P. E. Peppard, F. Han, S. C. Hong, F. Poli, G. Plazzi, R. O'Hara, E. Haffen, T. Roth, T. Young, and E. Mignot. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. *JAMA Neurology*, 70(7):891–902, 2013.

[44] M.H. Bonnet and D.L. Arand. Arousal components which differentiate the MWT from the MSLT. *Sleep*, 24(4):441–447, 2001.

[45] L.S. Bennett, J.R. Stradling, and R.J. Davies. A behavioural test to assess daytime sleepiness in obstructive sleep apnea. *Journal of Sleep Research*, 6(2):142–5, 1997.

[46] S. Mazza, S. Launois, J.-L. Pepin, and P. Levy. The Oxford Sleep Resistance Test. In *Handbook of Clinical Neurophysiology*, pages 73–77. 2005.

[47] O. Mairesse, D. Neu, Y. Rosseel, F. Van Acker, R. Cluydts, and P. Theuns. Comparative sensitivity of outcome variables of a software-based Behavioral Sleep Resistance Task. *Industrial Health*, 47(1):80–8, 2009.

[48] T. Åkerstedt and M. Gillberg. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52:29–37, 1990.

[49] J.A. Horne and S.D. Baulk. Awareness of sleepiness when driving. *Psychophysiology*, 41(1):161–5, 2004.

[50] M. Ingre, T. Åkerstedt, B. Peters, A. Anund, G. Kecklund, and A. Pickles. Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of Sleep Research*, 15(2):142–148, 2006.

[51] C.A. Czeisler, J.K. Walsh, K.A. Wesnes, S. Arora, and T. Roth. Armodafinil for treatment of excessive sleepiness associated with shift work disorder: A randomized controlled study. *Mayo Clinic Proceedings*, 84(11):958–972, 2009.

[52] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W.C. Dement. Quantification of sleepiness: a new approach. *Psychophysiology*, 10(4):431–436, 1973.

[53] A. W. Maclean, G. C. Fekken, P. Saskin, and J. B. Knowles. Psychometric evaluation of the Stanford Sleepiness Scale. *Journal of Sleep Research*, 1(1):35–39, 1992.

[54] T. H. Monk. A visual analog scale technique to measure global vigor and affect. *Psychiatry Research*, 27(1):89–99, 1989.

[55] A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Technical report, 1968.

[56] American Academy of Sleep Medicine (AASM). The AASM Manual for the Scoring of Sleep and Associated Events, 2007.

[57] S. K. L. Lal and A. Craig. Reproducibility of the spectral components of the electroencephalogram during driver fatigue. *International Journal of Psychophysiology*, 55(2):137–143, 2005.

[58] K. Kaida, T. Åkerstedt, G. Kecklund, J. P. Nilsson, and J. Axelsson. Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. *Industrial Health*, 45(4):520–6, 2007.

[59] R.-S. Huang, T.-P. Jung, and S. Makeig. Tonic Changes in EEG Power Spectra during Simulated Driving. In *Augmented Cognition*, pages 394–403, 2009.

[60] 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 Seep Research*, 5:12–15, 1996.

[61] S. K. L. Lal and A. Craig. Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology*, 39(3):313–21, 2002.

[62] C. Papadelis, Z. Chen, C. Kourtidou-Papadeli, P. D. Bamidis, I. Chouvarda, E. Bekiaris, and N. Maglaveras. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 118(9):1906–22, 2007.

[63] T. Hori. Spatiotemporal changes of EEG activity during waking-sleeping transition period. *International Journal of Neuroscience*, 27(1-2):101–114, 1985.

[64] J.A. Horne and L.A. Reyner. Counteracting driver sleepiness: Effects of napping, caffeine, and placebo. *Psychophysiology*, 33(3):306–309, 1996.

[65] A. Anund, G. Kecklund, B. Peters, Å. Forsman, A. Lowden, and T. Åkerstedt. Driver impairment at night and its relation to physiological sleepiness. *Scandinavian Journal of Work, Environment & Health*, 34(2):142–150, 2008.

[66] D. Sandberg, T. Åkerstedt, A. Anund, G. Kecklund, and M. Wahde. Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):97–108, 2011.

[67] M.R. Risser and J.C. Ware. Driving Simulation with EEG Monitoring in Normals and Obstructive Sleep Apnea Patients. In *Annual Proceedings of the Association for the Advancement of Automotive Medicine*, volume 43, pages 317–328, 1999.

[68] M. Thomas, D. Thorne, H. Sing, D. Redmond, T. Balkin, N. Wesensten, J. Russo, M. Welsh, L. Rowland, D. Johnson, R. Aladdin, R. Cephus, S. Hall, and G. Belenky. The relationship between driving accidents and microsleeps during cumulative partial sleep deprivation. *Journal of Sleep Research*, 7:275, 1998.

[69] V. Valley and R. Broughton. The physiological (EEG) nature of drowsiness and its relation to performance deficits in narcoleptics. *Electroencephalography and Clinical Neurophysiology*, 55:243–251, 1983.

[70] M. W. Johns. Assessing the drowsiness of drivers. Technical report, 2001.

[71] A. Muzet, T. Pébayle, J. Langrognet, and S. Otmani. AWAKE pilot study no.2: Testing steering grip sensor measures. Technical report, 2003.

[72] R. Schleicher, N. Galley, S. Briest, and L. Galley. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982–1010, 2008.

[73] W. W. Wierwille, L. A. Ellsworth, S. S. Wreggit, R. J. Fairbanks, and C. L. Kirn. Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. Technical report, 1994.

[74] W. W. Wierwille. Historical perspective on slow eyelid closure: whence PERCLOS? In *Technical Proceedings of Ocular Measures of Driver Alertness Conference, Herndon, VA (FHWA Technical Report No. MC-99-136), Federal Highway Administration, Office of Motor Carrier and Highway Safety, Washington DC*, pages 31–53, 1999.

[75] D.F. Dinges, M.M. Mallis, G. Maislin, and J.W. Powell. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Technical report, 1998.

[76] D.F. Dinges and R. Grace. PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Technical report, 1998.

[77] R. Grace, V.E. Byrne, D.M. Bierman, J.-M. Legrand, D. Gricourt, B.K. Davis, J.J. Staszewski, and B. Carnahan. A drowsy driver detection system for heavy vehicles. In *17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings*, pages I36/1–I36/8, 1998.

[78] Q. Ji. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, 2002.

[79] J. A. Stern, D. Boyer, and D. Schroeder. Blink rate: a possible measure of fatigue. *Human Factors*, 36(2):285–297, 1994.

[80] P.P. Caffier, U. Erdmann, and P. Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3-4):319–25, 2003.

[81] U. Svensson. *Blink behaviour based drowsiness detection - method development and validation*. PhD thesis, 2004.

[82] N. Galley, R. Schleicher, and L. Galley. Blink parameter as indicators of driver's sleepiness - possibilities and limitations. *Vision in vehicles*, pages 1–8, 2004.

[83] M. W. Johns. Alertness monitor, 2003.

[84] M. W. Johns. Measuring alertness, 2006.

[85] M. W. Johns. Incapacity monitor, 2007.

[86] M. W. Johns. Alertness sensing spectacles, 2007.

[87] V. E. Wilkinson, M. L. Jackson, J. Westlake, B. Stevens, M. Barnes, P. Swann, S. M. W. Rajaratnam, and M. E. Howard. The accuracy of eyelid movement parameters for drowsiness detection. *Journal of Clinical Sleep Medicine*, 9(12):1315–1324, 2013.

[88] N. Galley. Saccadic eye movement velocity as an indicator of (de) activation: A review and some speculations. *Journal of Psychophysiology*, 3:229–244, 1989.

[89] N. Galley. An enquiry into the relationship between activation and performance using saccadic eye movement parameters. *Ergonomics*, 41(5):698–720, 1998.

[90] J. Nishiyama, K. Tanida, M. Kusumi, and Y. Hirata. The pupil as a possible premonitor of drowsiness. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1586–1589, 2007.

[91] O. Lowenstein, R. Feinberg, and I. E. Loewenfeld. Pupillary movements during acute and chronic fatigue - A new test for the objective evaluation of tiredness. *Investigative Ophthalmology & Visual Science*, 2(1):138–158, 1963.

[92] S. Abtahi, B. Hariri, and S. Shirmohammadi. Driver drowsiness monitoring based on yawning detection. In *IEEE Conference on Instrumentation and Measurement Technology (I2MTC)*, 2011.

[93] S. Park, F. Pan, S. Kang, and C. D. Yoo. Driver drowsiness detection system based on feature representation learning using various deep networks. In *Computer Vision - ACCV 2016 Workshops*, pages 154–164, 2017.

[94] R. O. Mbouna, S. G. Kong, and M.-G. Chun. Visual analysis of eye state and head pose for driver alertness monitoring. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1462–1469, 2013.

[95] T. Nakamura, A. Maejima, and S. Morishima. Driver drowsiness estimation from facial expression features. In *International Conference on Computer Vision Theory and Applications*, 2014.

[96] H. W. Ko, W. G. Lee, and G. K. Lee. Arousal monitoring system using the change of skin impedance. In *KOSOMBE Spring Conference*, 1995.

[97] M. M. Bundele and R. Banerjee. Detection of fatigue of vehicular driver using skin conductance and oximetry pulse. In *11th International Conference on Information Integration and Web-based Applications and Services*, pages 739 – 744, 2009.

[98] M. Patel, S. K. L Lal, D. Kavanagh, and P. Rossiter. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications*, 38(6):7235–7242, 2011.

[99] J. Vicente, P. Laguna, A. Bartra, and R. Bailón. Drowsiness detection using heart rate variability. *Medical and Biological Engineering and Computing*, 54(6):927–937, 2016.

[100] D.F. Dinges and J.W. Powell. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, 17(6):652–655, 1985.

[101] M. Basner and D.F. Dinges. Maximizing sensitivity of the Psychomotor Vigilance Test (PVT) to sleep loss. *Sleep*, 34(5):581–591, 2011.

[102] J.D.F. Dinges, F. Pack, K. Williams, K.A. Gillen, J.W. Powell, G.E. Ott, C. Aptowicz, and J.A.I. Pack. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, 20(4):267–277, 1997.

[103] M. W. Johns, R. Chapman, K. Crowley, and A. Tucker. A new method for assessing the risks of drowsiness while driving. *Somnologie*, 12(1):66–74, 2008.

[104] C. C. Liu, S. G. Hosking, and M. G. Lenné. Predicting driver drowsiness using vehicle measures: recent insights and future challenges. *Journal of Safety Research*, 40(4):239–45, 2009.

[105] J. Verster and T. Roth. Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). *International Journal of General Medicine*, 4:359 – 371, 2011.

[106] F. Friedrichs and B. Yang. Drowsiness monitoring by steering and lane data based features under real driving conditions. In *18th European Signal Processing Conference (EUSIPCO)*, pages 209–213, 2010.

[107] P. M. Forsman, B. J. Vila, R. A. Short, C. G. Mott, and H. P. A. Van Dongen. Efficient driver drowsiness detection at moderate levels of drowsiness. *Accident Analysis and Prevention*, 50:341–350, 2013.

[108] H. P. A. Van Dongen and P. Forsman. Alertness monitoring systems and associated methods, 2014.

[109] C. Anderson, A.M. Chang, J.P. Sullivan, J.M. Ronda, and C.A. Czeisler. Assessment of drowsiness based on ocular parameters detected by infrared reflectance oculography. *Journal of Clinical Sleep Medicine*, 9(9):907–920, 2013.

[110] D. Hallvig, A. Anund, C. Fors, G. Kecklund, J. G. Karlsson, M. Wahde, and T. Åkerstedt. Sleepy driving on the real road and in the simulator - A comparison. *Accident Analysis and Prevention*, 50:44–50, 2013.

[111] P. Thiffault and J. Bergeron. Monotony of road environment and driver fatigue: A simulator study. *Accident Analysis and Prevention*, 35(3):381–391, 2003.

[112] T. D. Marcotte, E. Roberts, T. J. Rosenthal, and R. K. Heaton. Test-retest reliability of standard deviation of lane position as assessed on a PC-based driving simulator. In *Driving Assessment Conference*, 2003.

[113] T. Hoyoux. *Computer vision systems for automatic analysis of face and eye images in specific applications of interpretation of facial expressions*. PhD thesis, University of Liège, Belgium, 2017.

[114] T. Hayami, K. Matsunaga, K. Shidoji, and Y. Matsuki. Detecting drowsiness while driving by measuring eye movement - A pilot study. In *IEEE Conference on Intelligent Transportation Systems, ITSC*, pages 156–161, 2002.

[115] P. R. Tabrizi and R. A. Zoroofi. Open/closed eye analysis for drowsiness detection. In *1st International Workshops on Image Processing Theory, Tools, and Applications, IPTA*, 2008.

[116] B. Bhowmick and K. S. Chidanand Kumar. Detection and classification of eye state in IR camera for driver drowsiness identification. In *IEEE International Conference on Signal and Image Processing Applications*, pages 340–345, 2009.

[117] A. Picot, S. Charbonnier, and A. Caplier. Drowsiness detection based on visual signs: blinking analysis based on high frame rate video. In *IEEE Conference on Instrumentation and Measurement Technology*, pages 801–804. Ieee, 2010.

[118] D. Liu, P. Sun, Y. Xiao, and Y. Yin. Drowsiness detection based on eyelid movement. In *Second International Workshop on Education Technology and Computer Science*, pages 49–52. Ieee, 2010.

[119] A. Rahman, M. Sirshar, and A. Khan. Real time drowsiness detection using eye blink monitoring. In *National Software Engineering Conference (NSEC 2015)*, pages 1–7, 2015.

[120] H. Shuyan and Z. Gangtie. Driver drowsiness detection with eyelid related parameters by Support Vector Machine. *Expert Systems with Applications*, 36(4):7651–7658, 2009.

[121] P. Ebrahim, A. Abdellaoui, W. Stolzmann, and B. Yang. Eyelid-based driver state classification under simulated and real driving conditions. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 3190–3196, 2014.

[122] M. W. Johns, A. Tucker, R. Chapman, K. Crowley, and N. Michael. Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie*, 11(4):234–242, 2007.

[123] A. Anund. *Sleepiness at the wheel*. PhD thesis, Karolinska Institute, 2009.

[124] H. P. A. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26(2):117–26, 2003.

[125] S. M. Doran, H. P. A. Van Dongen, and D. F. Dinges. Sustained attention performance during sleep deprivation: evidence of state instability, 2001.

[126] H. P. A. Van Dongen, K. M. Vitellaro, and D. F. Dinges. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep*, 28(4):479–96, 2005.

[127] J. Taillard, P. Philip, O. Coste, P. Sagaspe, and B. Bioulac. The circadian and homeostatic modulation of sleep pressure during wakefulness differs between morning and evening chronotypes. *Journal of Sleep Research*, 12(4):275–282, 2003.

[128] P. Philip, P. Sagaspe, J. Taillard, C. Valtat, N. Moore, T. Åkerstedt, A. Charles, and B. Bioulac. Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep*, 28(12):1511–1516, 2005.

[129] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa. Validation of the Karolinska Sleepiness Scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7):1574–1581, 2006.

[130] L. Galley, E.-H. Hentschel, K.-P. Kuhn, and W. Stolzmann. Method and control device recognising, specific to a driver, inattentiveness of a driver of a vehicle, 2009.

[131] T. Vöhringer-Kuhnt. Wierwille's method of driver drowsiness evaluation revisited. In *International Conference on Traffic and Transport Psychology*, pages 5–9, 2004.

[132] G. Renner and S. Mehring. Lane departure and drowsiness - Two major accident causes - One safety system. In *Mobility for Everyone, 4th World Congress on Intelligent Transport Systems*, 1997.

[133] D. Sandberg, A. Anund, C. Fors, G. Kecklund, J. G. Karlsson, M. Wahde, and T. Åkerstedt. The Characteristics of Sleepiness During Real Driving at Night - A Study of Driving Performance, Physiology and Subjective Experience. *Sleep*, 34(10):1317–1325, 2011.

[134] W. Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–95, 1999.

[135] T.-P. Jung, S. Makeig, M. Stensmo, and T. J. Sejnowski. Estimating alertness from the EEG power spectrum. *IEEE Transactions on Biomedical Engineering*, 44(1):60–69, 1997.

[136] S. K. L. Lal, A. Craig, P. Boord, L. Kirkup, and H. Nguyen. Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research*, 34(3):321–328, 2003.

[137] S. F. Liang, C. T. Lin, R. C. Wu, Y. C. Chen, T. Y. Huang, and T. P. Jung. Monitoring driver's alertness based on the driving performance estimation and the EEG power spectrum analysis. In *27th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, volume 6, pages 5738–5741, 2005.

[138] B. T. Jap, S. Lal, and P. Fischer. Comparing combinations of EEG activity in train drivers during monotonous driving. *Expert Systems with Applications*, 38(1):996–1003, 2011.

[139] A. G. Correa and E. L. Leber. An automatic detector of drowsiness based on spectral analysis and wavelet decomposition of EEG records. In *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1405–1408, 2010.

[140] A. G. Correa, L. Orosco, and E. Laciar. Automatic detection of drowsiness in EEG records based on time analysis. *Medical Engineering and Physics*, 36:244–249, 2014.

[141] S. Yu, P. Li, H. Lin, E. Rohani, G. Choi, B. Shao, and Q. Wang. Support Vector Machine based detection of drowsiness using minimum EEG features. In *International Conference on Social Computing*, pages 827–835, 2013.

[142] S. Kar, M. Bhagat, and A. Routray. EEG signal analysis for the assessment and quantification of driver's fatigue. *Transportation Research Part F: Traffic Psychology and Behaviour*, 13(5):297–306, 2010.

[143] A. Picot, S. Charbonnier, and A. Caplier. On-line detection of drowsiness using brain and visual information. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 42(3):764–775, 2012.

[144] J.-C. Cexus. *Analyse des signaux non-stationnaires par transformation de Huang, Opérateur de Teager-Kaiser, et Transformation de Huang-Teager (THT)*. PhD thesis, 2006.

[145] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Procedures of Royal Society of London1*, 454:903–995, 1998.

[146] H. Sharabaty. *Diagnostic de la somnolence d'un opérateur: analyse automatique de signaux physiologiques*. PhD thesis, Université de Toulouse, 2007.

[147] M. Feldman. Time-varying vibration decomposition and analysis based on the Hilbert transform. *Journal of Sound and Vibration*, 295(3-5):518–530, 2006.

[148] M. Feldman. Theoretical analysis and comparison of the Hilbert transform decomposition methods. *Mechanical Systems and Signal Processing*, 22(3):509–519, 2008.

[149] M. Feldman. Hilbert transform in vibration analysis. *Mechanical Systems and Signal Processing*, 25(3):735–802, 2011.

[150] B. Wang, C. M. Wong, F. Wan, P. U. Mak, P. I. Mak, and M. I. Vai. Comparison of different classification methods for EEG-based brain computer interfaces: A case study. In *IEEE International Conference on Information and Automation (ICIA)*, pages 1416–1421, 2009.

[151] J. Herscovitch and R. Broughton. Sensitivity of the Stanford Sleepiness Scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep*, 4(1):83–92, 1981.