Monitoring drowsiness on-line using a single encephalographic channel

Antoine Picot, Sylvie Charbonnier, Alice Caplier

Chapter Number

Monitoring drowsiness on-line using a single encephalographic channel

Antoine Picot, Sylvie Charbonnier & Alice Caplier
Gipsa-Lab, Grenoble University
France

1. Introduction

Drowsiness is the transition state between awakening and sleep during which a decrease of vigilance is generally observed. This can be a serious problem for tasks that need a sustained attention, such as driving. According to a report of the American National Highway Safety Traffic Administration (Royal, 2002) driver drowsiness is annually responsible for about 56,000 crashes which is the reason why more and more researches have been developed to build automatic detectors of this dangerous state.
Both behavioural and physiological modifications occur during drowsiness. Reaction time is slower, vigilance is reduced and information processing is less efficient, which can generate abnormal driving. Moreover, as drowsiness is the transition between awakening and sleep, it induces an increase of the number and the duration of blinks and yawns. Changes in cerebral activity also happen and can be observed thanks to electroencephalography.
Researches on driver state monitoring has begun about thirty years ago and are still very active. The driver state monitoring systems can be classified into three kinds of system: those focusing on the vehicle behaviour, those focusing on the driver physical behaviour and those focusing on the driver physiological behaviour.
The first systems developed were the ones using sensors monitoring the vehicle behaviour (O'Hanlon & Kelley, 1974, Klein et al., 1980). The main features studied are the steering wheel movements, the lateral position of the car on the road, the standard deviation of lateral position (SDLP) and the time to line crossing (TLC). The purpose is to detect an abnormal behaviour of the car, due to the driver drowsiness. The problems encountered by this kind of methods are that the features used depend on the shape of the road and how one drives, which may change a lot from one driver to another (Renner & Mehring, 1997).
To overcome these problems, researches have focused on systems using sensors monitoring drivers' awareness. One widespread technique to monitor the driver state is the use of a video camera. Indeed, a lot of information can be extracted from the driver face to monitor fatigue such as gaze, frequency and duration of eye blinking and yawning or percentage of eyelid closure. A lot of examples using camera to monitor the driver state can be found in the literature (Grace et al., 1998; Ji & Yang, 2001; Vural et al., 2007).
These kinds of systems focus on the drivers' visual attention. Face, mouth and eye tracking algorithms are used to detect the face. Once the face, the eyes and the mouth are located, it is

easy to detect eye blinking or yawning and calculate their frequency and duration. Frequency and duration of yawning or eye blinking too high indicate a decrease of attention. The gaze can be calculated with the eyes and the face position (Smith et al., 2003) or using a stereoscopic camera (Ji et al., 2004). Then, it allows the driver to be warned when he is not looking at the road.

However, many differences can be observed between drivers, which makes it hard to monitor fatigue with only one feature (Karrer et al, 2004). An interesting way of merging the different features (eye blinking, yawning, gaze...) is used by Ji et al. (Ji et al., 2006). They use probabilistic networks which allow all features to contribute to the decision of the level of attention. Moreover, external factors (weather, hour of the day, etc...) can contribute in these networks to determine the level of attention.

However, video features are not the best indicators of drowsiness. According to Dinges (Dinges, 1995), the best indicators of fatigue are the physiological indicators. The electroencephalogram (EEG) and the electro-oculogram (EOG) are mainly used to study drowsiness. Yet, several researches have focused on other physiological indicators such as the electrocardiogram (ECG) to monitor drivers' heart rate (Törnros at al., 2000) or the drivers' temperature (Quanten et al., 2006).

The EOG is the measurement of the resting potential of the retina. It gives an accurate measurement of eyes movements. Many features can be extracted from this information such as eyelid opening and closing parameters, blinks frequency, blinks amplitude, blinks duration... According to Galley et al. (Galley et al., 2004), EOG is a relevant measure to monitor fatigue since some extracted features are really sensitive to drowsiness. One of the most efficient features extracted is the PERCLOS (PERcentage of eyelid CLOSure). This feature has been defined by Wierwille (Wierwille et al., 1994). It is the percentage of eyelid closure over the time. Knippling (Knippling, 1998) showed that PERCLOS is a good indicator of drowsiness that increases with fatigue.

Electroencephalography measures the electrical activity of the brain from electrodes placed over the scalp. Drowsiness appears into the EEG spectrum by an increase of activity in the frequency bands [8-12]Hz (alpha band) and [4-8]Hz (theta band) predominantly in the parietal and central regions of the brain. In the same time, a decrease of activity in the band [12-26]Hz (beta activity) can also be observed, as beta activity increases with cognitive tasks and active concentration. This has been shown in several studies (Santamaria & Chiapa, 1987; Akerstedt & Gillberg, 1990; Kay et al., 1994). EEG is so efficient in detecting drowsiness that it is often used as a reference indicator. In this case, the reference is built by expert doctors who visually observe the proportion of alpha and theta activity on a short-time window as in (Muzet et al., 2003). The analysis is done off-line and is time consuming.

In order to make the analysis of drowsiness in an automatic way, the EEG power spectrum can be computed using Fast Fourier Transform or using wavelets Transform but none of these techniques seems better than the other. The number of EEG channels used to monitor drowsiness fluctuates from a few to about thirty. The advantage of using a large number of EEG channels is to obtain spatial information on how the EEG energy is shifting from one frequency band to another (Makeig et al., 1996; Lin et al., 2005b). However, using only a few EEG channels is faster and easier to compute.

The features obtained from EEG can be used in many ways. Several studies proposed to monitor some ratios between different EEG power bands. De Waard and Brookhuis (De Waard & Brookhuis, 1991) suggest to monitor the ratio between the alpha activity and the

theta activity on the beta activity [(alpha+theta)/beta] (beta band: [12-26]Hz) which increases with driver drowsiness. Drowsiness decision can be made using EEG features used as inputs to a classifier. Ben Khalifa et al. (Ben Khalifa et al., 2004) used a connexionist method to detect drowsiness. They proposed to use a single parieto-occipital channel (P4-O2) to avoid ocular frontal artefacts. Relative power in the [1- 23]Hz band is computed using a short term fast Fourier transformation and divided in twenty three sub-bands of 1Hz. These features are then used as entries in a linear vector quantization (LVQ) neural network to detect artefacted states and in a self organizing map (SOM) neural network to classify states into two states: awake and drowsy. Lin et al. (Lin et al., 2005a) applied Principal Component Analysis (PCA) on the whole spectrum of a 2-channels EEG. A fast Fourier transform is then computed on the principal component to extract the power in 10 EEG bands from 1 to 40Hz. These features are used for training a linear regression model which estimates the driving performance, i.e. the deviation between the car and the centre of the road, assuming that driving performances are correlated with driver's alertness. They also used the two principal components obtained using an independent component analysis (ICA) on a 33-channels EEG instead using a PCA on a 2-channels EEG with their linear regression model to improve their estimation of driving performance. Most recently, Rosipal et al. (Rosipal et al., 2007) used hidden Gaussian mixture model to monitor drowsiness. A hierarchical Gaussian mixture model (hGMM) with two mixture components at the lower hierarchical level is used. Each mixture models the data density distribution of one of the two drowsiness cornerstones/classes represented by 4-second long EEG segments with low and high drowsiness levels. The spectral content of each EEG segment is transferred into a compact form of autoregressive model coefficients. Their study is performed on a large number of drivers.

However, even if some on-line techniques exist, they seem hard to implement: they either need to be trained on numerous data previously recorded on a driver or they require a large number of EEG channels. Yet, the fact that EEG is used as a drowsiness reference by expert doctors in physiological studies proves the utility of an on-line driver drowsiness detection system using EEG.

In this paper, we propose to develop a system to detect driver drowsiness that uses as few EEG channels as possible that works on-line and that does not need to be trained. Indeed, it seems that a relatively simple drowsiness detection system for drivers would be easier to implement in a car.

The method proposed is described in section 2. The results obtained on a significant database of 40 EEG recordings from 20 drowsy drivers are shown and discussed in section 3. Finally, conclusions and perspectives are presented.

2. Drowsiness detection method

After a short presentation on EEG and how drowsiness appears in EEG, the drowsiness detection method is presented step by step.

2.1 EEG and drowsiness

An electroencephalogram is a measurement of the electrical activity of the brain from electrodes placed on the scalp (Blinowska & Durka, 2006), according to the international 10-20 system shown in fig. 1.

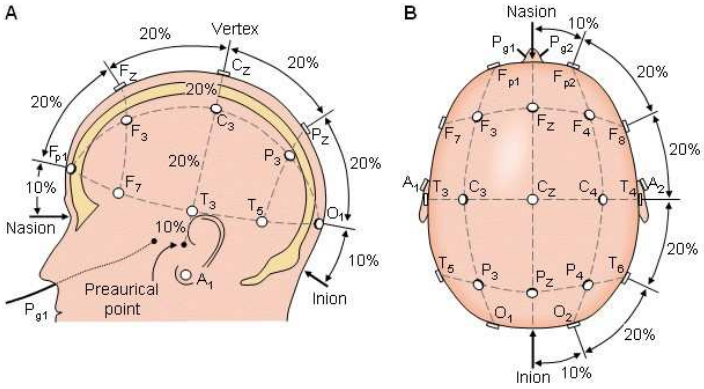

Fig. 1. 10-20 system

EEG is described in term of rhythmic activity and transients. The rhythmic activity is divided into frequency bands: delta (δ) activity ([0.5-4]Hz), theta (θ) activity ([4-8]Hz), alpha (α) activity ([8-12]Hz), beta (β) activity ([12-26]Hz) and gamma (γ) activity (over 26Hz). Most of the time, only the range [1-30]Hz is used because activity below or above this range is likely to be artefactual (under standard clinical recording techniques).
Drowsiness is characterized by an increase of alpha and theta activities, predominantly in the parietal (P) and central (C) regions of the brain, and a slowdown of blinks and eye movements (Akerstedt & Gillberg, 1990; Kay et al., 1994; Muzet et al., 2003). Even if different scales of drowsiness classification exist, none of them are standardized and there are no standardized rules to differentiate the levels of drowsiness (as the Rechtschaffen and Kales rules (Rechtschaffen & Kales, 1968) for the study of sleep). This may be due to the quite recent interest on drowsiness compared to the sleep analysis and the difficulty to collect drowsiness data.

Objective sleepiness score	α and θ cumulative duration	Blinks and eye movements
0	Negligible	Normal
1	Less than 5s	Normal
2	Less than 5s or Less than 10s	Slow Normal
3	Less than 10s or More than 10s	Slow Normal
4	More than 10s	Slow

Table1. OSS Criteria

There are two kinds of scales: subjective sleepiness scales like the Karolinska Sleepiness Scale (KSS) (Akerstedt & Gillberg, 1990) which allows drivers to directly evaluate their own drowsiness and Objective Sleepiness Scales (OSS) which is used by expert doctors to

evaluate drivers' drowsiness after driving. The OSS used in this study is the five levels scale from 0 (awake) to 4 (very drowsy) developed by Muzet (Muzet et al., 2003). Decisions are made every 20s and depend on the length of alpha and theta bursts as well as on the speed of eyes movements and blinks. The different criteria are presented in table 1. This scale is used by doctors to make decisions.

2.2 Method Principle

The purpose is to design a drowsiness detection algorithm which can work on-line, inspired by the OSS. The overview of the detection method is shown in fig. 2. First the EEG power spectrum is computed using a Short Time Fourier Transform (STFT) to calculate the relative power into the different EEG bands every second. Then, the relative power of the alpha band is median filtered using a sliding window to reject abnormal values. A Means Comparison Test (MCT) is computed at last to compare the energy to a reference level, learnt at the beginning of the recording while the patient is not supposed to be drowsy. MCT is normalized. A common threshold of detection can be proposed taking into account the acceptable level of false alarms and validated using experiments which has been presented in (Picot et al., 2008). Concomitantly, a Variances Comparison Test (VCT) is computed on the raw EEG data to detect high amplitude artefacts. Information on the occurrence of artefacts can be used as an index of reliability on the "drowsy decision".

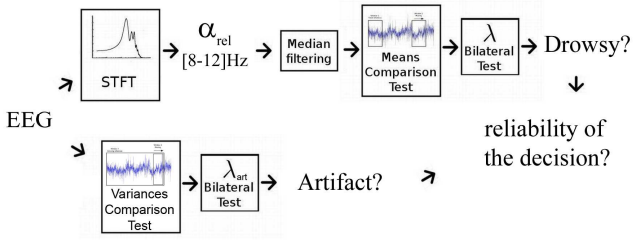

Fig. 2. Drowsiness detection method

2.2.1 EEG Power Spectrum

The EEG power spectrum is computed using a Short Time Fourier Transform (STFT). The power spectrum is computed every second on a window of two seconds using Welch's periodogram method (Welch, 1967). The overlapping window between the previous and next value is 1 second. Then, the relative powers in each band are calculated as the ratio of the power in one band and the power of the whole EEG spectrum. Only the range [1-30]Hz is used because activity below or above this range is likely to be artefactual. For example the α relative power is calculated as follows:

$$\alpha_{\text{relative\_power}} = \frac{\alpha_{\text{power}}}{\text{EEG}_{\text{power}}} \quad (1)$$

The relative power in the bands α ([8-12]Hz) , θ ([4-8]Hz) and β([12-26]Hz) are named α_{rel}, θ_{rel} and β_{rel} respectively.

2.2.2 Median filtering

Median filtering is used to smooth the α_{rel} signal and to reject abnormal values. The median is the value separating the higher half of a population from the lower half. Compared to the mean, the median is known for its robustness towards outliers, as far as the number of outliers is lower than half the length of the population. Here, the median of the relative powers is calculated every second, before performing MCT, using a sliding window of 10s.

2.2.3 Means Comparison Test

The method of MCT is inspired by Ragot (Ragot et al., 1990) and is applied on the relative powers in the alpha band. A moving window is compared to a fixed reference window as shown in fig. 3.

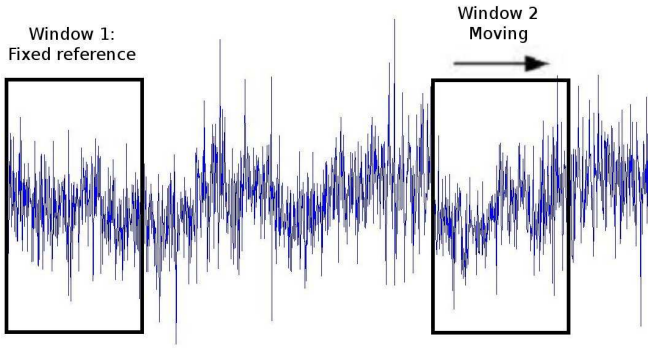

Fig. 3. Illustration of the windows for the MCT

The classical MCT has quite restrictive conditions due to the fact that the theoretical variances are unknown. Let us consider two independent populations of length n_1 and n_2, whose means are $\bar{x}_1$ and $\bar{x}_2$ and whose variances are s_1^2 and s_2^2. Then, the variable:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^{\,2} + n_2 s_2^{\,2}}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{2}$$

follows a $n_1 + n_2 - 2$ liberty degrees Student law. The equality of the two means can be tested by a bilateral test with a confidence threshold λ: $-t_{1-\lambda/2} < t < t_{1-\lambda/2}$.
If the two populations have the same length n (i.e. $n_1=n_2=n$) and their theoretical variances are equal, (2) can be formulated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^{\,2} + s_2^{\,2}}{n}}} \tag{3}$$

So, the variable t follows a n-1 liberty degrees Student law.
Moreover, if the populations are large enough, i.e. if n_1 and n_2 are equal or greater than 20, the test is performed with the variable:

$$u = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (4)$$

which then follows a centred reduced normal law. The means equality is then tested by a bilateral test with a confidence threshold λ: $-u_{1-\lambda/2} < u < u_{1-\lambda/2}$. Furthermore, the theoretical variances equality is no longer necessary.
Here, the test is computed on the relative power signals calculated every second. The reference, calculated from the fixed window, is calculated at the beginning of a recording with the assumption that, before driving, the driver is not drowsy. The mean value calculated during this period provides a reference for the not drowsy stage. The mean values calculated on line are compared to this "not drowsy" reference value. The length of the fixed reference window is n_1=60s and the one of the moving window is n_2=30s, so n_1 and n_2 are greater than 20 as α_{rel} is calculated every second. Hence, the formula (4) can be used. The threshold λ fixes the percentage of false alarms expected. The higher the threshold, the lower the percentage of false alarms. In this study, λ is empirically chosen and discussed in section 3.

2.3 Artefact detection

An artefact is an electric perturbation of the EEG signal due to patient movements or measurement problems. Artefacts pollute the whole EEG band and it is quite impossible to extract reliable EEG information when an artefact occurs. There are several types of artefacts. They may be due to ocular movements, face muscles movements or measurement devices problems such as electrode unstuck.
The proposed solution to detect high-amplitude artefacts corresponding to electrode unstuck is to apply a Variances Comparison Test (VCT) on the EEG. Artefact values are around mV whereas EEG signal values are around µV. High-amplitude artefacts pollute the whole EEG band as shown in the red box on the spectrogram in fig. 4.
The method of VCT is inspired by Ragot (Ragot et al., 1990) and is directly applied on the raw EEG. The principle is the same than the MCT: the variance of a moving window is compared to the variance of a reference window. Let us consider two independent populations with normal distributions. Their lengths are n_1 and n_2 and their experimental variances are s_1^2 and s_2^2. Then, the variable:

$$F = \frac{s_1^2}{s_2^2} \quad (5)$$

follows a Fisher law with $k_1=n_1-1$ and $k_2=n_2-1$ liberty degrees. The variances equality is then tested by a bilateral test with a confidence threshold λ_{art}: $F_{\lambda art/2} < F < F_{1-\lambda art/2}$.

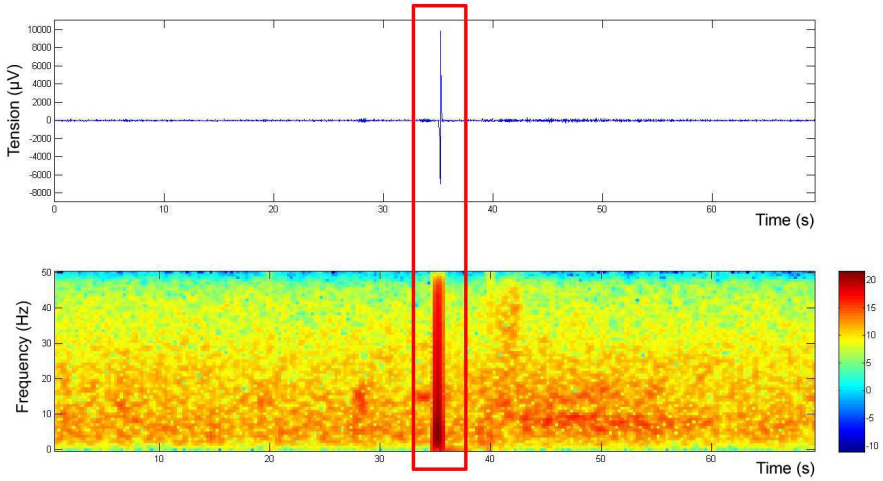

Fig. 4. Example showing a high-amplitude artefact

Here, the test is computed on the raw EEG data every 10s. A moving window is compared to a fixed reference window as shown in fig. 3. The reference variance is calculated on an artefact-free time window lasting one minute, chosen at the beginning of the recording,. Then, a moving window of 10s is compared to this reference every 10s. The threshold λ_{art} is empirically chosen and discussed in section 3.
The goal is to detect the artefact, not to reject it. As the system is working on a minimal number of EEG channels, it is not possible to recover lost EEG information of the artefact since missing information cannot be found somewhere else. Nevertheless, detecting the occurrence of an artefact provides information on the signal quality: whenever an artefact is detected, the concomitant relative power extracted from EEG should not be used to evaluate drowsiness.

2.4 Method relevance

The whole algorithm can be applied on-line. However, the sliding window of 10s used for median filtering induces a delay of 5s and the sliding window of 30s used for the MCT induces a delay of 15s. The artefact detection is computed in parallel with a sliding window of 10s which induces a delay of 5s. So, the decision provided by the algorithm is delayed by 20s from the signals recording. This latency in the decision will be taken into account when comparing the results to the expert's decisions.
The general purpose of this algorithm is the detection of drowsiness. The MCT detects α bursts, which are indicators of drowsiness as seen in section 1. The reference is calculated on a fixed window chosen at the beginning of the signal, supposing that the driver is completely awake when he starts driving. So, the mean calculated on the moving window is compared to a wakefulness reference. If the bilateral test is higher than the threshold, the driver is then considered as drowsy, otherwise he is considered as awake. Fig. 6 shows how

the signal is processed in the detection system. First the relative power in the α band (b) of the EEG (a) is calculated. Then, it is smoothed by median filtering (c). A MCT is performed (d) on the filtered signal and is thresholded to make the decision awake or drowsy (e).

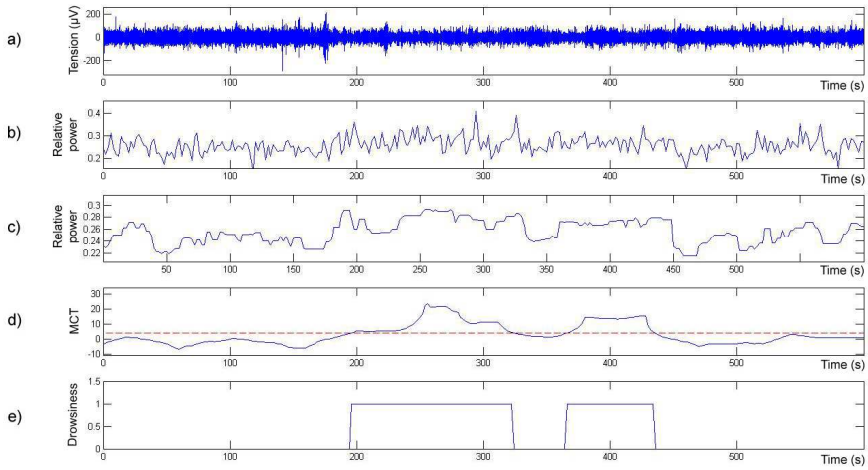

Fig. 6. Signal processing from EEG to drowsiness detection

High-amplitude artefacts pollute the EEG signals and generate isolated high abnormal values on the whole EEG band of the power spectrum. A median filter is used to smooth the α relative power signal and to reject abnormal isolated values to avoid false detection. Moreover, a VCT is calculated on the raw EEG signal to detect the occurrence of high-amplitude artefacts polluting the whole EEG band. The detection of these high-amplitude artefacts does not allow rejecting them but provides information on the quality of signal around this point. It means that if artefacts are found on a part of the signal, decisions on drowsiness in this part tend to be less reliable than if not.

The point with detecting α bursts in EEG signal is the difficulty to define a common threshold for all drivers because of the large inter-individual differences (Karrer et al., 2004). Here, the level of α_{rel} power in the "awake" state is learned on each driver from the reference window. Moreover, the output of MCT is a variable following a centred reduced normal law. So, the threshold used in the bilateral test has statistical meaning and is the same for all drivers.

In the same way, as the output of the VCT is a variable following a Fisher distribution, the threshold used to detect high-amplitude artefacts has a statistical meaning and is the same for all drivers.

3. Results and discussion

3.1 Database

The database used for the evaluation of the method has been provided by the CEPA (Centre d'Etudes de Physiologie Appliquée) from Strasbourg (France) using the driving simulator PAVCAS ("Poste d'Analyse de la Vigilance en Conduite Automobile Simulée"). PAVCAS is a moving base driving simulator composed of a mobile base with four liberty degrees (vertical and longitudinal movements, swaying and pitching) and a real-time interactive visualization unit. The visualisation unit reproduces very well the driving conditions on a freeway by day or night. Images are shown on five screens in front of the vehicle and are arranged in semicircle.
The database is composed of forty recordings from twenty subjects. Each subject was recorded while driving for 90 minutes, a first time perfectly rested and a second time suffering from sleep deprivation (the subject had slept 4 hours only) in diurnal conditions. The database is thus composed of 60 hours of driving data. Each recording includes four EEG channels (left frontal (F3), central (C3), parietal (P3) and occipital (O1)), one EOG channel and a video of the driver's face. Objective sleepiness was evaluated on each recording by an expert doctor using the scale described in section I. Data acquisition of physiological signals was performed at 250Hz.

3.2 Technical validation

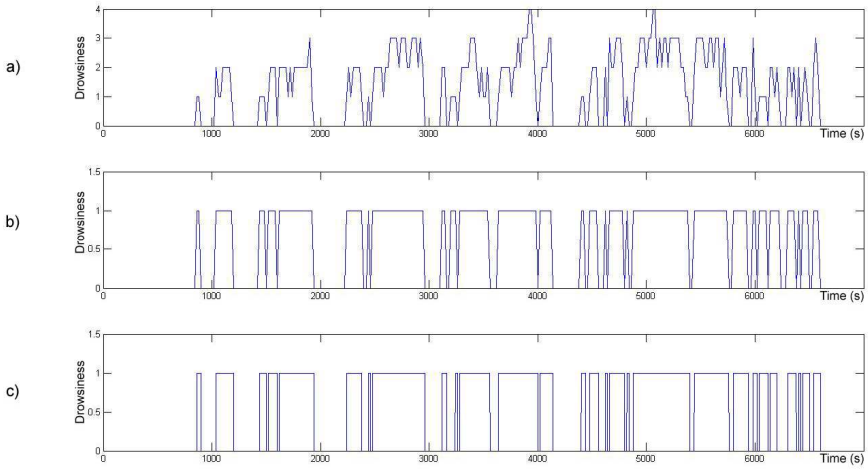

Fig. 9. Comparison between expert decision (a and b) and system decision (c)

The method proposed in this chapter provides a binary decision [awake; drowsy] while the database has been manually labelled using five levels. Moreover, the expert classified non overlapping intervals of 20s (epochs) while the automatic system makes a decision every second. To compare our results with the expert's decision, the following validation technique was used. The five expert decision levels were converted into a binary decision by considering as drowsy any decision superior or equal to 1 in the expert's scale as shown in

fig. 9. This figure shows the expert decision on a five levels scale (a) and on a binary scale (b) and the drowsiness detection obtained using our system (c).
Furthermore, each 20s epoch classified by the expert was directly compared to the system decision: if during the 20s interval, the system classifies at least 1s as "drowsy", then the decision for the epoch was "drowsy". Else it was "awake".
Epochs were then compared one by one and classified according to the table of confusion 2.

		Expert decision	
		awake	drowsy
Automatic decision	awake	True Negative (TN)	False Negative (FN)
	drowsy	False Positives (FP)	True Positive (TP)

Table 2. Table of confusion

The true positive rate (TP_{rate}) or detection rate is the ratio between the number of true "drowsy" automatic decisions and the number of "drowsy" expert decisions. The false positive rate (FP_{rate}) is the ratio between the number of false "drowsy" automatic decisions and the number of "awake" expert decisions. They are calculated according to (6) and (7).

$$TP_{rate} = \frac{TP}{TP + FN} \tag{6}$$

$$FP_{rate} = \frac{FP}{FP + TN} \tag{7}$$

The results are displayed as Receiver Operating Characteristic (ROC) curves (Hanley & McNeil, 1982), plotting TP_{rate} in function of FP_{rate}. The purpose is to have the highest TP_{rate} with the lowest FP_{rate}.

3.3 Results using alpha relative power

3.3.1 Results without artefact detection

The drowsiness detection algorithm was applied on the whole database, with a decision threshold λ (defined in section 2.4) varying from 1.5 to 5, on each of the four EEG channels. The results presented in fig. 10 are those obtained when the MCT is applied on the alpha relative power without considering artefact detection. "Star" markers correspond to the P3 channel, "circle" markers to the F3 channel, "square" markers to the C3 channel and "triangle" markers to the O1 channel. The head at the bottom right of Fig.10 reminds the reader of the position of each channel. For each channel, the results represented with the markers the further on the right corresponds to the smallest λ and those with the markers the further on the left to the biggest λ. It is coherent: increasing λ diminishes the FP_{rate} while decreasing the TP_{rate}.

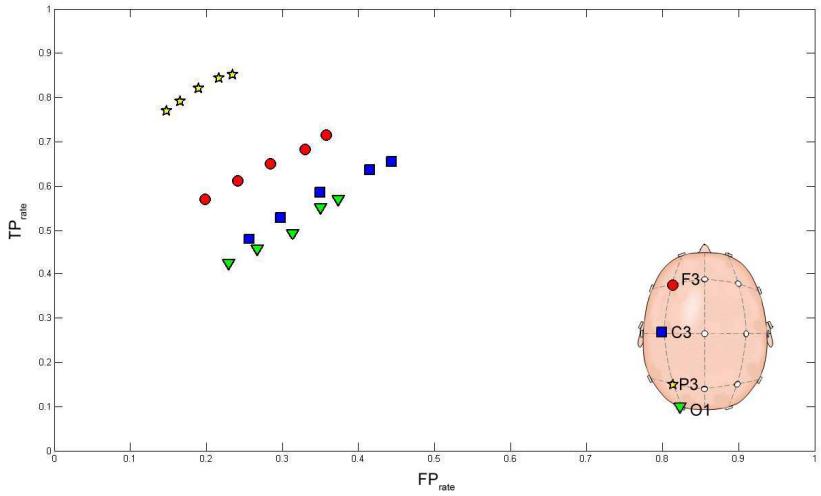

Fig. 10. Results obtained using different EEG channels

It is obvious from fig. 10 that the results are better when the P3 parietal channel is used, which is in concordance with results from the literature: drowsiness is characterized by an increase of α activity predominately in the parietal region of the brain. Indeed, results obtained with EEG recorded by C3, F3 or O1 are only slightly better than those that would be obtained with a random classifier. In the following sections, only the results obtained on P3 channel are shown.

The optimal result on P3 are TP_{rate}=82,1% and TP_{rate}=19,2% with λ=3. However changing the threshold does degrade the performances (TP_{rate}=85,1% and TP_{rate}=23,5% with λ=1,5 and TP_{rate}=76,9% and TP_{rate}=14,8% with λ=5), which proves that the method is not sensitive to the threshold value.

3.3.2 Results with artefact detection

An example of artefact detection is shown on fig. 11. The first signal (a) is the EEG raw data. The signal (b) is the result of the VCT. The dotted line corresponds to the threshold λ_{art}=6. The last signal (c) shows the result of the artefact detection (dotted line): when "zero" no high amplitude artefact is detected and when "one", an artefact is detected. The dotted line boxes underline high-amplitude artefacts.

First, this example shows that the detected artefacts correspond to high-amplitude electric perturbations of the EEG signal. As high-amplitude artefacts have not been evaluated by an expert on the dataset, it is not possible to quantify the performances of the artefact detection method. Nevertheless, a visual check of all the recordings shows that all the apparent high-amplitude artefacts have been detected. Fig. 12 shows the number of artefacts detected in the database (total number and corresponding percentage on the database) in function of the value of the threshold λ_{art} (used for artefact detection).

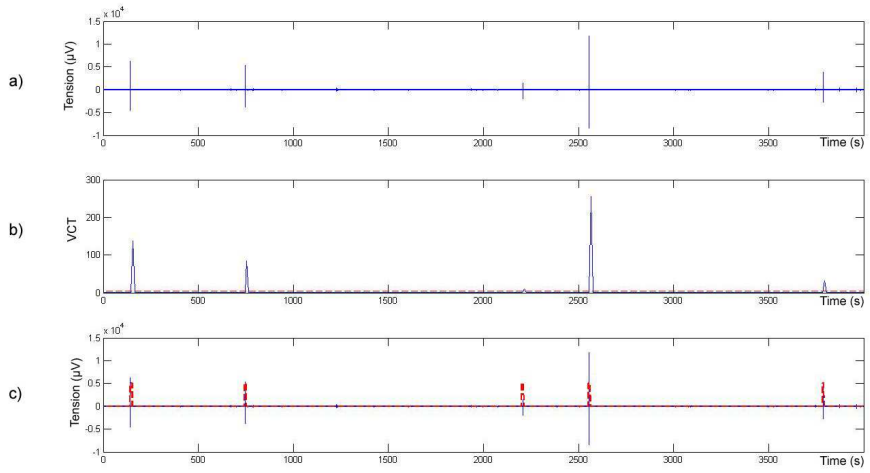

Fig. 11. Example of high amplitude artefact detection

It can be seen in fig. 12 that if the threshold is too small (λ_{art}<5), detection is really sensitive and a lot of points are rejected. Visually, this means a lot of false alarms. When increasing λ_{art}, the number of artefacts detected decreases quickly till λ_{art}=6, which visually seems an appropriate threshold. Indeed, for this threshold value, all the visible high-amplitude artefacts are detected without false alarms. At this point, one can see that high-amplitude artefacts represent only a small part of the dataset: about 2%.

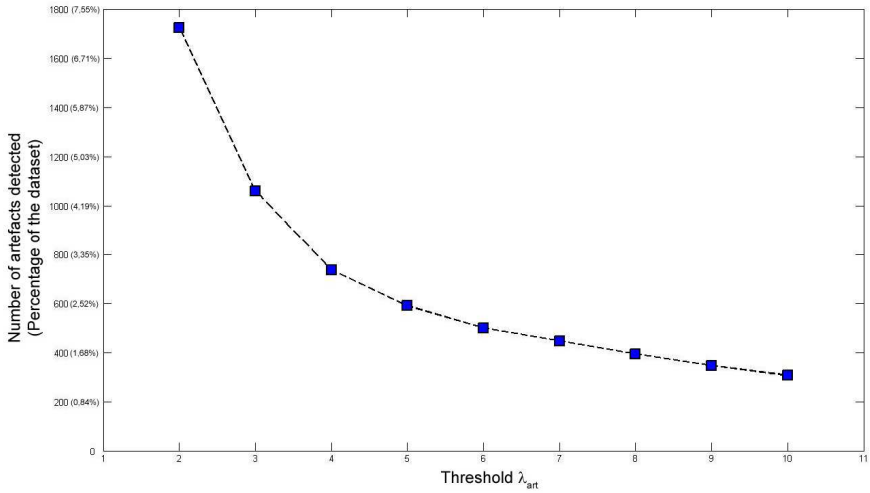

Fig. 12. Number of artefact detected in function of λ_{art}

Results obtained when no decision is made if artefacts are detected are displayed in fig. 13 with "circle" markers. The threshold used for the artefact detection is λ_{art} =6. "square" markers represent the results obtained without considering the artefact detection.

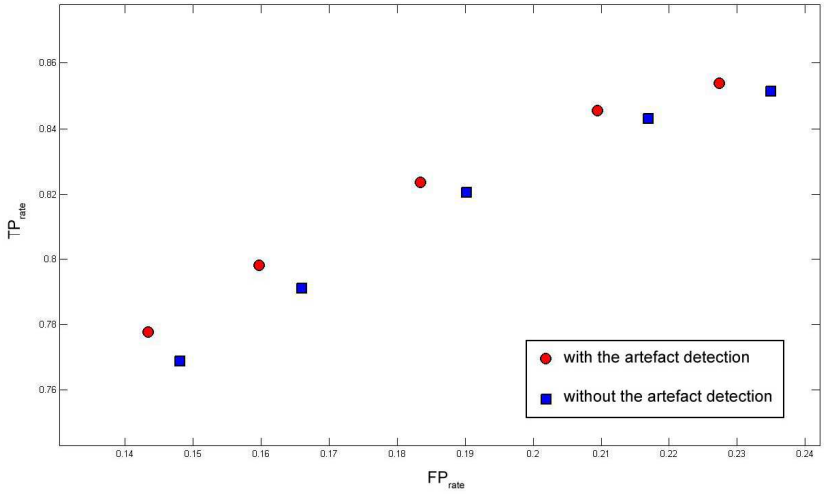

Fig. 13. ROC curve for drowsiness detection when artefact detection is used

It is obvious in fig. 13 that results are slightly improved: TP_{rate} is a bit increased while FP_{rate} is a bit decreased. Using the threshold λ=3, results increase from TP_{rate}=82,1% and FP_{rate}=19,2% to TP_{rate}=82,4% and FP_{rate}=18,3%. So, artefact detection decreases the number of false decisions. The fact that results are only slightly increased can be explained by the fact that high-amplitude artefacts represent about 2% only of the dataset.
Artefact detection will be taken into account in the following sections.

3.4 Results using other features than alpha relative power

Results presented in section 3.3 are now compared to results obtained with other features proposed in the literature. The results are displayed in fig. 14. Results from section 3.3, obtained using MCT on the median filtered α_{rel} signal, are represented by "star" markers. "Square" and "circle" markers represent results obtained using MCT and median filtering respectively on θ_{rel} and β_{rel} signals. Note that β activity increases with cognitive tasks and active concentration, so drowsiness is characterized by a decrease of the β activity. So, the detection algorithm using β as the main feature consider the driver as "drowsy" when the output of the MCT is lower than the threshold $-\lambda$ (varying from -5 to -1). "Triangle" markers correspond to results obtained with the combined signals $\alpha_{rel} | \theta_{rel}$. Decisions are made independently on α_{rel} and on θ_{rel} and then merged with a logical OR. The optimum threshold λ=3 was used for α_{rel}. Displayed results are obtained with a threshold varying from 1,5 to 5 for θ_{rel} detection. The idea to use both α_{rel} and θ_{rel} is inspired by table 1, where

it is assumed that drowsiness is characterized by an increase of the activity in one of the two frequency bands α and θ. "Hexagram" markers represent the (α+θ)/β feature. This feature has been suggested by De Waard and Brookhuis (De Waard & Brookhuis, 1991). As α and θ activity are supposed to increase with drowsiness whereas β activity is supposed to decrease, this feature should be increasing with drowsiness. In this case, MCT is computed on the sum of the signals α_{rel} and θ_{rel} divided by the β_{rel} signal and only one threshold is used to make the decision.

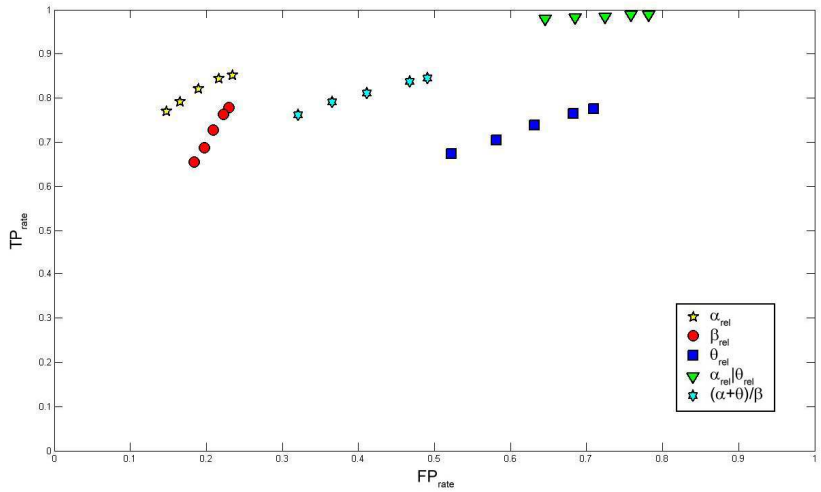

Fig. 14. ROC curves using different features for drowsiness detection

The best results are obtained with the drowsiness detection algorithm applied on the α_{rel} signal (TP_{rate}=82,4%, FP_{rate}=18,3%). Since the algorithm was tested with the same threshold on data recorded from 20 different persons, this tends to show that the method can be applied on different persons without adapting the tuning parameter.
The results obtained with $\alpha_{rel}|\theta_{rel}$ show that θ_{rel} is not relevant to detect drowsiness since the number of false positive increases tremendously when this information is added. This is confirmed by the results obtained with θ_{rel} only. The (α+θ)/β ratio and β_{rel} give correct results (TP_{rate}=76,2% and TP_{rate}=32,1% for (α+θ)/β and TP_{rate}=75,9% and TP_{rate}=24,1% for β_{rel}) but worse than the results obtained with the α_{rel} information only.
Now, if we compare the results obtained with the literature, the results obtained are as good as those found when using a trained algorithm. Lin et al. (Lin et al., 2005a) proposed to monitor driving performance, i.e. the capacity to maintain the car in the middle of the road computing a linear regression model on a 2-channels EEG. They obtained a correlation of r=0,88 between their model and the driving performances when the model is trained and tested on the same session. The correlation decreases to r=0,7 when trained and tested on different sessions. So, this method needs to be tuned for each driver as the model estimated for one driver does not work so well on another. Lin et al. increase these results using ICA on a 33-channel EEG (Lin et al., 2005a) to compute their linear regression model and obtain a

correlation of r=0,88 between their estimation and the driving performances on the testing session. Nevertheless, this model needs to be trained on a large amount of data and has been tested on 5 drivers only. Ben Khalifa et al. (Ben Khalifa et al., 2004) obtained 92% of true drowsiness detections by training a neural network on a the EEG spectrum of the P4-O2 channel but this result is obtained on the training set and decreases to 76% of true detections on the validation set. Moreover, these results are obtained on a set of only 4 drivers. At least, Rosipal et al. (Rosipal et al., 2007) obtained about 77% of true detections of drowsiness states by using hGMM on the spectral content of EEG transferred into a compact form of autoregressive model coefficients. This study has been performed on a large number of drivers and needs a period of training.
The advantage of the method proposed in this paper is that it does not need to be trained or adapted. The same threshold can be used for all drivers. Moreover, as the method has been tested on huge dataset, the results can be considered significant.

3.5 Results merging alpha and beta relative powers

From the previous section, the best results are obtained when α_{rel} or β_{rel} are used as features, which naturally gives the idea to merge these two features to increase the decision reliability
The technique used to merge α_{rel} and b_{rel} is fuzzy logic, which is based on the theory of fuzzy sets developed by Zadeh (Zadeh, 1965). Let us consider $\mu_{Dr}(\alpha_{rel})$ and $\mu_{Dr}(\beta_{rel})$, the membership functions, which represent the membership degree of α_{rel} and β_{rel}, independently considered, to the "drowsy" state. The purpose is to make a decision $Dr(\alpha_{rel}, \beta_{rel})$ using both $\mu_{Dr}(\alpha_{rel})$ and $\mu_{Dr}(\beta_{rel})$,. The driver is considered to be drowsy if both the decision made using α_{rel} and the decision made using β_{rel} is "drowsy". This is expressed thanks to the t-norm product as follows:

$$\mathrm{Dr}(\alpha_{\mathrm{rel}}, \beta_{\mathrm{rel}}) = \frac{\mu_{\mathrm{Dr}}(\alpha_{\mathrm{rel}})\mu_{\mathrm{Dr}}(\beta_{\mathrm{rel}})}{\mu_{\mathrm{Dr}}(\alpha_{\mathrm{rel}})\mu_{\mathrm{Dr}}(\beta_{\mathrm{rel}}) + \mu_{\mathrm{Aw}}(\alpha_{\mathrm{rel}})\mu_{\mathrm{Aw}}(\beta_{\mathrm{rel}})} \tag{8}$$

Note that $\mu_{Aw}(.)$ is the membership function of the "awake" state and is the complementary of $\mu_{Dr}(.)$. The denominator is used here to normalize $Dr(\alpha_{rel},\beta_{rel})$ between 0 and 1.
One has to define the membership function $\mu_{Dr}(\alpha_{rel})$ and $\mu_{Dr}(\beta_{rel})$. A study of the probability of being drowsy in function of the MCT's threshold λ on α_{rel} and β_{rel}, $P(dr|\alpha_{rel})$ and $P(dr|-\beta_{rel})$, is displayed in fig. 15. The "square" markers line displays the experimental $P(dr|\alpha_{rel})$ and the "circle" markers line displays the experimental $P(dr|-\beta_{rel})$. Probabilities are calculated as the percentage of true drowsiness detections obtained on periods when the relative power is over λ.
The membership function $\mu_{Dr}(\alpha_{rel})$ and $\mu_{Dr}(\beta_{rel})$ are then designed from the results presented in fig. 15. As α_{rel} and β_{rel} have a very similar behaviour, the same membership is used for α_{rel} and β_{rel}. This membership function is presented in fig. 16.
The driver is considered as "drowsy" when $Dr(\alpha_{rel},\beta_{rel})$ is larger than 0.5 The results obtained with this method are shown in fig. 17 with the "circle" markers. They are compared to the results obtained using the MCT on α_{rel} only ("square" markers).

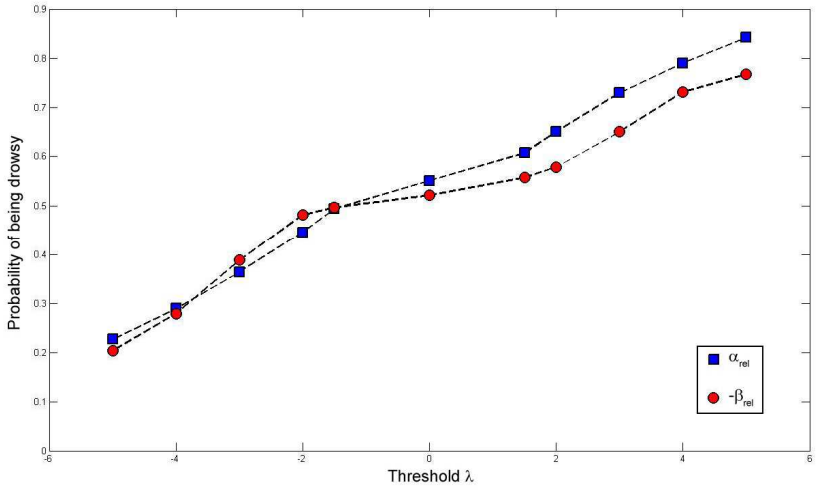

Fig. 15. Experimental probabilities of being drowsy in function of threshold λ

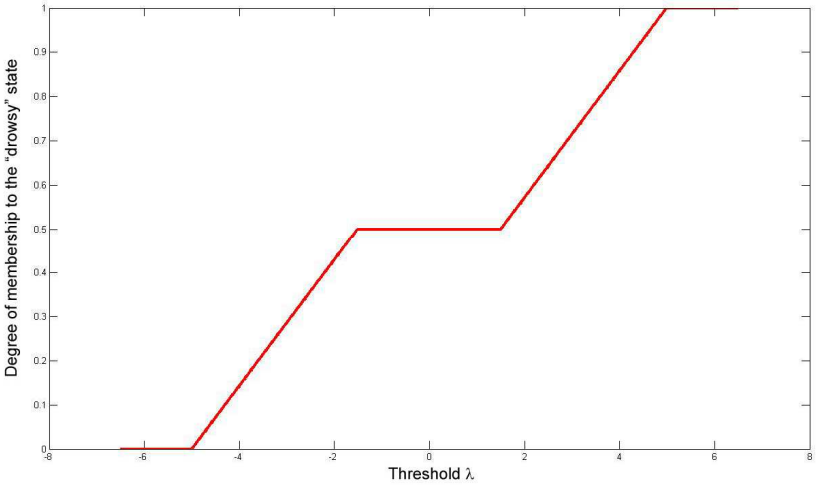

Fig. 16. Membership function in function of threshold λ

Results are improved using this fuzzy logic approach since FP_{rate} is increased and FP_{rate} is decreased. The results obtained with this method are TP_{rate}=84,6% and FP_{rate}=17,9% (TP_{rate}=82,4% and FP_{rate}=18,3% with λ=3 when using only α_{rel} information). This means that the β_{rel} information is relevant to detect drowsiness when combined with α_{rel}. Moreover, compared to the method proposed in section 3.5, there is no need to select an appropriate detection threshold for α_{rel} and β_{rel}. The fuzzy approach increased the detection reliability.

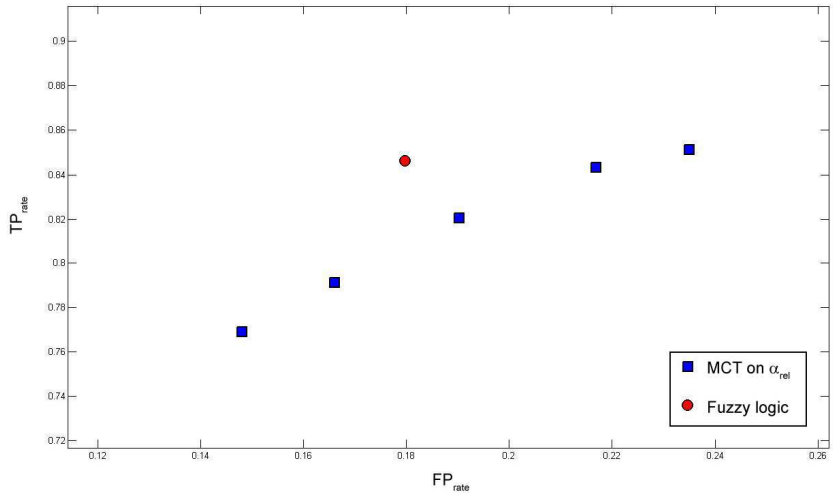

Fig. 17. Performances obtained using α_{rel}, β_{rel} and (α_{rel}, β_{rel}) merged with fuzzy logic

4. Conclusion

An algorithm to detect on line drivers' drowsiness from a P3 EEG channel has been presented in this paper. This algorithm is based on a means comparison test (MCT) applied on the EEG relative power calculated in the alpha band and in the beta band. The results of the MCT test are then merged using fuzzy logic. The algorithm can operate on-line with a short delay and does not need to be tuned. Performances obtained on a large data set recorded from 20 different drivers are 84,6% of true detection and 17,9% of false detection using the parietal EEG channel only. A high-amplitude artefact detection system has been developed and combined to the drowsiness detector. It enables periods of time when the EEG signal is unreliable to be detected on line. No decision is made by the drowsiness detector while the artefact detector classifies the EEG signal as unreliable. The artefact detector is tuned by a single threshold whose value is independent of the driver.

The next step of this work is to add an ``eye blinks and yawn'' detection system thanks to a high frame rate camera and to merge the decisions to obtain a highly reliable automatic drowsiness detector. Fuzzy logic could be a first step to merge this information.

5 Acknowledgements

The authors are grateful to the Centre d'Etudes de Physiologie Appliquée (CEPA) in Strasbourg (FR) for providing the data and their help, as well as the Laboratoire

d'Automatique, de Mécanique, et d'Informatique industrielles et Humaines (LAMIH) in Valenciennes (FR).

6. References

Akerstedt, T. and Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, Vol. 52, pp. 29–37

Ben Khalifa, K., Bédoui, M., Dogui, M., and Alexandre, F. (2004). Alertness states classification by SOM and LVQ neural networks. *International Journal of Information Technology*, Vol. 4, pp. 228–231

Blinowska, K. and Durka, P. (2006). Electroencephalography(EEG), In: *Wiley Encyclopedia of Biomedical Engineering*, Ed. Metin Akay, Wyley

DeWaard, D. and Brookhuis, K. (1991). Assessing driver status: a demonstration experiment on the road. *Accident Analysis and Prevention*, Vol. 23, No. 4, pp. 297–307

Dinges, D. (1995). An overview of sleepiness and accidents. *Journal of sleep research*, Vol. 4, No. 2, pp. 4–14

Galley, N., Schleicher, R., and Galley, L. (2004). Blink parameter as indicators of driver's sleepiness - possibilities and limitations. *Vision in Vehicles*, Vol. 10, pp. 189–196

Grace, R., Byrne, V.E., Bierman, D.M., Legrand, J.-M., Gricourt, D., Davis, B.K., Staszewski, J.J., Carnahan, B. (1998). A drowsy driver detection system for heavy vehicles, *Proc. of the 17th Digital Avionics Systems Conference*, Vol. 2, pp. I36/1-I36/8

Hanley, J. and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, Vol. 143, No. 1, pp. 29–36

Ji, Q., Lan, P., and Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on systems, man, and cybernetics - Part A: Systems and humans*, Vol. 36, No. 5, pp. 862–875

Ji, Q. and Yang, X. (2001). Real time visual cues extraction for monitoring driver vigilance. *Proc. of International Workshop on Computer Vision Systems*, pp. 107–124

Ji, Q., Zhu, Z., and Lan, P. (2004). Real time non-intrusive monitoring and prediction of driver fatigue. *IEEE Transport Vehicle Technology*, Vol. 53, No. 4, pp. 1052–1068

Karrer, K., Vohringer-Kuhnt, T., Baumgarten, T., and Briest, S. (2004). The role of individual differences in driver fatigue prediction. *The third International Conference on Traffic and Transportation Psychology*

Kay, A., Trinder, J., Bowes, G., , and Kim, Y. (1994). Changes in airway resistance during sleep onset. *Journal of Applied Physiology*, Vol. 76, pp. 1600–1607

Klein, R., Allen, R., and Miller, J. (1980). Relationship between truck ride quality and safety of operations: methodology development. Technical Report DOT HS 805 494, Systems Technology, Inc., Hawthorne, CA

Knippling, R. (1998). Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Technical Report FHWA-MCRT-98-006, Federal Highway Administration

Lin, C., Wu, R., Liang, S., Chao, W., Chen, Y., and Jung, T. (2005a). Eeg-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems*, Vol. 52, No. 12, pp. 2726–2738

Lin, C.-T., Wu, R.-C., Jung, T.-P., Liang, S.-F., and Huang, T.-Y. (2005b). Estimating driving performance based on eeg spectrum analysis. *EURASIP Journal on Applied Signal Processing*, Vol. 19, pp. 3165–3174

Makeig, S., Bell, A., Jung, T.-P., and Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, Vol. 8, pp. 145–151

Muzet, A., Pébayle, T., Langrognet, J., and Otmani, S. (2003). Awake pilot study no.2: Testing steering grip sensor measures. Technical Report IST-2000-28062, CEPA

O'Hanlon, F. and Kelley, G. (1974). A psychophysiological evaluation of devices for preventing lane drift and run-off-road accidents. Technical Report 1736-F, Federal Highway Administration

Picot, A., Charbonnier, S. and Caplier, A. (2008). On-line automatic detection of driver drowsiness using a single EEG channel. *30th Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3864-3867.

Quanten, S., Valck, E. D., Cluydts, R., and Berckmans, D. (2006). Thermoregulatory changes at driver sleepiness. *International Journal of Vehicle Design*, Vol. 42, pp. 87–100.

Ragot, J., Darouach, M., Maquin, D., and Bloch, G. (1990). *Validation de données et diagnostic.* Ed. Hermès, Traité des nouvelles technologies

Rechtschaffen, A. and Kales, A. (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subject*. National Institute of Health Publication, Washington DC, us government printing office edition

Renner, G. and Mehring, S. (1997). Lane departure and drowsiness - two major accident causes - one safety system. Technical report, Transport Research Laboratory

Rosipal, R., Peters, B., Kecklund, G., Akerstedt, T., Gruber, G., Woertz, M., Anderer, P. and Dorffner, G. (2007). EEG-Based Drivers' Drowsiness Monitoring Using a Hierarchical Gaussian Mixture Model, *Hum. Comp. Interaction*, Vol. 16, pp. 294-303

Royal, D. (2002). National survey of distracted and drowsy driving attitudes and behavior. Technical Report DOT HS 809 566, National Highway Traffic Safety Administration

Santamaria, J. and Chiappa, K. H. (1987). The EEG of Drowsiness in normal Adults. Journal of clinical Neurophysiology, Vol. 4, No. 4, pp. 327-382.

Smith, P., Shah, M., and Lobo, N. D. V. (2003). Determining driver visual attention with one camera. *IEEE Trans. on intelligent transportation systems*, Vol. 4, No. 4, pp. 205–218

Törnros, J., Peters, B., and Östlund, J. (2000). Heart rate measures as drowsiness indicators. Technical report, Swedish National Road and Transport Research Institute

Vural, E., Cetin, M., Ercil, A., Littlewort, G., .Bartlett, M. and Movellan, J., (2007). Drowsy driver detection through facial movement analysis, *Proc. ICCV*, pp. 6-18.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio & Electroacoust.*, Vol. 15, pp. 70–73

Wierwille, W., Ellworth, L., Wreggit, S., Fairbanks, R., and Kim, C. (1994). Research on vehicle-based driver status/performance monitoring. Technical report, NHTSA

Zadeh, L. (1965). "Fuzzy sets". *Information and Control* , Vol. 8, No. 3, pp. 338-353