A model to predict hypovigilance during a monotonous task

Larue, G. S. 1, Rakotonirainy A. 1, Pettitt A. N. 2

1Centre for Accident Research and Road Safety – Queensland; 2Faculty of science and technology, Queensland University of Technology
email: g.larue@qut.edu.au

Abstract

The driving task requires sustained attention during prolonged periods, and can be performed in highly predictable or repetitive environments. Such conditions could create drowsiness or hypovigilance and impair the ability to react to critical events. Identifying vigilance decrement in monotonous conditions has been a major subject of research, but no research to date has attempted to predict this vigilance decrement. This pilot study aims to show that vigilance decrements due to monotonous tasks can be predicted through mathematical modelling. A short vigilance task sensitive to short periods of lapses of vigilance called Sustained Attention to Response Task is used to assess participants’ performance. This task models the driver’s ability to cope with unpredictable events by performing the expected action. A Hidden Markov Model (HMM) is proposed to predict participants’ hypovigilance. Driver’s vigilance evolution is modelled as a hidden state and is correlated to an observable variable: the participant’s reaction time. This experiment shows that the monotony of the task can lead to an important vigilance decline in less than five minutes. This impairment can be predicted four minutes in advance with an 86% accuracy using HMMs. This experiment showed that mathematical models such as HMM can efficiently predict hypovigilance through surrogate measures. The presented model could result in the development of an in-vehicle device that detects driver hypovigilance in advance and warn the driver accordingly, thus offering the potential to enhance road safety and prevent road crashes.

Keywords

Monotony, Fatigue, Vigilance, Hidden Markov Models

Introduction

Drowsiness at the wheel has been identified globally as a major cause of road crashes. Inattention and fatigue are reported as contributing factors in 6% and 5% of fatal crashes respectively in Australia between 1992 and 2006 [1]. It is difficult to reliably measure the influence of such contributing factors so that such estimates are likely to be underestimated. This is supported by the survey conducted by McCartt et al. [2] where 55% of 1000 drivers had reported to have driven while drowsy and 23% had fallen asleep while driving without having a crash.

Boredom, fatigue, monotony, sleep deprivation are factors that induce sleepiness and drowsiness. It results in decreased attention, impaired information processing ability and impairs decision-making capability. These factors increase crash risk due to driver inability to react to emergency-type situations [3]. Most research on vigilance-related impairments focuses on sleep-deprived participants [4, 5]. However there is evidence from crash data and from simulated driving studies that vigilance decrement could occur during daytime especially on monotonous roads [6]. Driver hypovigilance is often attributed to fatigue but can emerge independently of time on task and is more frequent in monotonous road environments where task demand and stimulus variability are low and moderates sustained attention [6, 7].

Driver’s self-assessment questionnaires have been used to evaluate their vigilance state. Such a subjective approach is not applicable on monotonous roads [8] suggesting the need for an objective mathematical model to predict vigilance decrement during driving. The most reliable assessment of vigilance is obtained by electroencephalography (EEG) [9]. However such a device is too obtrusive to be deployed in vehicles. Driving performance is impaired during vigilance decrement and surrogate measures from the driver, the car and the environment can be used to assess such impairment. This paper presents a pilot
study designed to assess the feasibility of predicting vigilance decrement on monotonous roads. A low task demand, lab-based vigilance task is used to isolate and simulate impairments due to monotony in a vigilance task. The Sustained Attention to Response Task (SART) is a well-known vigilance task satisfying such requirements. Our aim is to predict hypovigilance during a short, monotonous vigilance task using surrogate measures (reaction times).

Background

Vigilance

Vigilance is defined as the state of readiness to detect and respond to certain specified small changes occurring at random time intervals in the environment [10]. To a broader sense, vigilance can be defined as the ability to sustain attention to a task for a period of time [11]. Vigilance fluctuates and is an issue in terms of road safety when decreasing. This particularly applies to monotonous environments where driving is largely reduced to a visual vigilance task (lane keeping task).

Factors that have an effect on vigilance can be divided into two categories: (i) endogenous and (ii) exogenous factors. Endogenous factors are associated with long-term fluctuation of alertness which emanate from within the organism, whereas exogenous factors are linked to the task itself or the interaction between the driver and the outside environment. Among the endogenous factors are both physical and mental fatigue, sleep deprivation as well as task duration. Personality dimension (age, gender, mood and particularly sensation seeking level), time of day (circadian rhythms), caffeine and other stimulants and cognitive task demands are also endogenous factors [12]. Exogenous factors include complexity and monotony of the task, environmental factors such as noise, ambient temperature, frequency and variation of stimulation [13]. This is particularly the case when driving on a highly predictable highway where, because of lack of stimuli (or repetitive ones), the driver pays less attention to the road situation [14]. These numerous factors result in a complex and strongly interrelated phenomena regulating vigilance. The impacts of the different factors leading to changes in vigilance performance are not of the same order. Stressors such as heat, noise and circadian effects are of low impact on the performance compared to fatigue, monotony and/or boredom [15].

Each individual has their personal optimal level of stimulation and arousal required to perform well. The Sensation Seeking Scale was developed to provide such a measure. High sensation seekers are expected to experience vigilance decrement faster than any other group. Sensation seeking drivers need varied, complex sensations and experiences, and are able to take physical and social risks to achieve such experiences [6]. This leads to risk taking driving, and negative reactions to monotonous driving [16].

Vigilance tasks are the paradigm used to study sustained attention and its vigilance decrement. Vigilance level is often assessed automatically by an algorithm through the estimated performance (from 0 to 1) to the task. We classify vigilance levels into three categories as described by Dutta et al. [17]:

- alert: 0.7-1
- intermediate (medium): 0.3-0.7
- drowsy or hypovigilant: 0-0.3

In this experiment an adaptation of the SART is used where participants are asked to respond to non-targets and not respond to targets. In such an experiment the vigilance as assessed by performance has been shown to depend on the level of monotony and is correlated to reaction times (RTs) [18]. The SART was chosen as it is a well-established task to measure vigilance level during continuous tasks. We hypothesise that the SART exhibits the same characteristics as driving on monotonous roads. Indeed like the SART, driving on monotonous roads consists of sustained monitoring of visual stimuli with a preparedness to respond to infrequent critical events such as an impending crash. The SART has been shown to lead to insufficient level of arousal and insufficiently strong representation of the task goals which might be a reasonable model for what happens when a train driver passes through a signal at danger (failure to react to a no-go) [19]. Driving a train is typically routine, minimising the need for effortful attention on the part of the train driver. Motorway driving has similar demands, of maintaining vigilance and alertness in the face of monotony and routine responding. Also it has been observed that the
SART activates the right frontoparietal lobe of the brain, which shows that it places demands on the vigilant attention system [20].

**Mathematical model for prediction**

Vigilance decrement can manifest quite early on [21] and change quite abruptly during monotonous vigilance tasks. This can be well described by discrete modelling. Vigilance, defined as the accuracy of target detection, is categorised as presented before. We aim to predict vigilance through surrogate variables that are correlated to the vigilance state. Sleep research has shown that such performance models must be able to deal with inter-individual differences to be implemented reliably in operational settings. Bayesian forecasting is widely used to overcome this limitation. Indeed such models can handle these differences even when prediction is applied to individuals not studied beforehand [22]. Among Bayesian models, Hidden Markov Models (HMMs) have been used to model numbers of real-life problems, such as driver manoeuvre recognition [23]. They combine independence assumptions making the model numerically computable with field knowledge that vigilance decrement is the cause of reaction time variations [24].

A Hidden Markov Model is designed to model a sequence of $T$ observations data (at time $t=1,2,...,T$) which is the consequence of an unobserved (hidden) variable [25]. Here the unobserved variable is the vigilance level $V_{igt}$ at time $t$. This variable is the cause of other random variables, the surrogate measure $RT_t$ at time $t$ in this study. These variables must have the following conditional independence properties for each time $t$ [26]:

- given $V_{igt}$, the sequences $\{V_{igt}, RT_{igt}\}$ and $\{V_{igt+1}, RT_{igt+1}\}$ are independent, where the notation $A_{a:b} = \{A_a, \ldots, A_b\}$ is used (Markov property of order one)
- given $V_{igt}$, $RT_t$ is independent of the sequence $\{V_{igt}, RT_{igt}\}$, where the notation $A_{a:-b} = \{A_1, \ldots, A_{b-1}, A_{t+1}, \ldots, A_T\}$ is used.

In the case of a HMM with discrete states and discrete observations sequences, the model is completely characterised in terms of [25, 27]:

- number of states in the model, say $N$. Here the random variable $V_{igt}$ takes its values in the set $S=\{\text{fully awake, intermediate, drowsy}\}$ so that $N=3$.
- number of distinct observation symbols per state. Here it is the reaction times values once categorised.
- transition probability matrix giving the probability to go from the state $S_i$ at time $t$ to the state $S_j$ at time $t+1$.
- observation symbol probability distribution.
- initial state distribution.

The training of the HMM is done through Bayesian learning from the given hidden and observation sequences. If the hidden state is not available during training, the Baum-Welch algorithm (adaptation of the EM-algorithm applied to HMM training) can be used. Then the model can be used for prediction (see Figure 1). The Viterbi algorithm is used to infer the value of the vigilance state given the reaction times [25]. This algorithm determines the states sequence, respecting the transition probabilities, that is the most likely to occur with the model used. Then predicting the next vigilance state can be done using the transition probability matrix.

![Figure 1: Prediction methodology with HMMs](image-url)
Methods

Participants

Forty students of the Queensland University of Technology (QUT), 8 males and 32 females (mean age = 22.6 years, SD = 9.2), volunteered to participate in this study. All subjects provided written consent for this study, which was approved by the QUT ethics committee. Students undertaking the first year psychology subject received course credit for their participation.

Experimental design

Two five minutes adaptations of a continuous sustained attention task (SART) [19] were run on an IBM compatible computer using E-Prime. The conditions varied in terms of task monotony, with two different settings for target appearance: (i) probability 0.11 (low target probability) and (ii) probability 0.5 (high target probability). The first one creates a monotonous condition where a response can be predicted and leads to automatic responses. The second one, with a markedly higher stimulation, is a non-monotonous condition and results in a non-automatic response mode, associated with a lower response predictability [7].

Experimental conditions

This experiment was designed by Michael and Meuter [7]. 225 single digits (ranging from 1 to 9, height of 29 mm) were displayed randomly for 250 ms in the middle of a computer screen. An inter-stimulus interval of 1150 ms was used with a mask (height 29 mm) consisted of an “X”. The chosen target stimulus was the display of number 3. When a stimulus different from the target stimulus was displayed, the participant was asked to press the spacebar as fast as possible, and when the target number was displayed, action required was to withhold the response (that is to say not press the spacebar).

Procedure

Participants were tested individually in a quiet room, between 9am and 3pm, in a session lasting approximately 45 minutes. They were randomly assigned in two groups, each of which performed five short vigilance tasks, as follows. Each participant performed a monotonous then a non-monotonous task, followed by one of various types of monotonous tasks (this task formed part of a larger study and will not be further described here). Finally, there was a repetition of the monotonous and non-monotonous tasks, participants of the second group performing this sequence in a counterbalanced order with time. Prior to each condition, participants received written instructions on the computer screen. The instructions asked them to respond as quickly as possible to all stimuli, and this with the highest accuracy possible. On completion, participants filled out short questionnaires: the Sensation Seeking Scale - Form V (SSS), the General Health Questionnaire (GH-28) to screen and eliminate participants for psychiatric morbidity (found to impair performance using the SART) and a general background questionnaire (control sleep pattern and caffeine consumption).

Data analysis

The software Matlab version 7.4.0.287 was used to analyse data. Responses to target are used to assess vigilance fluctuations. They are converted into error rates in fixed time windows, defined as the fraction of targets not detected by the subject (i.e. lapses) within a fixed window. Due to the small number of targets in the monotonous setting, a window size of 45 stimuli (targets and non-targets) was chosen to obtain an average number of five targets in the window in the monotonous setting. This window size corresponds to approximately one minute. The window size was chosen to be the same for the non-monotonous task. Pearson's linear correlation coefficient between the reaction times (resp. performance) of two consecutive time windows is computed to test whether assumptions required during HMM modelling are reasonable. Performance is then divided into states as described in the ‘Vigilance’ subsection. The predictor reaction time is computed as the mean response time to non-targets. Reaction times are normalised per participant and then categorised.
The sensation seeking level of the participant is categorised into one of the following classes: low (less than one standard deviation (S.D.) in the available participants sample), normal (within one S.D.) or high (greater than one S.D.) [28].

Six different HMMs are fitted to take into account the impact of the monotony of the task (monotonous or not) and the sensation seeking scale (low, medium and high level). Vigilance states and reaction times are known when the model is trained. That way, computing the joint distribution is only a matter of counting the different transitions from the different performance states and the probability of observation of the different reaction times for each vigilance state (Bayesian learning) [29].

A stratified 10-fold cross-validation is performed to assess the robustness of the modelling. In this technique data are divided into 10 folds. The model is trained on 9 and tested on the remaining one. This is repeated so that each fold is used as a test sample [30]. A stratified cross-validation was used to avoid putting high and low sensation seekers in the same fold.

The most probable performance state sequence at time $t$ using the reaction times data until time $t$ is computed with the Viterbi algorithm. This gives the probable vigilance state at this time. Future vigilance states are then inferred up to four minutes in advance using the transition probability matrix. The model's accuracy is evaluated through the capacity to detect hypovigilance occurrences reliably. Therefore sensitivity and specificity are reported. Sensitivity measures the proportion of actual drowsy states which are correctly identified as such while specificity measures the proportion of non-drowsy states which are correctly detected. Their harmonic mean, the $F$-statistic, is also provided.

**Results**

The correlation between two consecutive performance measures was $\rho = 0.70$ while the correlation between two consecutive reaction times was $\rho = 0.18$. This shows that vigilance evolution is progressive and depends on the previous state. Particularly, there was no need to use a Markov property of order higher than one. Such observation was not true in the case of reaction times. Reaction times are not equivalent to the performance level though they depend on it (a reaction time value does not correspond to a specific vigilance state). This supports the choice of HMMs, their assumptions being compatible with the data.

The non-monotonous setting of the SART did not create hypovigilance (drowsy state), with only two occurrences appearing when considering all the participants. On the other hand the monotonous setting resulted in 104 occurrences of hypovigilance. Therefore there was no need to detect hypovigilance on the non-monotonous setting, and only results on the monotonous setting were further analysed. Reaction times were categorised in a varied number of categories, and best results in terms of prediction were obtained for 19 categories. The values of the transition probabilities for the corresponding HMM are shown in Table 1. In this monotonous setting, only low sensation seeking participants may stay in the alert state. On the other hand each participant – independently of their sensation seeking level - is highly likely (66% for medium sensation seekers) to stay in the drowsy state once they reach it.

<table>
<thead>
<tr>
<th></th>
<th>Alert</th>
<th>Medium SS</th>
<th>High SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alert</td>
<td>74</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Medium</td>
<td>19</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>Drowsy</td>
<td>7</td>
<td>61</td>
<td>66</td>
</tr>
</tbody>
</table>

The trained HMM has been used to make predictions using reaction times until time $t$ up to four minutes in advance. The accuracy of these predictions is reported in Table 2. The prediction of vigilance at time $t$ has an $F$-statistic value of 79.4% (73.1% and 86.9% for the sensitivity and specificity respectively). This $F$-statistic increases as prediction steps increase up to four minutes ($t+4$) reaching 86.3% (100% and...
75.9% for the sensitivity and specificity respectively). This increase is due to an increase in sensitivity (while specificity decreases) and results from the high likelihood to finish the experiment in the drowsy state.

<table>
<thead>
<tr>
<th>Prediction step (minutes)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0</td>
<td>73.1</td>
<td>86.9</td>
<td>79.4</td>
</tr>
<tr>
<td>+1</td>
<td>88.5</td>
<td>76.0</td>
<td>81.8</td>
</tr>
<tr>
<td>+2</td>
<td>95.5</td>
<td>72.3</td>
<td>82.3</td>
</tr>
<tr>
<td>+3</td>
<td>97.9</td>
<td>73.5</td>
<td>83.9</td>
</tr>
<tr>
<td>+4</td>
<td>100</td>
<td>75.9</td>
<td>86.3</td>
</tr>
</tbody>
</table>

Discussion

This short vigilance task shows that the monotony of the task can lead to important vigilance impairment. This impairment is not observed in the non-monotonous setting. Sensation seeking level changes the way participants cope with the monotonous setting. High and medium sensation seekers are not able to maintain high vigilance whereas low sensation seekers can. Also high and medium sensation seekers tend to have an immediate and fast decrease in vigilance, going from an alert state to the drowsy one directly with 39% and 26% probability respectively. By contrast vigilance decrement for low sensation seekers is less abrupt and goes through the intermediate vigilance level (7% probability to go straight from the alert state to the drowsy state). These results on sensation seeking level’s impact on vigilance decrement are in line with previous research conducted on a driving simulator by Thiffault & Bergeron [6] where steering wheel movements were used as a measure of driving performance.

The vigilance decrement can be accurately detected and predicted up to four minutes in advance through surrogate measures (here reaction times) using HMMs with an F-statistic around 80%. Although the increase in the F-statistic as the prediction step increases is counterintuitive, it can be explained in this experiment. Independently of the sensation seeking level, participants are highly likely to finish the experiment in the drowsy state when the setting is monotonous. Therefore, it is easier to predict the vigilance state closer to the end of the experiment, which results in better predictions.

Limitations

Models were trained according to the Sensation Seeking Scale level, so that a population modelling approach has been used in this study. Adapting models to each participant should improve these results. Also, the sample of participants is heavily biased by age, gender and possibly intellectual capacity compared to the wider population due to the sampling population being university students. Nevertheless, generalisation of the results found in this pilot study seems reasonable due to the simplicity of the task involved.

Conclusion

We show on a short vigilance task that monotony can quickly lead to critical vigilance impairment. Such impairment depends importantly on the sensation seeking level of the participant and is detected through task performance. In view of predicting hypovigilance during driving, this vigilance decrement has to be detected through surrogate measures. Indeed, the most reliable and most often used method to assess vigilance is electroencephalography, which cannot be implemented in a real car. This experiment shows that the vigilance decrement can be predicted using reaction times as surrogate measures with 80% to 86% accuracy and up to four minutes in advance. Such results support the idea to use HMMs to predict hypovigilance during driving, using surrogate measures. Different measures, such as lane-keeping, where steering wheel movements or eye-tracking performance, have been shown in the literature to be altered when the driver vigilance is impaired. Such further research could be implemented in an in-vehicle device to predict driver vigilance decrement and therefore prevent crashes.
References


Peer-reviewed full paper