Evaluation of the performance of an experimental somnolence quantification system in terms of reaction times and lapses

Clémentine François, Jérôme Wertz, Murielle Kirkove, and Jacques G. Verly, *IEEE Senior Member*

Abstract— **Somnolence is known to be a major cause of various types of accidents, and ocular parameters are recognized to be reliable physiological indicators of somnolence. We have thus developed an experimental somnolence quantification system that uses images of the eye and that produces a level of somnolence on a continuous numerical scale. The aim of this paper is to show that the level of somnolence produced by our system is well related to the level of performance of subjects accomplishing three reaction-time tests in different sleep conditions. Twenty seven subjects participated in the study and images of their right eye were continuously recorded during the tests. Levels of somnolence, reaction times (RTs), and percentages of lapses were computed for each minute of test. Results show that the values of these three parameters increase significantly with sleep deprivation. We determined the best threshold on our scale of somnolence to predict lapses, and we also shown that correlations exist with some of the ocular parameters. Our somnolence quantification system has thus significant potential to predict performance decrements of subjects accomplishing a task.**

I. Introduction

Somnolence (or, synonymously, drowsiness) is known to be a major cause of various types of accidents, and especially on roads. It would be responsible for 20 to 30% of road accidents in general [1], and for one third of fatal accidents on highways in France [2]. The causes of somnolence are diverse, e.g. sleep deprivation, sleep disorders, alcohol, some medications, or performing a monotonous task. Moreover, from a medical point of view, some people have a higher propensity than normal to somnolence. Indeed, 6 to 11% of the population suffers from severe chronic excessive daytime sleepiness [3]. Somnolence is thus a major problem of public health and safety.

Somnolence is different from fatigue; it is the intermediate state between wakefulness and sleep, and it is characterized by a loss of vigilance [4]. There are various traditional ways to assess somnolence; e.g., we can consider the dichotomy of objective methods vs. subjective methods [5]. Subjective methods are mainly based on questionnaires, and objective methods are mainly based either on performance measures or on physiological measures. Since somnolence is a physiological state, it seems particularly meaningful to use physiological methods to characterize it. Among these, the most significant ones rely on polysomnography and/or oculography. Polysomnography [6] is viewed by some practioners as the reference in the domain, but it is very sensitive to artifacts, and it is not very practical for everyday use. Ocular parameters are recognized to be good and reliable physiological indicators of somnolence [7,8], so that oculography seems to be the most sensible way to characterize somnolence in practice.

We have thus developed an experimental somnolence monitoring system (software/algorithms) based on the physiological state of a subject. This system uses ocular parameters extracted from images of the eye to determine a level of somnolence on a continuous numerical. The ultimate goal of this system is to prevent somnolence-related accidents.

The aim of the study described in this paper is to verify that, for a number of subjects, the level of somnolence produced by our system is well "correlated" with the level of performance of these subjects in the accomplishment of a task.

II. Material and Method

A. Data acquisition

Twenty seven (27) healthy volunteers (12 M, 15 F, mean age 24.3, range 19-32 years) participated in the experiment, which included performing three visual reaction-time (RT) tests – each of 15 minutes duration – in different sleep conditions over two days.

Following a normal night sleep, each subject takes the first RT test in our laboratory between 8:00 and 10:00 am on Day 1. The subject then wears an actimeter, but is otherwise free to carry out his normal activities (except sleeping), this until 11:00 pm, when the subject returns to our laboratory. During this night, the subject is not allowed to sleep (and he no longer wears the actimeter). On Day 2, the subject takes the second RT test between 2:00 and 4:00 am, and, after breakfast, the third RT test between 11:00 am and 1:00 pm (after at least 28h of sleep deprivation). The subject is asked not to take any stimulant (coffee, tea…) from 6:00 pm on Day 1 until the end of the last RT test.

This protocol was approved by the Ethics Committee of the University of Liège.

Research supported by Région Wallonne.

C. François is with the INTELSIG Laboratory, Department of Electrical Engineering and Computer Science, University of Liège, Belgium (e-mail: cfrancois@ulg.ac.be).

J. Wertz is with the INTELSIG Laboratory, Department of Electrical Engineering and Computer Science, University of Liège, Belgium (e-mail: j.wertz@ulg.ac.be).

M. Kirkove is with the INTELSIG Laboratory, Department of Electrical Engineering and Computer Science, University of Liège, Belgium (e-mail: m.kirkove@ulg.ac.be).

J. G. Verly is with the INTELSIG Laboratory, Department of Electrical Engineering and Computer Science, University of Liège, Belgium (e-mail: jacques.verly@ulg.ac.be).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

The conditions before the night without sleep (RT test 1), are referred to as "not-sleep-deprived", and those during and after this night (RT tests 2 and 3) as "moderately-sleep-deprived" and "sleep-deprived", respectively.

During each RT test, we recorded (1) images of the right eye of the subject using a portable system with an infrared camera placed near the eye and (2) data coming from the RT test. We also recorded polysomnographic (PSG) data, but these data are not used here.

B. *Somnolence quantification*

It is well-known that several ocular parameters (OPs) are indicative of the level of alertness/somnolence of a person [7,8]. Most of these 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 eyeballs (including saccades).

In this paper, we only consider OPs that are related to the eyelids. Furthermore, we assume that, at some intermediate point of processing, one has determined the positions of the upper and lower eyelids in each image of the eye. 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 of one (1) minute. Each OP produces a numerical value per 1-minute window. Here are some examples of OPs that we extract from images of the eye:

- the mean of the durations of blinks;
- the PERCLOS 70 (which is the proportion of time in a 1-minute window that the eye is at least 70% closed);
- the mean of the velocities for eyelids closings;
- the mean of the amplitudes of blinks.

We thus obtain a set of several OPs for each 1-minute window of test. We then use our somnolence quantification system to produce a level of somnolence for each 1-minute window based on each set of OPs. The level of somnolence that is produced by our system is a numerical value between 0 (well awake) and 10 (very somnolent). In this paper, we refer to our scale of somnolence as the "experimental Somnolence Scale (xSS)". A level of somnolence on this scale is then abbreviated as "xSS score".

C. *Performance quantification*

The visual reaction-time (RT) test that we use is our own implementation of a Psychomotor Vigilance Test (PVT). A stimulus is presented to the subject on a screen during 400 ms and is then repeated randomly throughout the test. The subject is asked to press a button on a device as quickly as possible after he detects the start of each stimulus. The duration of the test is 15 minutes.

The data recorded for each stimulus are the time (t_s) when the stimulus starts and, if applicable, the time(s) (t_r) when the subject reacts (i.e. presses the button). When there is a response, the RT is defined as RT = t_r-t_s, where t_r corresponds to the first response after the start of the stimulus (and before the next stimulus). When there is no response, the RT is marked with "nil" and we talk about a "no response" or an error of omission. If the subject responds with a delay of at least 2000 ms from the start of the stimulus, we say that the response is "significantly delayed", and that it is an error. We use the term "lapse" to refer jointly to a "significantly delayed response" and to a "no response". In Section III, when we talk about reaction times (RTs), we only consider responses within 2000 ms from the start of the stimulus.

D. *Statistical methods*

Several one-way ANOVA analyses were performed on the reaction-time data and on the levels of somnolence determined by our system to look for the presence of significant differences between subjects and between the three sleep conditions (not-sleep-deprived, moderately-sleep-deprived, and sleep-deprived). Statistical significance was accepted for a p-value less than 0.01.

Sensitivities and specificities were also computed and we produced a Receiver Operating Characteristic (ROC) curve by thresholding the level of somnolence determined by our system in order to predict lapses.

III. RESULTS

A. *Effect of sleep deprivation on reaction times*

Several one-way ANOVA analyses were performed on reaction time (RT) data to distinguish differences between subjects and sleep conditions. These analyses lead to the following results. There are significant differences of RTs between subjects, regardless of their sleep condition ($F(27,240)=33.22$, $p<0.01$), and between subjects when they are sleep-deprived ($F(27,80)=15.44$, $p<0.01$). There are also significant differences of RTs between the three sleep conditions ($F(3,2160)=140.97$, $p<0.01$). If we compare the mean of all RTs for all subjects, we can indeed observe that it increases from 387.8±101.2 ms in not-sleep-deprived condition (RT test 1) to 461.4±160.5 ms in sleep-deprived condition (RT test 3). Similarly, the total number of lapses for all subjects in not-sleep-deprived condition increases from 10 for 2325 stimuli (0.43%) to 147 for 2321 stimuli (6.33%) in sleep-deprived condition. This shows that there is a real decrease in performance with sleep deprivation.

B. *Effect of sleep deprivation on levels of somnolence*

Several one-way ANOVA analyses were also performed on levels of somnolence (xSS scores) determined by our system to distinguish differences between subjects and sleep conditions. Even though, the study involved 27 subjects, these analyses were performed on the data from 21 of them (because of difficulties in automatically extracting the OPs for the others). They lead to the following results. There are significant differences of xSS scores between subjects regardless of their sleep condition ($F(21,42)=12.94$, $p<0.01$), and between subjects when they are sleep-deprived ($F(21,14)=22.69$, $p<0.01$). There are also significant differences of xSS scores between the three sleep conditions

(F(3,294)=119.36, p<0.01). As we did for the effect of sleep deprivation on RTs, we can also compare the mean of all xSS scores for all 21 subjects across the different sleep conditions. We observe that the mean xSS score for all 21 subjects increases from 1.48±1.25 in not-sleep-deprived condition (RT test 1) to 3.96±2.83 in sleep-deprived condition (RT test 3). We decided to go further in this analysis by splitting the last sleep condition in "sleep-deprived and not lapsing" (representing 231 1-minute windows out of 294) and "sleep-deprived and lapsing" (representing 63 1-minute windows out of 294). The means and standard deviations of the xSS scores for these two other conditions are respectively 3.00±2.10 and 7.47±2.36. This shows that there is a significant increase in the level of somnolence with sleep deprivation, and even more during a decline of performance such as a lapse.

C. *Link between levels of somnolence and outputs of reaction tests*

For a given RT test, we can compute, for each 1-minute window, (1) several values of RTs (one value for each response of a subject within 2000 ms from the start of a stimulus happening in this window), and (2) the value of the xSS score for this window. Since there are 15 butting 1-minute windows per RT test, we obtain 15 xSS scores for each given RT test and several RT values for each xSS scores. We do the same for all RT tests of all subjects. To present the results in a simple, bar-plot way, we define ten successive unit-length intervals (or bins) for the xSS scores, i.e. [0,1[, [1,2[… We can then associate each RT (value) to a unique xSS bin, and compute the mean RT for each bin. Figure 1 shows the mean RT and its standard deviation for each bin along the xSS score axis. The number above each bin represents the number of responses falling into the bin.

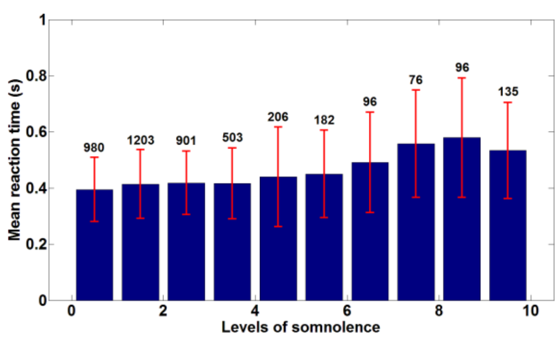

Figure 1. Mean reaction time (RT) as a function of xSS scores.

Figure 2 is similar to Figure 1 but it shows the mean percentage of lapses for each xSS bins (instead of the mean RT). For each 1-minute window, we compute one value representing the percentage of lapses. We can then associate each value of percentage of lapses to a unique xSS bin and compute the mean percentage of lapses for each bin. The number above each bin corresponds to the number of minutes falling into the bin.

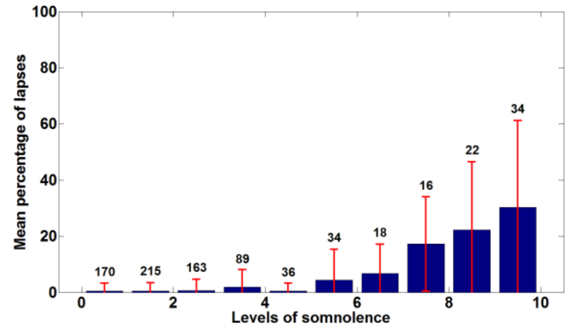

Figure 2. Mean percentage of lapses as a function of xSS scores.

D. *Our scale of somnolence as predictor of lapses*

To find the best threshold to predict lapses with our somnolence scale, we used integer thresholds from 0 to 10 on the xSS scores, and we computed values of sensitivity and specificity based on confusion tables. These values were calculated from data from all 21 subjects in all sleep conditions. We then use the common Receiver Operating Characteristic (ROC) curve for representing these quantities. The resulting curve is shown in Figure 3.

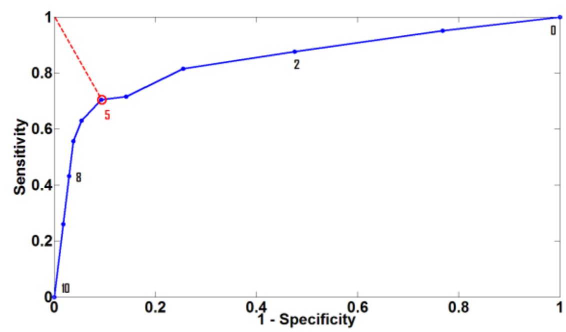

Figure 3. ROC curve with "at least one lapse" as parameter.

"Sensitivity" is also called "true positive rate" or "probability of correct detection". "1-specificity" is also called "false positive rate" or "probability of false detection". One can then easily understand that the more a point on the curve approaches the upper left corner of the graph, the more it has a high probability of correct detection and the more it has a low probability of false detection. A threshold of 5 on our scale of somnolence seems to be the best to predict lapses, and this threshold corresponds to a sensitivity of 70.37% and to a specificity of 90.66%.

E. *Relations with ocular parameters*

For a given 1-minute window, we can compute a value for each ocular parameter (OP), an xSS score, and the mean RT for this window. We can thus analyze the relations between values of OPs and both xSS scores and values of mean RT. In this paper, we decided to examine two well-known OPs: mean of the durations of blinks and PERCLOS 70. We thus associated each value of each OP with the corresponding xSS score and the corresponding mean RT value, and we represent the results on several scatter plots in Figure 4. In this figure, the two plots at the top show the mean of the durations of blinks (or mean blink duration) in relation with, on the left, xSS score, and, on the right, mean RT. Similarly, the two plots at the bottom of the figure show PERCLOS 70 in relation with, on the left, xSS score and, on the right, mean RT. The red line in each plot represents the straight line of best fit, i.e. the straight line that is the best approximation of the data. The equation of this line is determined by the least-square method.

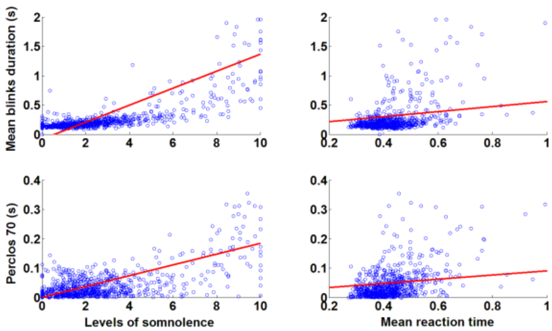

Figure 4. Relations between two OPs and (1) xSS scores and (2) mean RTs.

IV. Discussion

On average for all subjects, the reaction times and the percentages of lapses increase significantly across the three successive tests. Similarly, the levels of somnolence determined by our system also increase significantly with sleep deprivation on average for all subjects. However, one can notice that the mean level of somnolence in the third condition (sleep-deprived) remains quite low compared to the upper bound of our scale. Nevertheless, if one looks at the mean level of somnolence 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 somnolence quantification system reflects well performance decrements like lapses.

From Figure 1, one can observe that the mean reaction time increases with the level of somnolence determined by our system except for the 10^{th} bin. This exception can be explained by the fact that we only consider reaction times within 2000 ms after the start of each stimulus. Moreover, Figure 2 also shows that the mean percentage of lapses increases with the levels of somnolence, and that its highest value corresponds to the 10^{th} bin.

The Receiver Operating Characteristics (ROC) curve presented in Figure 3 suggests that a threshold of 5 on our scale of somnolence (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 present dataset. 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 somnolence-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.

Figure 4 illustrates the relations between two well-known ocular parameters (OPs) (mean blink duration and PERCLOS 70) with (1) xSS scores and (2) mean RTs. From the different graphs in this figure, we can conclude that both OPs are well correlated with xSS scores, but slightly less correlated with mean RTs. Concerning this last conclusion, one should take into account the fact that we do not consider lapses here and that high values of OPs are probably more related to lapses. Moreover, with these results, one can also highlight the fact that a single OP is not sufficient to predict performance decrements, and that one needs to look at a combination of OPs. And this is exactly what our system does.

V. Conclusion

The above experiments indicate that the level of somnolence determined by our system based on images of the eye is well “correlated” with the level of performance of a subject accomplishing a task. We have indeed shown that, in the case of a reaction-time task, mean reaction times and percentages of lapses increased with levels of somnolence and with sleep deprivation. We have also demonstrated that a threshold of 5 on our scale of somnolence (from 0 to 10) is the best for predicting lapses. Our somnolence quantification system has thus significant potential for predicting performance decrements due to somnolence and, ultimately, for preventing somnolence-related accidents.

References

[1] Klauer et al., “The impact of driver inattention on nearcrash/crash risk: An analysis using the 100-Car Naturalistic Driving Study data,” *NHTSA*, 2006.

[2] Association des Sociétés Françaises d’Autoroutes, “Somnolence au volant – une étude pour mieux comprendre,” Juin 2010.

[3] M.F. Vecchierini, et al., “La somnolence diurne excessive et les hypersomnies centrales primaires: données épidémiologiques,” *Médecine du sommeil,* 2010.

[4] M.W. Johns, “A new perspective on sleepiness,” *Sleep and Biological Rhythms*, 2010, vol. 8, pp. 170-179.

[5] T. Åkerstedt, and M. Gillberg, “Subjective and objective sleepiness in the active individual,” *Intern. J. Neuroscience,* 1990, Vol. 52, pp. 29-37.

[6] A. Anund, G. Kecklund, B. Peters, A. Forsman, A. Lowden, and T. Åkerstedt, “Driver impairment at night and its relation to physiological sleepiness,” *Scand. J. Work Environ. Health*, 2008, vol. 34(2), pp. 142-150.

[7] D.F. Dinges, and R. Grace, “PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance,” *US Dept. Transportation, Federal Highway Admin., Washington, DC, Tech. Rep. Publication No. FHWA-MCRT-98-006,* 1998.

[8] R. Schleicher, N. Galley, S. Briest, and L. Galley, “Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?” *Ergonomics*, Jul. 2008, vol. 51, pp. 982-1010.