Recognition of Sleep Dependent Memory Consolidation with Multi-modal Sensor Data

Akane Sano

Massachusetts Institute of Technology
Media Lab
Cambridge, MA, USA
akanes@media.mit.edu

Rosalind W. Picard

Massachusetts Institute of Technology
Media Lab
Cambridge, MA, USA
picard@media.mit.edu

Abstract—**This paper presents the possibility of recognizing sleep dependent memory consolidation using multi-modal sensor data. We collected visual discrimination task (VDT) performance before and after sleep at laboratory, hospital and home for N=24 participants while recording EEG (electroencepharogram), EDA (electrodermal activity) and ACC (accelerometer) or actigraphy data during sleep. We extracted features and applied machine learning techniques (discriminant analysis, support vector machine and k-nearest neighbor) from the sleep data to classify whether the participants showed improvement in the memory task. Our results showed 60-70% accuracy in a binary classification of task performance using EDA or EDA+ACC features, which provided an improvement over the more traditional use of sleep stages (the percentages of slow wave sleep (SWS) in the 1st quarter and rapid eye movement (REM) in the 4th quarter of the night) to predict VDT improvement.**

Keywords—memory consolidation; visual discrimination task (VDT), EEG, EDA, actigraphy, classification, galvanic skin response, GSR

I. INTRODUCTION

Past studies have shown that sleep can enhance memory consolidation. Some studies have shown the relation to REM sleep [1] [2]. Stickgold et al. showed that consistent and significant performance improvement on a Visual Discrimination Task (VDT) became proportional to the amount of sleep in excess of six hours, and subjects with an average of eight hours then exhibited a correlation in performance to the sleep stages: SWS (Slow Wave Sleep) in the first quarter of the night, and REM (Rapid Eye Movements) in the last quarter [3]. To our knowledge, no prior studies have attempted to classify whether sleep-dependent memory consolidation occurred by using automated analysis of sensor data during sleep. This paper examines whether EEG (electroencephalogram), EDA (electrodermal activity) and Actigraphy data during sleep can predict task performance improvement on the VDT.

II. METHODS

A. Measurement

Twenty-four healthy university students (ages 18-22, 16 males) participated in 3 nights of measurements, one night in a "homey" sleep laboratory, one night in a hospital sleep lab, and one night at home to measure physiological changes

related to task performance on a VDT, measured before and after sleep. The physiological measures consisted of EDA (a measure of sympathetic nervous system activity), skin temperature and actigraphy (all three measured from the wrist at 32Hz) each night and EEG (C3, C4, O1 and O2 under international 10-20 method, 100Hz) for the nights in the homey and hospital sleep labs. The Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects (COUHES) approved both studies.

Each night (PM), participants trained on a different version of the VDT, slept, and were tested the next morning (AM). Sleep in the homey and hospital sleep labs was also monitored with standard PSG (polysomnography), consisting of 30-second epochs of sleep stages labeled by experts (Wake, REM, Non-REM 1, 2 and SWS).

We evaluated task improvement by overnight change (PM-AM) in VDT performance (a lower score is better performance). We obtained standard PSG sleep staging as well as subjective sleep quality evaluations on a scale of 1 to 4. Out of all data, only 15 nights of data from 15 participants (10 males) in the hospital had accurately time-synched EDA data with concurrent PSG, so while we can analyze EDA and actigraphy for all nights, we can analyze its relationship to PSG sleep stage information for only 15 nights.
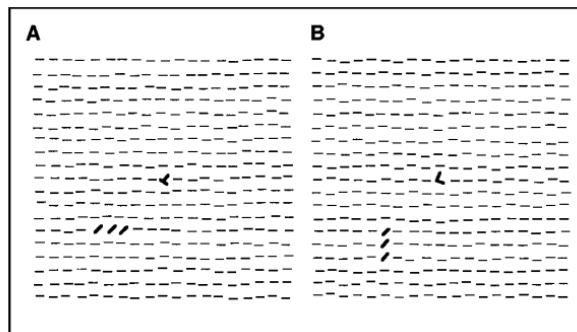

Fig. 1. Sample target screens for the visual discrimination task.

The VDT was a task that has been used routinely in sleep studies by Stickgold et al. (2000). Each target screen contained a rotated "T" (A) and "L" (B) at the fixation point

and a horizontal (A) or vertical (B) array of three diagonal bars in the lower-left quadrant of the visual field.(Reprinted with permission from (Stickgold et al., 2000)

Figure 1 shows one trial of the VDT method, consisting of 5 screens that appear in typically less than 1 second. The first screen is a fixation screen (black with a white centered crosshair) that remains until the participant hits a key. This is followed by a 16-ms target screen, a 0-400-ms blank "interstimulus interval" (ISI), and then a 16-ms mask screen. The ISI varies over the course of the task, starting at 400 ms and is progressively shortened to 0ms over the 25 blocks of 50 trials. Each participant was asked to determine two features of the target screen: whether the capital letter in the center of the screen was "T" or "L" and whether an array of three diagonal bars in one quadrant of the screen is horizontal or vertical. By interpolating the ISI at which 80% accuracy is achieved on the horizontal versus vertical decision, a 'threshold' ISI (in ms) was extracted from each session of the VDT. A lower threshold is a better performance. Overnight improvement or deterioration on the VDT was then calculated as a subtraction of the AM VDT threshold from the PM threshold. For example, 30 ms indicates the person performed 30 ms better in the AM than in the PM, while a negative value signals deterioration from PM to AM. One VDT session consisted of 25 blocks with 50 trials in each.

B.   Feature Extraction

Figure 2 shows a sample representation of one night's data from one participant when the PSG was fully synchronized.

We computed the following features for building a machine learning classifier.

a) EEG

We calculated power spectrum density of the frequency band (delta, theta, alpha and beta) of the quarters of the night for electrode locations C3 and C4.   We also computed the features using the average amplitude at electrodes C3 and C4 over the whole night and per epoch.

b) EDA

The EDA was processed first by low-pass filtering (cutoff frequency 0.4 Hz, 32nd order FIR filter) before computing the features. We normalized the amplitude of the EDA by dividing all values by the maximum amplitude over the night, then obtained the first derivative of the filtered EDA, then determined where the slope exceeds a value of 0.5 micro Siemens per second. We detected EDA "peaks" based on those that exceeded this $0.5\mu S/s$ threshold and counted the number of peaks per each 30-second epoch.   We also computed the mean, standard deviation, median of the normalized EDA amplitude (normalized by the maximum EDA amplitude over the night) before sleep and during sleep. For EDA peaks, we computed the total number for the night, the mean, standard deviation, and median of the number of EDA peaks per 30 s epoch over the night, the averaged number of peaks, the % of epochs with EDA peaks for each sleep stage and the mean # of EDA peaks per epoch.
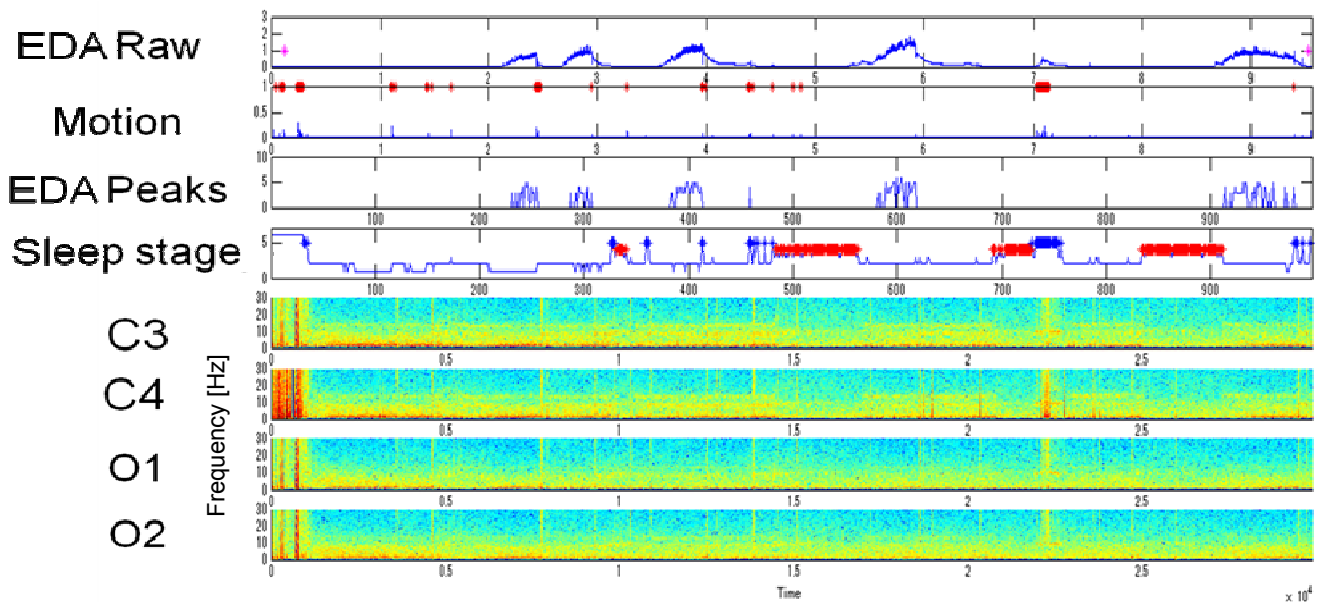

Fig.2 Raw EDA, Actigraphy (derived from 3-axis accelerometer data, red marks means wakefulness), detected regions with EDA peaks, manually scored sleep stages from EEG(red marks mean REM sleep), and four EEG channels for one night measured from a healthy adult.

Previously, we have shown that EDA peaks are much more likely to occur during SWS or Non-REM 2 sleep [4].

EDA data that corresponded to non-sleep epochs (as determined by actigraphy, see below) were removed from the analysis before computing features related to sleep.

c) Actigraphy or Accelerometer (ACC)

Sleep and non-sleep epochs were determined using standard zero-crossing detection and Cole's function applied to the accelerometer data [5]. From this motion information, we further computed sleep latency, sleep duration, the % of wake in each quarter of the night, and the mean and standard deviation of the motion level.

d) Sleep stages

The sleep stages were scored by standard criteria [6]. The features we used included the % of each sleep stage over the night, the sleep efficiency derived from the EEG (the percentage except wake and others during sleep), the time to first deep sleep, and the percentage of each sleep stage for each quarter of the night

C. Classifiction

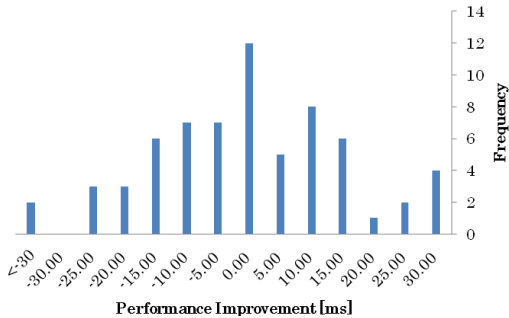Fig. 3 shows the distribution of performance improvement across the 24 participants x 3 nights = 72 nights.

Fig.3 VDT performance improvement (N=24, 3 nights)

We grouped the participants into the following three groups.

1) the highest and lowest 33% of VDT improvement
2) the highest and lowest 20% of VDT improvement
3) the highest and lowest 20 % of VDT improvement only in hospital and laboratory nights (because we have PSG only for those nights).

For each of the three groups of data, we compared 6 methods:
A) Support vector machine with linear kernal
B) Support vector machine with Gaussian kernel
C) PCA and linear discriminant analysis
D) PCA and support vector machine with linear kernal
E) PCA and support vector machine with Gaussian kernel
F) PCA and k nearest neighbors (k=1-5)

Each method was run with 4 variations of features (a-d) for the nights of hospital, laboratory and home and run with 7 variations of features (a-g) for only the nights of hospital and laboratory,

a) All features
b) EDA features only
c) ACC features only
d) EDA + ACC features
e) EEG features only
f) Sleep stage features only
g) EEG + EDA + ACC features

and compared classification accuracy with the 10-fold cross validation. In this method, we trained the model with 90% of the data, tested with the remaining 10% and repeated this procedure 10 times, each time leaving out a different 10% of the data.

III. RESULTS

As a baseline, because of prior published findings on how SWS and REM interact with VDT improvement, we examined the classification result using the percentage of SWS in the 1st quarter and percentage of REM in the 4th quarter of the night (Fig. 4). The accuracy of this published method for predicting VDT performance is mostly around 60% or below it.
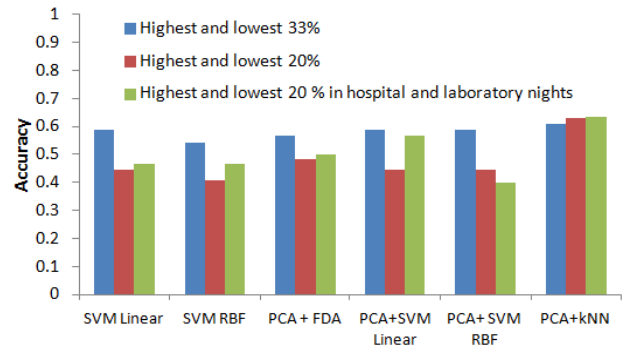

Fig.4.Accuracy of classification of VDT performance using the percentage of SWS in the 1st quarter and REM in the 4th quarter of the night

Figure 5 shows the comparison of classification accuracy using new physiological features and our approach using classification methods for the highest and lowest 33% of VDT performance improvement. The features from EDA alone showed the highest accuracy, around 60-70% beating every one of the other three feature set combinations, while being tested on all six machine learning systems.
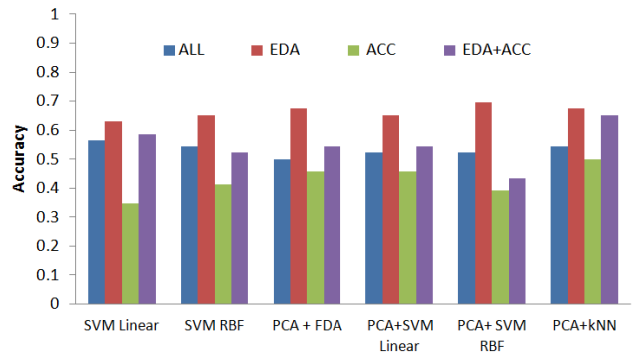
Fig.5.Accuracy of classification using the highest and lowest 33% of VDT performance improvement

Figure 6 shows the comparison of classification accuracy with our new machine learning approach to this problem for the highest and lowest 20% of VDT performance improvement. The features from EDA again showed the highest accuracy, 74%, this time either by appearing solo as the top performer or in three cases appearing in combination with ACC data.
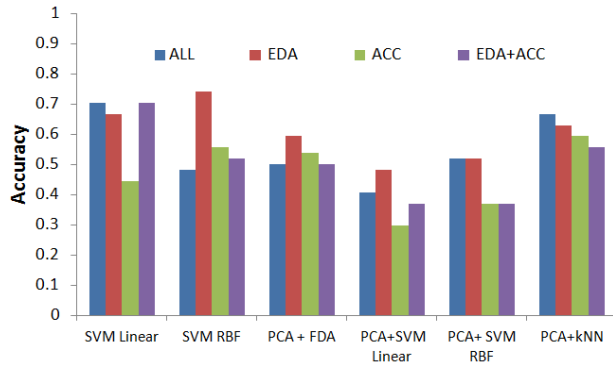

Fig.6 Accuracy of classification using the highest and lowest 20% of VDT performance improvement
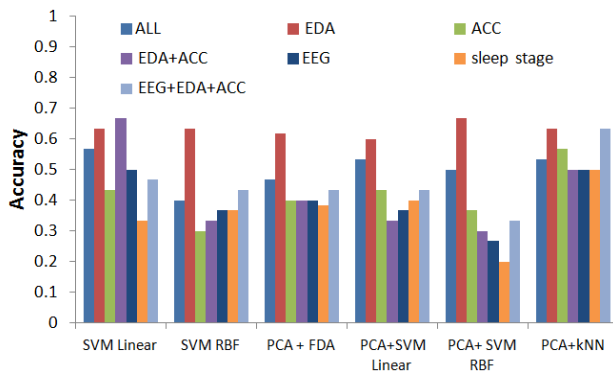

Fig.7 Accuracy of classification using the highest and lowest 20 % of VDT performance improvement only in hospital and laboratory nights

Figure 7 shows the comparison of classification accuracy with features and classification methods for the highest and lowest 20% of VDT performance improvement for the nights where we also had PSG at the hospital and laboratory. The features from EDA solo or EDA + ACC showed the highest accuracy, 67%, followed by EEG + EDA + ACC. Thus, EDA was again a part of all of the top performing sets of features.

In all the comparisons here, overall, either solo EDA features or EDA + ACC features improved the classification accuracy compared to use of sleep stages and to use of only EEG. Why would EDA changes play a significant role in prediction of performance on a VDT task? This cannot be fully explained by the finding that EDA activity can show some signs of SWS or NREM2 sleep. While this question remains unanswered, this phenomenon is intriguing to look into more, in part because it's known that EDA during wake is increased with greater engagement and arousal, which then is believed to help predict memory [7]. Now we also see that for sleep, the use of electrodermal physiology is showing patterns that invite greater exploration.

REFERENCES

[1] J. M. Siegel, "The REM sleep-memory consolidation hypothesis.," *Science (New York, N.Y.)*, vol. 294, no. 5544, pp. 1058–63, Nov. 2001.

[2] A. Karni, D. Tanne, B. S. Rubenstein, J. J. Askenasy, and D. Sagi, "Dependence on REM sleep of overnight improvement of a perceptual skill.," *Science (New York, N.Y.)*, vol. 265, no. 5172, pp. 679–82, Jul. 1994.

[3] R. Stickgold, D. Whidbee, B. Schirmer, V. Patel, and J. A. Hobson, "Visual discrimination task improvement: A multi-step process occurring during sleep.," *Journal of cognitive neuroscience*, vol. 12, no. 2, pp. 246–54, Mar. 2000.

[4] A. Sano and R. W. Picard, "Toward a taxonomy of autonomic sleep patterns with electrodermal activity.," *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2011, pp. 777–80, Jan. 2011.

[5] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity.," *Sleep*, vol. 15, no. 5, pp. 461–9, Oct. 1992.

[6] A. Rechtschaffen,A.A. Kales. "A Manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects*".* Washington, D.C: U.S. Government Printing Office; 1968.

[7] W. Boucsein, "Electrodermal Activity (The Springer Series in Behavioral Psychophysiology and Medicine)," Springer, 2012