Robust evaluation of time since awakening using force platform posturography

Quang-Cuong Pham, Marco Túlio de Mello, Fernanda Veruska Narciso, Marcos Mônico Neto, Cristiane Westin Teixeira, Leandro Stetner Antonietti, Silvério Aparecido Garbuio, Silvio de Araújo Fernandes Júnior, Renata Guedes Koyama, Marcos Duarte*, Sílvio de Araújo Fernandes Júnior, Renata Guedes Koyama, Marcos Duarte*, Sérgio Tufik

Abstract Introduction: Sleepiness is responsible for a considerable proportion of traffic accidents. It is thus an important traffic safety issue to find a robust, objective and practical way to estimate the amount of time a person has been awake. To attempt to meet this goal, we investigated the relationship between sleepiness and posture control. Methods: Subjects were kept awake for 36 hours and posturographic data during quiet standing were collected every two hours by means of a force platform. The standing surface (rigid surface or foam surface) and visual (eyes open or eyes closed) conditions were manipulated. Results: In the more challenging conditions (with foam surface and/or eyes closed), the body sway variables derived from the center of the pressure measurement increased significantly when time since awakening became greater than 21 h in almost all subjects. Conclusion: Based on this result, we propose a practical protocol that could robustly assess whether time since awakening was greater than 21 h.

Keywords Sleepiness, Time since awakening, Posturography, Postural control.

Introduction

Sleepiness and fatigue are responsible for a considerable proportion of traffic accidents (Pinho et al., 2006; Noce et al., 2008; Williamson et al., 2014), particularly in truck and inter-city bus drivers, who must drive for long periods of time. It is thus an important traffic safety issue to find a robust, objective (as opposed to subjective reports such as the Stanford Sleepiness Scale) (Hoddes et al., 1973) and practical way to estimate the amount of time a person has been awake (time since awakening or TA). Posturography is an interesting candidate (Häggström et al., 2006; Morad et al., 2007; Forsman et al., 2010; Smith et al., 2012; Tietavainen et al., 2013) because it yields objective measures and can be performed by means of cheap and easily transportable devices such as force platforms. Basically, posturography is the measurement of human posture (Winter, 1995). When a human subject is asked to stand as still as possible, he actually exhibits small, involuntary, movements called postural sway (Winter, 1995; Duarte and Zatsiorsky, 2000). This sway can be quantified from the trajectories of the Center of Pressure (COP), which is typically recorded by force platforms (Winter, 1995), or alternatively quantified from body acceleration (Mancini et al., 2012). Other possible methods for detecting fatigue related to sleepiness include using electroencephalography and electrooculography (Hirvonen et al., 2010; Liu et al., 2013; Hallvig et al., 2014).

It can be hypothesized that the amount of sway is related to fatigue, which in turn is related to sleepiness and TA. However, simple measurements of the amount of sway, such as the standard deviation, area, and velocity of COP trajectories during conventional normal upright standing, are not always significantly correlated with TA and novel measurements, such as fractal dimension and critical time from a stabilogram diffusion plot, have been proposed because they achieved a higher correlation with TA (Häggström et al., 2006; Morad et al., 2007; Smith et al., 2012; Tietavainen et al., 2013).

We believe that standing still in normal conditions is too simple a task to reveal any clear effect of TA on the amount of sway. We therefore designed an experiment where the subjects had to stand still in more challenging conditions, namely, on a foam surface and/or without visual feedback. Indeed, it has been shown that together or separately, these experimental conditions significantly increase the difficulty of the task (Woollacott and Shumway-Cook, 2002), and more specifically, that standing with closed eyes enhances the sleepiness evaluation by posturography (Morad et al., 2007). Our hypothesis is that under the more challenging conditions we suggest (standing on foam surface and/or with eyes
closed), the effect of TA on the amount of sway will be clearly observable. Based on the experimental results, we proposed a protocol to evaluate TA from posturographic data. We statistically validated this protocol on a subject-by-subject basis and argue that it can be rapidly developed for real-life application.

**Methods**

**Subjects, material, protocol**

Fourteen male subjects participated in this experiment, which conformed to the Declaration of Helsinki. This work was approved by our local ethical committee (number 0268/09), and it was carried out with the understanding and written consent of each subject. The subjects were healthy and reported neither sleep disorders (assessed by polysomnography) nor postural deficits; their ages, heights and weights were 23±4 years, 1.79±0.07 m and 73±14 kg, respectively.

The subjects slept at the laboratory (from 11 p.m.) the night before the experiment started. They were awakened at 7 a.m. and then kept awake and monitored for 36 consecutive hours. Posturographic data were collected by means of a force platform (AMTI, USA) with a sampling frequency of 100 Hz. Data collection was performed every two hours, from TA = 2 h to TA = 36 h, resulting in 18 data collection times. Each data collection consisted of 8 trials per subject, yielding 18 × 8 = 144 trials in total for each subject.

In each trial, the subject was instructed to stand on the force platform in a fully erect position and to stay as still as possible for 40 seconds. We tested two standing surface conditions and two visual conditions: Rigid Surface (R) and Foam Surface (F), Eyes Open (O) and Eyes Closed (C). This procedure resulted in four experimental conditions: RO, RC, FO and FC. Condition RO is the “normal” or “control” condition that has been tested in previous studies (Hæggström et al., 2006; Morad et al., 2007). The 8 trials of each data collection time point were performed in the following order: two RO trials, two RC trials, two FO trials and two FC trials.

In the Eyes Open trials, the subject had to fixate his gaze on a cross drawn on the wall ~ 2 m in front of him at eye-level. In the Foam Surface trials, viscoelastic foam (Ball Dynamics, USA; model: AIREX Balance Pad, USA; dimension: 41×49×6 cm³) was placed on the force platform and the subject was asked to stand on top of the foam.

**Data processing and statistics**

Matlab 7.8 software (MathWorks, USA) was used for processing the data. The COP time series were first filtered using a fourth-order zero-lag low-pass Butterworth filter with cut-off frequency of 10 Hz. The first and last five seconds of each trial were then removed, resulting in a 30 s time series (see figure 1 for an example of the COP trajectories).

Several simple variables, which are directly related to the amount of sway, were considered. However, because they yielded similar results, we showed only results corresponding to the variable AREA, the area of the ellipse computed by Principal Component Analysis such that there is a 95% probability that a new observation will lie inside the ellipse (Oliveira et al., 1996). In addition, for direct comparison with previous findings reported by Hæggström et al. (2006), we...
calculated the same novel variables used by them: Critical Time (CT), which they called “time interval of open-loop control”; and Fractal Dimension (FD). The critical time variable is defined as the time interval where there is a general change in the trend of the stabilogram diffusion plot as illustrated in Figure 2. In general, the fractal dimension variable is a statistical measure of the complexity in the data and, using the definition employed by Hæggström et al. (2006), it will be calculated as the slope of a straight line fitted to the first region (up to the critical time) in the stabilogram diffusion plot in a log-log scale. In comparison to the original proposition of the stabilogram diffusion plot applied to the COP analysis (Collins and De Luca, 1993), this calculation corresponds to two times the Hurst exponent in that region.

To statistically assess whether TA has an effect on these variables, we separated the data collection points into two bins: the “early” bin with TA < 21 h (comprising 10 data collection points, from 2 h to 20 h) and the “late” bin with TA > 21 h (comprising 8 data collection points, from 22 h to 36 h). Several prior studies on driving performance and surveys of traffic accidents reported critical thresholds as a result of extended wakefulness of approximately 19-20 h (Dawson and Reid, 1997; Howard et al., 2007; Noce et al., 2008). Normality and homogeneity of variances of the variables were verified using the Shapiro-Wilk test and the Levene statistic, respectively, and two-way ANOVA tests (the first factor was the subject with 14 levels, one level per subject; the second factor was the TA, with 2 levels: “early” and “late”) were then applied to the values of each variable. The level of significance of the tests was set at p = 0.01. The eta-squared ($\eta^2$) effect size of each variable was also quantified.

Note that contrary to Hæggström et al. (2006), we did not apply any deconvolution to our data to remove circadian rhythm to avoid the necessity of measuring the circadian rhythm for each subject. However because the sizes of the bins were large, any significant global trends of the variables would still be detected.

**Results**

The values of several variables were plotted for four typical subjects in Figure 3. Intuitively, a variable will yield clear discrimination between the “early” and “late” bins of trials if (i) the variability within each bin (intra-bin variabilities) is low and (ii) the difference of the average values of the two bins (inter-bin difference) is high. One can observe graphically that in condition RO, requirement (i) – cf. subject LUVE – and requirement (ii) – cf. subject ANPE – were not fulfilled by the variable AREA. This agrees with previous work that reported that simple posturographic variables were not sufficient to evaluate TA in normal visual and standing surface conditions (Hæggström et al., 2006).

As expected (Woollacott and Shumway-Cook, 2002), increasing the difficulty of the task led to larger values of simple variables; for instance, the average values of AREA across subjects and data collection points were $1.9\pm1.8 \text{ cm}^2$ in condition RO, $2.7\pm2.3 \text{ cm}^2$ in condition RC, $8.4\pm3.9 \text{ cm}^2$ in condition FO and $18.5\pm8.5 \text{ cm}^2$ in condition FC.

In agreement with our main hypothesis, increasing values of the simple variables correlated with a stronger effect of TA on these variables. For instance, one could observe in Figure 3 that the two requirements mentioned above (low intra-bin variabilities and high inter-bin differences as shown by the superimposed horizontal segments) were fulfilled in conditions RC, FO and FC. Table 1 quantifies these observations. Although TA has a significant effect on AREA in all conditions ($p < 0.01$), the effect size $\eta^2$ (given by $\frac{SS_{TA}}{SS_{TOTAL}}$) was much higher in condition FC ($\eta^2 = 0.18$) than in condition RO ($\eta^2 = 0.03$).

We also found that the novel variables proposed in Hæggström et al. (2006) did not allow for the clear discrimination between the “early” and the “late” bins. For instance, TA had no significant effect on variable CT in the normal condition, and increasing the task difficulty did not notably improve the quality of the discrimination (see Figure 3 and Table 1). TA had a significant effect on CT in condition RC and on FD in conditions RO and RC, but the effect sizes (respectively $\eta^2 = 0.02$, 0.03 and 0.02) were much smaller than that corresponding to variable AREA in condition FC ($\eta^2 = 0.18$).

**Figure 2.** Example of the stabilogram diffusion plot for the COP trajectory of one subject in the eyes open and rigid surface test conditions at the beginning of the experiment. The critical time variable is found at the intersection between two linear fits adjusted to the data as illustrated.
For a more detailed comparison of the variable AREA in condition FC and variable FD in condition RO (the novel variable from Hæggström et al. (2006) that performed the best in our data), we plotted the intra-bin variabilities and the inter-bin differences for each of the 14 subjects in Figure 4. One can note that for AREA/FC, the algebraic inter-bin differences (white bars) were positive for 13 out of the 14 subjects, meaning that AREA was, on average, higher in “late” trials than in “early” trials in 92.9% of the subjects (for the other variables/conditions, see the last column of Table 1). Moreover, for most subjects, the inter-bin differences were at the same level or even larger in many subjects than the intra-bin variabilities (black and grey bars).

In contrast, for FD/RO, 3 of the 14 subjects behaved differently from the majority. Additionally, for most subjects, the inter-bin differences were smaller than...
the intra-bin variabilities (see also the values of $\eta^2$ reported above).

**Protocol for evaluating time since awakening**

Based on the observations of the previous section, we here propose a practical protocol for determining whether TA is larger than a certain threshold $T_{\text{cut}}$.

By convention, if $TA < T_{\text{cut}}$, we will say that the test result is *negative* and if $TA > T_{\text{cut}}$, we will say that it is *positive*.

For simplicity, we describe and assess the robustness of the protocol for a single variable and a single condition (the most discriminating pair variable/condition according to the previous section: AREA/FC), but it is possible to combine multiple variables and multiple conditions to further improve the results (see Discussion).

**Description of the protocol**

- **Database construction phase**: We suppose that the *baseline* amount of sway of each subject is known to the experimenter. For

### Table 1. Effect of TA on some variables (ANOVA test). Note: The last column gives the percentage of subjects for whom the effect of TA was opposite to that of the majority of the subjects.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cond</th>
<th>Effect of TA</th>
<th>F-value</th>
<th>Effect size $\eta^2$</th>
<th>&quot;Dissenting&quot; subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREA</td>
<td>RO</td>
<td>signif.</td>
<td>17.8</td>
<td>0.03</td>
<td>21.4%</td>
</tr>
<tr>
<td>AREA</td>
<td>RC</td>
<td>signif.</td>
<td>93.4</td>
<td>0.12</td>
<td>0%</td>
</tr>
<tr>
<td>AREA</td>
<td>FO</td>
<td>signif.</td>
<td>113.3</td>
<td>0.15</td>
<td>0%</td>
</tr>
<tr>
<td>AREA</td>
<td>FC</td>
<td>signif.</td>
<td>170.8</td>
<td>0.18</td>
<td>7.1%</td>
</tr>
<tr>
<td>CT</td>
<td>RO</td>
<td>not signif.</td>
<td>6.3</td>
<td>0.01</td>
<td>28.6%</td>
</tr>
<tr>
<td>CT</td>
<td>RC</td>
<td>signif.</td>
<td>12.3</td>
<td>0.02</td>
<td>21.4%</td>
</tr>
<tr>
<td>CT</td>
<td>FO</td>
<td>not signif.</td>
<td>1.0</td>
<td>0.001</td>
<td>43.9%</td>
</tr>
<tr>
<td>CT</td>
<td>FC</td>
<td>not signif.</td>
<td>0.3</td>
<td>0.0003</td>
<td>43%</td>
</tr>
<tr>
<td>FD</td>
<td>RO</td>
<td>signif.</td>
<td>16.6</td>
<td>0.03</td>
<td>21.4%</td>
</tr>
<tr>
<td>FD</td>
<td>RC</td>
<td>signif.</td>
<td>12.4</td>
<td>0.02</td>
<td>21.4%</td>
</tr>
<tr>
<td>FD</td>
<td>FO</td>
<td>not signif.</td>
<td>3.7</td>
<td>0.007</td>
<td>35.7%</td>
</tr>
<tr>
<td>FD</td>
<td>FC</td>
<td>not signif.</td>
<td>1.4</td>
<td>0.003</td>
<td>50%</td>
</tr>
</tbody>
</table>

Figure 4. Subject-by-subject assessment of discrimination quality. Top: Variable AREA, condition FC. Bottom: Variable FD, condition RO.

The standard deviations of the variable in the "early" and "late" bins (intra-bin variabilities) were plotted in black and grey, respectively. The algebraic differences between the averages in the "early" and "late" bins (inter-bin differences) were plotted in white. Note that the bars were normalized by dividing by $c$, where $c$ is the average height of the black and grey bars across subjects. Note that there is one subject (HEAN) for whom the average AREA was smaller in the "late" than in the "early" bin.
example, this can be done by measuring the subject’s sway several times in normal awake conditions (e.g., when TA ≤ 10 h), computing and averaging the values of the variable across the trials, and finally, recording the average value µ in the subject’s database record (in a bus company, for example, this would be done upon hiring a new driver and the database would be updated regularly thereafter). This is what we meant by “limited prior knowledge of the subjects’ posturographic history” (see Introduction), as opposed to protocols that require posturographic data for an entire 36 h period.

• Test phase: In the test phase, the subject is measured N times in a row. The values of the variable are then computed and averaged across the N trials (for practical reasons, N should not exceed 3 or 4, but the larger the N, the more reliable the test). If the average value of the variable is larger than k times the value m in the database (in practice, the coefficient k will be determined according to a certain trade-off, see below), then the test is declared positive. Otherwise, it is declared negative.

Assessment of the robustness of the protocol

We statically assessed the robustness of this protocol based on our experimental data as follows.

• Database construction phase: The baseline value m was defined as the average value of the variable computed across the first 10 trials (5 data collection points with TA = 2 h … 10 h × 2 reps).

• Test phase: We considered the combinations of N trials taken from those with T_{INI} ≤ TA < T_{CUT}. For a given combination, if the average value of the variable computed across N trials of the combination is larger than km, then the combination is incorrectly labelled as positive. We call such a combination a false positive. The ratio of false positives over the total number of N-combinations is called the false positive rate (FPR). Similarly, we considered the combinations of N trials taken from those with T_{CUT} < TA ≤ T_{END}. For a given combination, if the average value of the variable computed across N trials of the combination is larger than km, then the combination is correctly labelled as positive. We call such a combination a true positive. The ratio of true positives over the total number of N-combinations is called the true positive rate (TPR).

One can observe that the protocol is robust if its FPR is low and its TPR is high. There is, however, a trade-off between these two quantities; for example, when the coefficient k increases, the FPR decreases while the TPR decreases. To assess the robustness of the protocol, we plotted the TPR versus the FPR (yielding the Receiver Operating Characteristic or ROC curve). The robustness of the protocol could then be quantified by the Area Under the ROC Curve (AUC), with AUC = 1 for perfect discrimination and AUC = 0.5 for indifferent discrimination (Su and Liu, 1993).

The ROC curves were plotted in Figure 5 for N = 2 (left plot) and N = 4 (right plot). For both plots, T_{INI} = 12 h, T_{CUT} = 21 h and T_{END} = 30 h, i.e.,

![Figure 5. Receiver Operating Characteristic (ROC) curves for our proposed protocol, which aims at discriminating trials with TA = 12 h … 20 h (negative trials, cf. the beginning of section Protocol…) from those with TA = 22 h … 30 h (positive trials). Left plot for N = 2 and right plot for N = 4 (see Protocol…). The individual ROC curves for the 14 subjects were plotted in thin lines. The values of the Area Under the average ROC Curve (AUC) are also given. The only curve below the non-discrimination line corresponds to subject HEAN (see also Figure 2).]
we wanted to discriminate trials with $TA = 12 \text{h} \ldots 20 \text{h}$ from those with $TA = 22 \text{h} \ldots 30 \text{h}$ (note that setting $T_{\text{ini}} = 12 \text{h}$ and $T_{\text{end}} = 30 \text{h}$ is more realistic and yields a harder challenge for the protocol than setting $T_{\text{ini}} = 2 \text{h}$ and $T_{\text{end}} = 36 \text{h}$ because the first trials tend to have a small AREA and the last trials tend to have a large AREA).

One can see that our protocol allowed very clear discrimination in trials with $TA = 12 \text{h} \ldots 20 \text{h}$ from those with $TA = 22 \text{h} \ldots 30 \text{h}$. For $N = 4$, the AUC of the average ROC curve was 0.89. For $N = 2$, the discrimination power was lower but still very good (AUC = 0.85). Note that all the individual ROC curves, except that of subject HEAN (cf. also Figure 3), were well above the no-discrimination line for both $N = 4$ and $N = 2$.

Depending on need, one may choose different operating regimes. For $N = 4$, if one needs a low average FPR, e.g., 10%, this corresponds to $k = 1.19$ and an average TPR across subjects of 81%. Alternatively, if one needs a high average TPR, e.g., 90%, this corresponds to $k = 1.10$ and an average FPR of 18% (these results were obtained excluding subject HEAN).

**Discussion**

By using the Foam Surface/Eyes Closed condition, we increased the difficulty of the quiet-standing task and thereby, were able to clearly discriminate between “early” ($TA < 21 \text{h}$) and “late” ($TA > 21 \text{h}$) trials. The variable we used (AREA) is directly related to the amount of sway displayed by the subject, and as such, is easy to compute and interpret in physiological terms. The reliability of the results was particularly high; 13 out of 14 subjects (or 92.9%) had a higher average AREA in the “late” than in the “early” trials in condition FC (and even 100% in conditions RC and FO). This variable (in the experimental conditions we considered) was more reliable than the novel variables proposed in the literature (Hæggström et al., 2006).

Based on the experimental results, we proposed a protocol to determine whether $TA > 21 \text{h}$. The robustness of this protocol was statistically demonstrated on a subject-by-subject basis. We insist on the practical character of this protocol; the database construction phase requires only 10 trials in the normal awake condition ($TA \leq 10 \text{h}$), and the test phase may include as little as two 40 s trials.

A limitation of this study is its small sample size. This sample size is due to difficulty recruiting subjects for complex data collection which requires the subjects to stay awake for 36 hours. However, the proposed protocol is based on a subject-by-subject analysis, and we are confident that the present results are robust.

Our experiments and protocol can be further improved in several ways. First, combining multiple variables and conditions (e.g., AREA/FC, AREA/FO and FD/RO) using standard statistical tools, such as Linear Discriminant Analysis (Su and Liu, 1993) or multi-dimensional classification algorithms, should improve the discrimination quality. Second, assuming a higher discrimination quality, it should be possible to use a dichotomy search to determine $TA$ with higher precision. Third, taking into account circadian rhythm in our experiments (for instance, by awakening subjects at times of day other than 7 a.m.) may lead to a more precise understanding of the effects of $TA$. Finally, our protocol needs to be tested in a real-life scenario.

**Acknowledgments**

We would like to thank the São Paulo Research Foundation (FAPESP), Research, Innovation and Dissemination Centers (CEPID-FAPESP) and the National Council for Scientific and Technological Development (CNPq) for financial support for this research. We would also like to thank the Multidisciplinary Center for the Study of Drowsiness and Accidents (CEMSA), the Center for Psychobiology and Exercise Studies (CEPE), and the Research Incentive Fund Association (AFIP) for their partnership.

**References**


Hallvig D, Anund A, Fors C, Kecklund G, Akerstedt T. Real driving at night: predicting lane departures from physiological...


Authors

Quang-Cuong Pham, Marcos Duarte*
Biomedical Engineering, Universidade Federal do ABC – UFABC, Av. dos Estados, 5001, CEP 09210-580, Santo André, SP, Brazil.

Marco Túlio de Mello, Fernanda Veruska Narciso, Marcos Mônico Neto, Cristiane Westin Teixeira, Leandro Stetner Antonietti, Silvio de Araújo Fernandes Júnior, Renata Guedes Koyama, Sérgio Tufik
Centro Multidisciplinar em Sonolência e Acidentes, São Paulo, Brazil.

Marco Túlio de Mello, Sérgio Tufik
Centro de Estudos em Psicobiologia e Exercício, São Paulo, Brazil.

Silvério Aparecido Garbuio, Sérgio Tufik
Instituto do Sono, São Paulo, Brazil.

Cristiane Westin Teixeira, Leandro Stetner Antonietti, Silvério Aparecido Garbuio, Renata Guedes Koyama, Sérgio Tufik
Associação Fundo de Incentivo a Pesquisa, São Paulo, Brazil.

Marco Túlio de Mello, Fernanda Veruska Narciso, Marcos Mônico Neto, Silvio de Araújo Fernandes Júnior, Sérgio Tufik
Universidade Federal de São Paulo, São Paulo, Brazil.