Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
J Intell Transp Syst. Author manuscript; available in PMC 2018 October 17.

Published in final edited form as:
J Intell Transp Syst. 2017 ; 21(5): 422–434. doi:10.1080/15472450.2017.1369060.

Using kinematic driving data to detect sleep apnea treatment adherence

Anthony D. McDonald[a], **John D. Lee**[b], **Nazan S. Aksan**[c], **Jeffrey D. Dawson**[c], **Jon Tippin**[c], and **Matthew Rizzo**[d]

[a]Texas A&M University, College Station, TX, USA

[b]University of Wisconsin-Madison, Madison, WI, USA

[c]The University of Iowa, Iowa City, IA, USA

[d]University of Nebraska Medical Center, Omaha, NE, USA

Abstract

People spend a significant amount of time behind the wheel of a car. Recent advances in data collection facilitate continuously monitoring this behavior. Previous work demonstrates the importance of this data in driving safety but does not extended beyond the driving domain. One potential extension of this data is to identify driver states related to health conditions such as obstructive sleep apnea (OSA). We collected driving data and medication adherence from a sample of 75 OSA patients over 3.5 months. We converted speed and acceleration behaviors to symbols using symbolic aggregate approximation and converted these symbols to pattern frequencies using a sliding window. The resulting frequency data was matched with treatment adherence information. A random forest model was trained on the data and evaluated using a held-aside test dataset. The random forest model detects lapses in treatment adherence. An assessment of variable importance suggests that the important patterns of driving in classification correspond to route decisions and patterns that may be associated with drowsy driving. The success of this approach suggests driving data may be valuable for evaluating new treatments, analyzing side effects of medications, and that the approach may benefit other drowsiness detection algorithms.

Keywords

driving; drowsiness; machine learning; sleep disorders; symbolic aggregate approximation

Introduction

American drivers spend nearly 20 hours in their cars every week (Williams, 2009). Throughout these hours, drivers must maintain vehicle control and their safety through deliberate interaction with a steering wheel and pedals. In addition to these exchanges, drivers frequently interact with safety devices and entertainment options. Together these

CONTACT Anthony D. McDonald McDonald@tamu.edu 4075 Emerging Technologies Building, 3131 TAMU, 101 Bizzell Street, College Station, TX, 77843, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/gits.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interactions represent a rich behavioral dataset. Recent advancements in on-road Data Acquisition Systems (DAS) have made wide-scale collection of this behavior data possible (Shrestha, Lovell, & Tripodis, 2017). The 100-car Naturalistic Driving Study demonstrated that such behavioral data has significant potential to enhance understanding of vehicle crashes and their causes (Dingus et al., 2006). Given these insights, it is natural to wonder if the reach of naturalistic driving data extends beyond the connection between driver actions and driving performance to driver state.

The problem of driver state detection is often viewed in the context of designing driver support systems (Doshi, Morris, & Trivedi, 2011; McCall, Wipf, Trivedi, & Rao, 2007; Raksincharoensak et al., 2010; Rigas, Goletsis, Bougia, & Fotiadis, 2011; Rosenfeld et al., 2015). Yet, driver states often extend beyond the vehicle cockpit, particularly medical conditions like obstructive sleep apnea (OSA). OSA is a chronic disorder associated with recurrent obstructive respiratory events and fragmented sleep that results in excessive daytime sleepiness and increased cardiovascular morbidity and mortality. OSA affects approximately 15% of middle-aged adults (Young et al., 2009) and is particularly pervasive in professional drivers (Moreno et al., 2004). A recent large meta-analysis showed that OSA drivers have a mean crash risk ratio of 2.72, having a 172% greater chance of a crash relative to the general population (Sassani et al., 2004; Stephen Tregear, Reston, Schoelles, & Phillips, 2009).

Despite its severity, OSA is a treatable condition. Continuous positive airway pressure (CPAP) is the treatment of choice for patients with clinically significant OSA (Kushida, Littner, & Hirshkowitz, 2006). Proper use of CPAP reduces crash risk and improves simulated driving performance in as little as two weeks of treatment (Barbé et al., 2007; Findley, Smith, Hooper, Dineen, & Suratt, 2000; Orth et al., 2005; Turkington, Sircar, Saralaya, & Elliott, 2004). Specifically, Tregear, Reston, Schoelles, and Phillips (2010) found that crash risk declines by 72% with CPAP. Unfortunately, CPAP treatment is often uncomfortable leading to reduced acceptance and adherence to CPAP therapy in many patients (Kribbs et al., 1993). In one study, only 72% of patients who started on CPAP agreed to continue it after the first night (Rauscher, Popp, Wanke, & Zwick, 1991). Patterns of use vary widely in patients who accept CPAP; about half use it an average of 6 hours in 90% of nights, while the rest use it less than 4 hours in 2–79% of nights (Weaver et al., 1997).

The connection between untreated sleep apnea and driving performance and the high likelihood of CPAP abandonment represent a promising opportunity to explore the use of driving behavioral data to address health problems outside the vehicle. This study addresses the opportunity by using driving data to detect CPAP adherence. The study describes a data reduction technique, symbolic aggregate approximation (SAX) (Lin, Keogh, Wei, & Lonardi, 2007), which can be used to convert the task of CPAP adherence prediction to a supervised learning problem. The work will illustrate a potential solution to this supervised learning problem, demonstrate the effectiveness of the approach, and assess the sensitivity of parameter selection.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Methods for driving data reduction

Naturalistic driving data provides a unique opportunity to observe driver behavior beyond simulator studies and controlled on-road experimentation. Unfortunately, the collection and analysis of naturalistic driving data is difficult. From a data perspective, there are two central issues with naturalistic driving data: the size and the structure. The size of a naturalistic dataset is problematic because it makes visualization and comparison between multiple drives difficult. The structure of driving data is problematic because it is not amenable in its raw form to traditional statistical models such as Linear Regression. These problems lead naturalistic driving researchers to focus their analysis on safety critical events and the short-time windows surrounding them (Dingus et al., 2006; Gordon et al., 2011; Guo, Klauer, Hankey, & Dingus, 2010). This approach creates a potential omitted variable bias in subsequent analyses and it ignores meaningful data (Jovanis, Aguero-Valverde, Wu, & Shankar, 2011). When driving data are considered over a longer time horizon, data are typically reduced using simple aggregating functions such as the mean and standard deviation calculated over a window of time (Dozza, Bärgman, & Lee, 2013). To balance the goals of preserving meaningful data and addressing “Big Data” problems, researchers might consider more advanced time-series data reduction techniques (Das, Zhou, & Lee, 2012).

Time-series data reduction methods

Time-series data reduction is accomplished by shifting perspectives from viewing the data as a series of observations to a sum of simple functions. In this new perspective, the key intuition is that a small subset of this sum of simple functions can represent the majority of the information of the time series. Many simple functions have been considered for time-series data decomposition including sine waves, square waves, and constant values. These three simple functions represent the three most common methods of time-series data reduction: Discrete Fourier Transforms (DFTs), Discrete Wavelet Transforms (DWTs), and Piecewise Aggregate Approximation (PAA).

DFTs represent a time series as a sum of sine or cosine functions (Agrawal, Faloutsos, & Swami, 1993). DFTs are popular in the driving impairment detection literature because they are simple to apply and able to filter out noise in complex signals such as steering behavior (Das et al., 2012; Jap, Lal, Fischer, & Bekiaris, 2009; Krajewski, Sommer, Trutschel, Edwards, & Golz, 2009; Lal, Craig, Boord, Kirkup, & Nguyen, 2003). DFTs are limited for two reasons: they assume stationarity and are not localized in time. Stationarity refers to the fact that the distribution of the time series does not change over time. This assumption causes DFTs to perform poorly in the reduction of time series like stock prices that have a positive or negative trend in their mean. Time localization refers to the fact that DFT components are fit to an entire signal at once. This broad fitting causes DFTs to struggle when capturing emergent effects. DWTs, specifically Haar wavelet transforms, represent time series as a sum of square wave functions. They are advantageous relative to DFT because they are localized in time and can represent emergent phenomenon that arise and dissipate in the middle of a time series (Chan & Fu, 1999; Keogh, Chakrabarti, Pazzani, & Mehrotra, 2001). The primary disadvantage of DWTs is that they are only defined for time series of lengths equivalent to integral powers of 2 (i.e. 2^n). PAA represents a time series as a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

series of means or constant values. This conversion is accomplished by dividing the time series into equal sized windows, taking the mean of the samples within each window, and then creating a reduced representation of the original time series by simply concatenating the means into a vector (Keogh et al., 2001). Despite its simplicity, PAA produces the same reduction results as DWT with optimal parameter selection and appropriate length. This parameter selection is the primary limitation of PAA. The method requires setting two parameters: the window size and the percentage of overlap between the windows, which introduces a tradeoff between replication of the original signal and data reduction.

Each of the three primary time-series data reduction methods have been modified to extend their capacity, however these extensions do not fundamentally alter their performance. One exception is SAX. SAX is a modification to PAA that converts the output of PAA from a series of constants to a series of letters. This conversion is accomplished by adding additional steps to PAA, which bin the *y*-axis into quantiles of the normal distribution, assign a letter to each quantile, and then convert the PAA output into letters based on each mean's associated *y*-axis quantile. This process is illustrated in Figure 1 for a sample of data.

The output of SAX is a series of letters that correspond to quantiles of the normal distribution ("dfghhhhhii" in Figure 1). Unfortunately, the addition of normalization and the symbolic conversion create two additional parameters, relative to PAA, which are subject to optimization. These parameters are global versus local normalization and the size of the alphabet to use.

Finding the preferred reduction method

Given this set of reduction methods, it is natural to compare them and find a clear optimal solution for naturalistic data reduction. Surprisingly, this process is quite difficult. The difficulty arises because the criteria for evaluation are ill-defined and all conclusions are highly data dependent. This latter point is demonstrated by Keogh and Kasetty who show that different outcomes of the comparison between DFT and DWT can be achieved by using different datasets (Keogh & Kasetty, 2003). The data dependency issue is especially complex in the driving domain because the data are quite diverse. For example, Figure 2 shows a sample of speed and lateral acceleration data for a single drive. In the figure, there are several clear differences between the two data sources namely, the acceleration data is centered about zero and is significantly noisier than the speed data. These differences suggest that even within driving data, conclusions for a single variable may not generalize to others.

The data dependency and lack of clear evaluation metrics suggest that a data reduction method should be selected within the goals of a particular study. In this work, our broad goal is to develop predictive models of CPAP adherence using reduced driving data as input. The most useful data reduction method in this case would reduce the data while preserving the information most critical to discern between CPAP adherence and abandonment. This requirement suggests that SAX should be preferred over PAA (and by extension DWT) because the reduction performance is relatively similar and the symbolic output of SAX amenable to a larger range of data mining methods compared to the continuous PAA output (Lin et al., 2007). An additional consideration in this case is the feasibility of an

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

abandonment detection model in a real-world system. Such a system would most likely take daily or hourly driving data, reduce these data, apply the prediction model, and produce a prediction of adherence. Ideally, the prediction would be supported by objective measures that are easily interpretable for healthcare providers as well as individuals without access to medical data such as occupation supervisors and fleet managers. The requirement of simple interpretability suggests that SAX should be preferred to DFT because DFT reduces the data to a set of phases and magnitudes. These phases and magnitudes—describe behavior in the frequency domain rather than time domain—and are not easily related to driving behavior. In contrast, patterns of SAX data retain the shape of the time series and thus are easier to relate to behavior. Thus, this paper utilizes SAX as the data reduction method component of the CPAP adherence detection model.

Reducing driving data with SAX

The goal of this work is to develop a predictive model of CPAP adherence using driving data as input. The previous section suggests that SAX should be a preferred method for this type of data reduction. The next step in the development process is to apply SAX to driving data. This section describes data used in this work, the parameter settings of SAX, and the application of SAX to the data.

Naturalistic driving study data

The data considered are from an evaluation of CPAP therapy. One hundred and four participants completed the 3.5-month data collection process. The participants represent two groups: participants diagnosed with OSA and healthy control participants. The work here focuses solely on the participants with OSA. These patients represent 69 of the 104 drivers, 47 of which are male with a mean age of 46 (SD = 7.7). Drivers were compensated for their time and effort at the completion of data collection. A summary of the demographic and anthropometric data from the participants is shown in Table 1.

An in-vehicle data acquisition system (IV-DAS) was installed in each participant’s personal vehicle to record GPS information, speed, three-axis accelerometer, and accelerator input at 10 Hz. The IV-DAS also collects intermittent video of the driver’s face and the road. Data were collected for a period of two weeks prior to OSA participants receiving CPAP therapy and for 3 months following the start of CPAP therapy. The data were partitioned into individual drive files defined by ignition engagement and disengagement. In addition to driving data, participants’ CPAP usage was monitored daily. The CPAP data included nightly usage and several sleep quality measures.

The present analyses focus on CPAP usage, speed, and accelerometer data. Speed and accelerometer data were selected because they are the most robust data in the dataset, are inexpensive and unobtrusive to collect, and are likely diagnostic of the main crash and near crash types associated with drowsy driving: lane departures and rear-end collisions (MacLean, Davies, & Thiele, 2003). Lane departures are often preceded by a period of inactivity and followed by a large lateral acceleration as a driver attempts to avoid crashing. Rear-end collisions are often preceded by large changes in longitudinal acceleration and speed due to drivers applying the brake pedal strongly to avoid crashing. Alert drivers should

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

show more gradual deceleration and fewer peak acceleration events. In order to capture these phenomena, acceleration data were analyzed in three ways. Lateral and longitudinal acceleration measures were included separately in the data and then combined together in a vector sum (combined acceleration).

Setting the parameters of SAX

One of the limitations of SAX is that it introduces a number of free parameters into the reduction process. The size of the window, percentage of overlap between windows, alphabet size, and normalization are all adjustable parameters. The window size can range from the maximum resolution of the data to the duration of the entire dataset. The amount of overlap between windows can vary between 0 and 100%. The alphabet size defines the resolution of binning. The normalization can either be completed globally or locally within each window. The means and standard deviations for global normalization can be calculated across as entire sample or within groups of the sample such as a mean for each participant. Global normalization makes each SAX letter have consistent meeting, which facilitates comparisons between signals where amplitude has significant meaning such as speed. Local normalization preserves the shape of the signal and allows for comparison between shapes regardless of their location and amplitude.

The selection of these parameters is largely a subjective process although settings can be objectively evaluated through a sensitivity analysis. The parameters selected in this case were adapted from McDonald et al. (2012) and McLaurin et al. (2014). The parameters used in this case were window size of 3 s, no overlap between windows, and global normalization across the entire dataset. The window size was the minimum resolution that was robust to data dropouts from the recording device. The non-overlapping windows and global normalization correspond with McDonald et al. (2012). Alphabet size for acceleration was set to 4, which aligns with McLaurin et al. (2014). Alphabet size for speed was determined through a sensitivity analysis that compared a range of alphabet sizes from 9 to 23 letters.

Application of SAX

SAX time-series analysis was applied separately to speed, lateral acceleration, longitudinal acceleration, and the combined acceleration measure. The result of the four applications of SAX was a set of four words corresponding to speed, acceleration, lateral acceleration, and longitudinal acceleration, which characterized each drive. These words were merged with the CPAP usage from the previous evening thus forming a reduced dataset containing the date of each drive, the CPAP usage, and the set of four words describing the drive.

Model development

After the application of SAX, the problem has been reduced to an analysis of symbolic sequences. In any symbolic sequence analysis, the goal is to find a way of measuring the similarities between sequences. The simplest way to accomplish this goal is to compare two sequences directly and assess their differences based on each letter. Unfortunately, this technique is not amenable to sequences of different lengths and offsets. For example, consider two drivers who live next door to one another driving to a traffic signal at the end of

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

their street. One driver will have to drive for an extra few seconds before she reaches the other driver's home. After this point, the drives might be exactly equivalent. If compared directly, the drives may look very different because of this small offset. A more informed approach would break the drive into smaller sequences and analyze their frequencies in the respective datasets (Leslie, Eskin, & Noble, 2002). In the previous example, this conversion makes the two drives very similar except for the frequency of the sequence that defines the drive between the two houses.

One underlying assumption of the example above is that the duration of each meaningful sequence is apparent in the data. In the OSA case, this assumption fails because specific patterns of data have not been associated with CPAP use or abandonment. In order to compensate for this, a range of sequence lengths, from one to ten letters, and their respective frequencies were examined for each of the four words. The result of this process is a dataset that contains columns for the date of the drive, the CPAP use in minutes, and one column for each sequence observed in the data.

Figure 3 demonstrates the process of creating this dataset for a sequence length of 3. At each step of the process, a sliding window is moved one letter to the right. If the sequence in the window is already in the dataset, then the frequency of that sequence is incremented by one. If the sequence has not been observed, it is added to the dataset with a frequency of one.

Additional data processing

The sequence frequency dataset contained one row for each of the 28,972 drives in the data, labeled with the CPAP usage from the previous night. This arrangement is somewhat problematic because often multiple drives occurred in one day for each driver and therefore the drive rows are not consistent with the daily CPAP use labels. To compensate for this differential, the sequence frequency data was aggregated to the day level. Thus, if the sequence "Speed_aaa" was observed once in three different drives on a given day, then the aggregate data would show "Speed_aaa" with a frequency of three.

Each row of aggregated data was labeled with a continuous measure of CPAP use during the previous evening specified in minutes. This continuous specification does not integrate well with most machine learning algorithms, which perform best with binary labels. Therefore, it was necessary to discretize the CPAP use measure into two levels, use and lack of use. The discretization consists of defining a threshold for each condition and then simply relabeling the data according to the thresholds. In this case, two thresholds were applied: more than 240 minutes (4 hours) of use and 0 minutes of use, respectively. Nights when participants used their CPAP, but did so for less than 240 minutes were removed from the dataset. These limits were based on the clinical cutoff for acceptable adherence (Sawyer et al., 2011). The goal of the bounds and data removal was to create a clear separation between the adherence and non-adherence that ensured nights classified as positive CPAP use complied with recommended CPAP practices. After this reduction and the day aggregation, the final dataset contained 4,038 instances (days), corresponding to 1,838 days where participants used CPAP and 2,200 when they did not. The instances included data from 22,870 drives, approximately 80% of the original dataset. The data were further partitioned into training and testing datasets so that the generalizability of the algorithm could be assessed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Generalizability in this context is an assessment of how the algorithm would perform on new data. The training and testing data were separated by driver (i.e. drivers appeared either in the training or testing dataset, but not both). This separation prevents the algorithm from making classifications based on driver habits or other factors unrelated to CPAP adherence. The training set contained approximately 90% of the drivers (62 drivers) and the testing set contained the remaining 10%. The testing set was withheld from all feature (measure) and algorithm selection decision-making processes.

Feature reduction

The feature creation process produces many irrelevant features. For example, an unexpected and rare occurrence such as a squirrel running into the road may cause a unique speed pattern in one of the driving days. Since the pattern is unique, it will be associated with CPAP use or disuse alone, and may be used to inaccurately differentiate between the two labels by an algorithm trained on the data. Other uninformative features are generated by common events such as stops at traffic signals, which will occur frequently regardless of CPAP adherence. These features are a hindrance to learning and cost processing time. In order to compensate for these two limitations sparse features, those found in less than 1% of days, were removed from the data and then an information-gain approach to feature selection was applied to the remaining data.

Information-gain-based feature selection involves ranking features by the amount they reduce entropy in the original data labels and then selecting the best *n*-features (Yang & Pedersen, 1997). Entropy reduction can be envisioned as a process where knowledge of certain information affects conclusions drawn based on other information. For example, the probability of a particular result in a die roll may be uncertain, but knowing that the die has the same number on all sides would reduce this uncertainty to 0. In contrast, knowing that the die is red or blue would provide little insight and thus no information gain. In this context, informative features represent patterns of driving associated CPAP adherence or CPAP non-adherence.

The 500 most informative features were retained. The final algorithm training dataset contained 3,757 days (2,103 days of CPAP use) each described by 500 different features corresponding to segments observed in speed, lateral acceleration, longitudinal acceleration, and combined acceleration lasting between 3 and 30 seconds. An example of the complete data reduction and feature generation process is shown in Figure 4.

Modeling results

This work sought to explore the feasibility of detecting CPAP lapses with driving data using an algorithm that provides explanatory power for the underlying phenomenon of OSA induced drowsy driving. These two goals are evaluated separately below.

Model training

The primary goal of this work is to create an algorithm that can identify CPAP usage from driving data. However, the interpretability of this algorithm is also important because it provides insight into OSA-induced drowsy driving. The measures here are inherently noisy

Author Manuscript

and may undermine algorithms that are sensitive to noise such as *k*-nearest neighbor and rule-based methods (Kotsiantis, Zaharakis, & Pintelas, 2007). The goals of interpretability and robustness to noise eliminate many modeling approaches including a Support Vector Machine with a string kernel, which would generally be associated with sequence frequency features (Leslie et al., 2002). One modeling approach that satisfies these requirements is random forests.

Random forests, developed by Breiman (2001), are a series of decision trees (Kotsiantis et al., 2007) trained on random bootstrapped samples of training data, and randomly selected feature subsets. Classification is performed by a majority vote among the predictions of the trees for each new instance. A key feature of random forests is that they naturally produce a measure of variable importance. This measure is based upon the frequency that a feature appears in the algorithm and its predictive power in each appearance. The variable importance compensates for the loss of interpretability of a multi-tree algorithm by demonstrating key variables associated with classification. In this study, important features represent driving patterns that differentiate between CPAP adherence and non-adherence.

Sensitivity analysis

One of the limitations of SAX is the amount of input parameters. Multiple SAX parameters could influence the results, in particular, the alphabet size. In order to understand the impact of this parameter and select an optimum, random forest models were fit to data reduced with SAX alphabet sizes between 9 and 23 letters. The model's prediction performance was assessed with the area under the receiver operating characteristic curve (AUC). The AUC is a measure of classifier performance that is insensitive to distributions of positive and negative examples in the dataset. An AUC of 0.5 (the line with a slope of 1 in the receiver operating characteristic (ROC) plot) represents the performance of a random classifier and an AUC of 1.0 represents a perfect classifier (Fawcett, 2004).

Figure 5 shows the AUC at each SAX alphabet size along with a 95% bootstrapped confidence interval based on 2,000 replicates. While none of the differences between the models are statistically significant based on the bootstrapped confidence interval, there is a clear trend in the prediction performance that peaks at an alphabet size of 13 letters and drops off after 17 letters. This trend suggests that alphabet sizes between 13 and 17 more adequately capture driving patterns associated with untreated sleep apnea. Furthermore, the results suggest that 13 should be retained as the alphabet size for the final algorithm.

Prediction results

The sensitivity analysis suggests that the final algorithm should be based on a SAX alphabet with 13 letters. The ROC curve for this final algorithm is shown in Figure 6. The AUC for the algorithm is 0.74, the lower bound of the confidence interval is 0.68, and the upper bound of the confidence interval is 0.80. The algorithm AUC and the lower bound of the confidence interval are higher than 0.50, which suggests that the algorithm predicts CPAP adherence significantly better than random. Moreover, this result indicates that the algorithm is generalizable and that it is detecting meaningful associations in the data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Variable importance

The quality of the algorithm predictions demonstrated by the evaluation criteria suggests that this algorithm identifies meaningful associations between driving data and CPAP use. Thus, exploring the variable importance of the features used to train the algorithm can provide valuable insight. Figure 7 shows a plot of the variable importance for the ten most important features in the random forest CPAP use algorithm. The features in Figure 7 are identified by their associated variable followed by the appropriate sequence. Thus, the most important variable is the sequence, "bbbbbbbbbb" for the speed variable. Note that speed is on a scale from "a" to "m," where "a" represents speeds over 90 kph and "m" represents speeds less than 6 kph, and that the acceleration values are on a scale of "a" to 'd," where "a" represents accelerations over 0.14 g an "d" represents 0 g of acceleration. The full table of speed and acceleration mappings is illustrated in Table 2.

Several interesting observations emerge from Figure 7. Many of the variables listed correspond to sustained periods of constant speed or acceleration. For example, "speed_bbbbbbbbbb" and "speed_bbbbbbbbb" correspond to 27 s and 30 s of sustained speed between 76 and 87 kph, or 45 and 54 mph. These speeds correspond to common speed limits on state highways. The prevalence of these factors may indicate that drivers with OSA change their route-choice based on their level of drowsiness or that drives involving highway travel are more likely to also include drowsiness events. This latter correlation is consistent with the drowsy driving literature (Brown, 1994). In contrast, the "latacc_aaa" variable corresponds to peak lateral acceleration. This pattern may be associated with turns or lane changes which may be common in more alert drivers.

Variable frequency analysis

The variables observed as most important in the prediction and classification of CPAP adherence warrant further analysis. One extension of the variable importance is to analyze the frequency of the variables across CPAP use definitions. This type of analysis enhances the variable importance measurement by illustrating the states where the ten most important features are common. Figure 8 shows the normalized frequencies of the ten most important features. The frequencies in these plots have been normalized relative to the number of drives in each condition and the overall distribution of treated and untreated days. Bars to the left of 0 in the chart indicate features are more common in the drivers that adhered to CPAP and the bars to the right of 0 show that the features are more common in drivers that did not adhere to CPAP. These results seem to confirm the hypothesis that drivers who adhere to CPAP make different route choices as compared to drivers that do not.

The data in Figure 8 can be further segregated by pre and post-CPAP adherence to provide further insight into the impact of CPAP use on driving pattern frequency. Figure 9 shows this breakdown. The figure shows the frequency of each pattern of speed and acceleration for drives prior to the start of CPAP (pre-CPAP), drives after the start of CPAP use where drivers did not use their CPAP (post-CPAP non-adhering), and drives after the start of CPAP use where drivers adhered (post-CPAP adhering). The figure illustrates that in most cases the pre-CPAP and post-CPAP non-adhering frequencies are aligned and that they differ significantly from the post-CPAP adhering frequencies. This result provides credence to the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

decision to combine pre-CPAP and post-CPAP non-adhering cases in the training data for the model. It also confirms the previous literature findings that non-adherence to CPAP impairs drivers to similar levels of untreated OSA.

Analysis of healthy controls

The random forest CPAP adherence detection algorithm partitions participants based on CPAP adherence and non-adherence. A natural extension of this algorithm would be to apply it to differentiating healthy individuals from individuals with untreated sleep apnea. If the algorithm successfully differentiates healthy drivers, it suggests that it is sensitive to the differences between healthy drivers and those with untreated OSA. One method to accomplish this task is to apply SAX to the healthy control data and then use the model to predict the state of the healthy controls. The model predictions should identify the controls with treated OSA participants in most cases, because studies indicate that proper adherence to CPAP therapy reduces crash risk to nearly the level of healthy drivers (Tregear et al., 2010). Table 3 shows the prediction results and accuracy of the predictions for the control participants after applying SAX and generating features following the same methods as the OSA data. Data from all 35 control participants were used in the predictions.

The table illustrates that the model is accurate on 81.41% of the predictions. This level of accuracy suggests that the algorithm effectively differentiates healthy drivers from drivers with OSA.

Discussion

OSA is a significant health and safety concern because it is associated with increased risks for cardiovascular diseases and automobile crashes. The link between untreated OSA and automobile crashes suggests that data collected from vehicles may be effective for developing algorithms identifying untreated and undiagnosed sleep apnea. The purpose of this study was to evaluate the effectiveness of using kinematic driving data and SAX to develop an algorithm for detecting CPAP treatment adherence. Such algorithms could be combined with unobtrusive data recording devices to help healthcare practitioners remotely monitor treatment adherence and effectiveness. Alternatively, the algorithms could be employed by professional driving companies to monitor their fleet of drivers and ensure that they are healthy and adhering to prescribed treatment or identify if drivers patterns of driving align with individuals diagnosed with OSA.

The study showed that a random forest algorithm containing speed- and acceleration-based SAX features (i.e. driving patterns) could detect CPAP treatment adherence significantly better than a random classifier. The important features of the algorithm appear to align with planning decisions and driving maneuvers. The success of this algorithm suggests that driving data is an effective tool for differentiating treatment adherence in patients with OSA. Subsequent analyses suggest that the algorithm may also be used to differentiate healthy control participants from those with untreated OSA. Furthermore, the results suggest that SAX is an effective approach for feature generation in driving algorithms. For example, this approach may be used to supplement other drowsiness detection algorithms such as the sleep apnea component of the Bayesian Network model in Ji, Zhu, and Lan (2004). Future analyses should explore this approach on a larger dataset with a more exhaustive evaluation of SAX parameters and comparisons with other time-series feature generation techniques.

Despite its success, this method is limited in several respects. The size of the dataset and scope of recruiting criteria may limit the conclusions based on this model. In addition, the binary labeling and exclusion of CPAP use between 0 and 4 hours may limit the ability of the model to classify data in a more realistic setting. The model structure is also limited, as it does not capture trends in CPAP use over time. These limitations could be addressed with additional data and a temporal modeling framework such as a Hidden Markov Model (HMM). The HMM framework allows for uncertainty of CPAP prediction which might be used to capture degrees of CPAP usage and directly models the temporal dependencies of CPAP use. These extensions should be explored in future work.

Conclusion

This article considers driving data from OSA patients and demonstrates that cars can detect compliance with CPAP treatment adherence and differentiate healthy drivers. This work represents a preliminary step toward integrating driving data with other health indicators, which may be valuable for evaluating new devices, analyzing side effects of medications, and augmenting current drowsiness detection algorithms. Future work should consider expanding the dataset size, evaluating other driving behavioral indicators, more complex modeling architectures, and integration with other drowsiness detection algorithms. In addition to OSA, the analysis process here could be applied to evaluate other chronic conditions such diabetes and age-related illness that may impact driving behavior.

Acknowledgments

The data presented in this paper were collected as part of a project sponsored by the National Institutes of Health (NIH) and parts of this work were presented at the 57th annual meeting of the Human Factors and Ergonomics Society. The authors would like to acknowledge the considerable efforts of the data collection team.

References

Agrawal R, Faloutsos C, Swami A. Efficient similarity search in sequence databases. In: Lomet DB, editorProceedings of the 4th International Conference, FODO ‘93 Chicago, Illinois, USA, October 13–15, 1993. Vol. 8958546. Chicago, IL: Springer Berlin Heidelberg; 1993. 69–84.

Barbé F, Sunyer J, de la Peña A, Pericas J, Mayoralas LR, Antó JM, Agustí AGN. Effect of continuous positive airway pressure on the risk of road accidents in sleep apnea patients. Respiration; International Review of Thoracic Diseases. 2007; 74(1):44–49. DOI: 10.1159/000094237 [PubMed: 16804289]

Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.

Brown ID. Driver Fatigue. Human Factors. 1994; 36(2):298–314. DOI: 10.1177/001872089403600210 [PubMed: 8070794]

Chan K, Fu AW. Proceedings of the 15th International Conference on Data Engineering. Sydney, NSW, Australia: 1999. Efficient time series matching by wavelets; 126–133.

Das D, Zhou S, Lee JD. Differentiating alcohol-induced driving behavior using steering wheel signals. IEEE Transactions on Intelligent Transportation Systems. 2012; 13(3):1355–1368. DOI: 10.1109/TITS.2012.2188891

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dingus T, Klauer S, Neale V, Petersen A, Lee SE, Sudweeks J, Knipling RR. The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment. Washington D.C: 2006. Retrieved from http://trid.trb.org/view.aspx?id=783477

Doshi A, Morris BT, Trivedi MM. On-road prediction of driver's intent with multimodal sensory cues. IEEE Pervasive Computing. 2011; 10(3):22–34. DOI: 10.1109/MPRV.2011.38

Dozza M, Bärgman J, Lee JD. Chunking: A procedure to improve naturalistic data analysis. Accident Analysis & Prevention. 2013; 58:309–17. DOI: 10.1016/j.aap.2012.03.020 [PubMed: 23000042]

Fawcett T. ROC graphs: Notes and practical considerations for researchers. Machine learning. 2004; 31(1):1–38.

Findley LJ, Smith C, Hooper J, Dineen M, Suratt PM. Treatment with nasal CPAP decreases automobile accidents in patients with sleep apnea. American Journal of Respiratory and Critical Care Medicine. 2000; 161(3):857–859. [PubMed: 10712333]

Gordon TJ, Kostyniuk LP, Green PE, Barnes MA, Blower D, Blankespoor AD, Bogard SE. Analysis of crash rates and surrogate events. Transportation Research Record: Journal of the Transportation Research Board. 2011; 2237(1):1–9. DOI: 10.3141/2237-01

Guo F, Klauer SG, Hankey JM, Dingus TA. Near Crashes as Crash Surrogate for Naturalistic Driving Studies. Transportation Research Record: Journal of the Transportation Research Board. 2010; 2147(1):66–74. DOI: 10.3141/2147-09

Jap BT, Lal S, Fischer P, Bekiaris E. Using EEG spectral components to assess algorithms for detecting fatigue. Expert Systems with Applications. 2009; 36(2):2352–2359. DOI: 10.1016/j.eswa.2007.12.043

Ji Q, Zhu Z, Lan P. Real-time nonintrusive monitoring and prediction of driver fatigue. IEEE Transactions on Vehicular Technology. 2004; 54(4):1052–1068.

Jovanis PP, Aguero-Valverde J, Wu K, Shankar V. Analysis of naturalistic driving event data omitted-variable bias and multilevel modeling approaches. Transportation Research Record: Journal of the Transportation Research Board. 2011; 2236:49–57.

Keogh E, Chakrabarti K, Pazzani M, Mehrotra S. Dimensionality reduction for fast similarity search in large time series databases. Knowledge and Information Systems. 2001; 3(3):263–286. DOI: 10.1007/PL00011669

Keogh E, Kasetty S. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Mining and Knowledge Discovery. 2003; 7:349–371. Retrieved from http://link.springer.com/article/10.1023/A:1024988512476.

Kotsiantis SB, Zaharakis ID, Pintelas PE. Supervised machine learning: A review of classification techniques. Informatica. 2007; 31:501–520.

Krajewski J, Sommer D, Trutschel U, Edwards D, Golz M. Fifth international driving symposium on human factors in driver assessment, training, and vehicle design. Big Sky: MT; 2009. Steering wheel behavior based estimation of fatigue; 118–124.

Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, Dinges DF. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. The American Review of Respiratory Disease. 1993; 147(4):887–95. DOI: 10.1164/ajrccm/147.4.887 [PubMed: 8466125]

Kushida C, Littner M, Hirshkowitz M. Practice parameters for the use of continuous and bilevel positive airway pressure devices to treat adult patients with sleep-related breathing disorders. Sleep. 2006; 29(3):1–6. Retrieved from http://www.medfusion.net/templates/groups/2719/3668/OSA4.pdf.

Lal SKL, Craig A, Boord P, Kirkup L, Nguyen H. Development of an algorithm for an EEG-based driver fatigue countermeasure. Journal of Safety Research. 2003; 34(3):321–328. [PubMed: 12963079]

Leslie C, Eskin E, Noble WS. Pacific symposium on biocomputing. Vol. 7. Lihue, Hawaii: 2002. The spectrum kernel: A string kernel for SVM protein classification; 566–575.

Lin J, Keogh E, Wei L, Lonardi S. Experiencing SAX: A novel symbolic representation of time series. Data Mining and Knowledge Discovery. 2007; 15(2):107–144. DOI: 10.1007/s10618-007-0064-z

MacLean AW, Davies DR, Thiele K. The hazards and prevention of driving while sleepy. Sleep medicine reviews. 2003; 7(6):507–521. [PubMed: 15018093]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

McCall JC, Wipf DP, Trivedi MM, Rao BD. Lane change intent analysis using robust operators and sparse bayesian learning. IEEE Transactions on Intelligent Transportation Systems. 2007; 8(3): 431–440. DOI: 10.1109/TITS.2007.902640

McDonald AD, Lee JD, Aksan NS, Dawson JD, Tippin J, Rizzo M. The language of driving: advantages and applications of symbolic data reduction for naturalistic driving data analysis; 92nd Annual Meeting of the Transportation Research Board; Washington D.C. 2012.

McLaurin E, McDonald AD, Lee JD, Aksan N, Dawson J, Tippin J, Rizzo M. Variations on a theme Topic modeling of naturalistic driving data; Proceedings of the Human Factors and Ergonomics Society Annual Meeting; Chicago, IL. 2014. 2107–2111.

Moreno CRC, Carvalho FA, Lorenzi C, Matuzaki LS, Prezotti S, Bighetti P, Lorenzi-Filho G. High risk for obstructive sleep apnea in truck drivers estimated by the berlin questionnaire prevalence and associated factors. Chronobiology International. 2004; 21(6):871–879. [PubMed: 15646234]

Orth M, Duchna H-W, Leidag M, Widdig W, Rasche K, Bauer TT, Kotterba S. Driving simulator and neuropsychological [corrected] testing in OSAS before and under CPAP therapy. The European Respiratory Journal. 2005; 26(5):898–903. DOI: 10.1183/09031936.05.00054704 [PubMed: 16264053]

Raksincharoensak P, Khaisongkram W, Nagai M, Shimosaka M, Mori T, Sato T. Integrated driver modelling considering state transition feature for individual adaptation of driver assistance systems. Vehicle System Dynamics. 2010; 48(sup1):55–71. DOI: 10.1080/00423111003668229

Rauscher H, Popp W, Wanke T, Zwick H. Acceptance of CPAP therapy for sleep apnea. Chest. 1991; 100(4):1019–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9445527. [PubMed: 1914550]

Rigas G, Goletsis Y, Bougia P, Fotiadis DI. Towards driver's state recognition on real driving conditions. International Journal of Vehicular Technology. 2011; 2011:1–14. DOI: 10.1155/2011/617210

Rosenfeld A, Bareket Z, Goldman CV, Kraus S, Leblanc DJ, Tsimoni O. Learning driver's behavior to improve adaptive cruise control. Journal of Intelligent Transportation Systems. 2015; 19(1):18–31. Retrieved from https://www.tandfonline.com/doi/full/10.1080/15472450.2014.889960.

Sassani A, Findley LJ, Kryger M, Goldlust E, George C, Davidson TM. Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome. Sleep. 2004; 27(3): 453–458. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15164898. [PubMed: 15164898]

Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: Clinical and empiric insights for developing CPAP adherence interventions. Sleep Medicine Reviews. 2011; 15(6):343–356. DOI: 10.1016/j.smrv.2011.01.003 [PubMed: 21652236]

Shrestha D, Lovell DJ, Tripodis Y. Hardware and software for collecting microscopic trajectory data on naturalistic driving behavior. Journal of Intelligent Transportation Systems. 2017; 21(3):202–213.

Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine. 2009; 5(6):573–81. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2792976&tool=pmcentrez&rendertype=abstract. [PubMed: 20465027]

Tregear S, Reston J, Schoelles K, Phillips B. Continuous positive airway pressure reduces risk of motor vehicle crash among drivers with obstructive sleep apnea: Systematic review and meta-analysis. Sleep. 2010; 33(10):1373–80. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2941424&tool=pmcentrez&rendertype=abstract. [PubMed: 21061860]

Turkington PM, Sircar M, Saralaya D, Elliott MW. Time course of changes in driving simulator performance with and without treatment in patients with sleep apnoea hypopnoea syndrome. Thorax. 2004; 59(1):56–59. [PubMed: 14694250]

Weaver TE, Kribbs NB, Pack AI, Kline LR, Chugh DK, Maislin G. Night-to-night variability in CPAP use over the first three months of treatment. Sleep. 1997; 20:278–283. [PubMed: 9231953]

Williams D. The Arbitron National In-Car Study. 2009; 10019 Retrieved from http://www.arbitron.com/downloads/InCarStudy2009.pdf.

Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. In: Fisher DH, editorProceedings of the Fourteenth International Conference on Machine Learning. Vol. 97. San Francisco, CA: Morgan Kaufmann; 1997. 412–420.

Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, Hla KM. Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. WMJ: Official Publication of the State Medical Society of Wisconsin. 2009; 108(5):246–249. DOI: 10.1016/j.str.2010.08.012.Structure [PubMed: 19743755]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

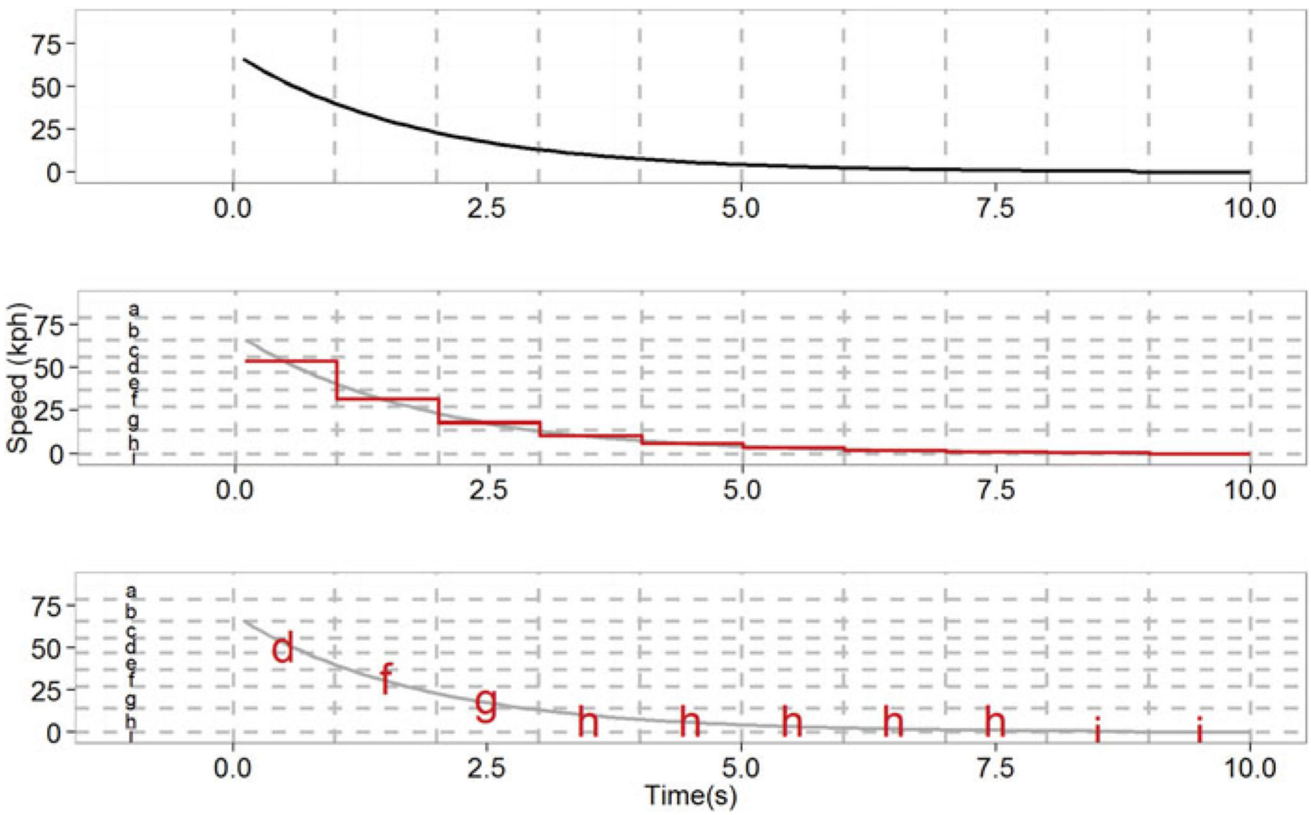


Figure 1.
Demonstration of SAX steps for a 10 s sample of data.

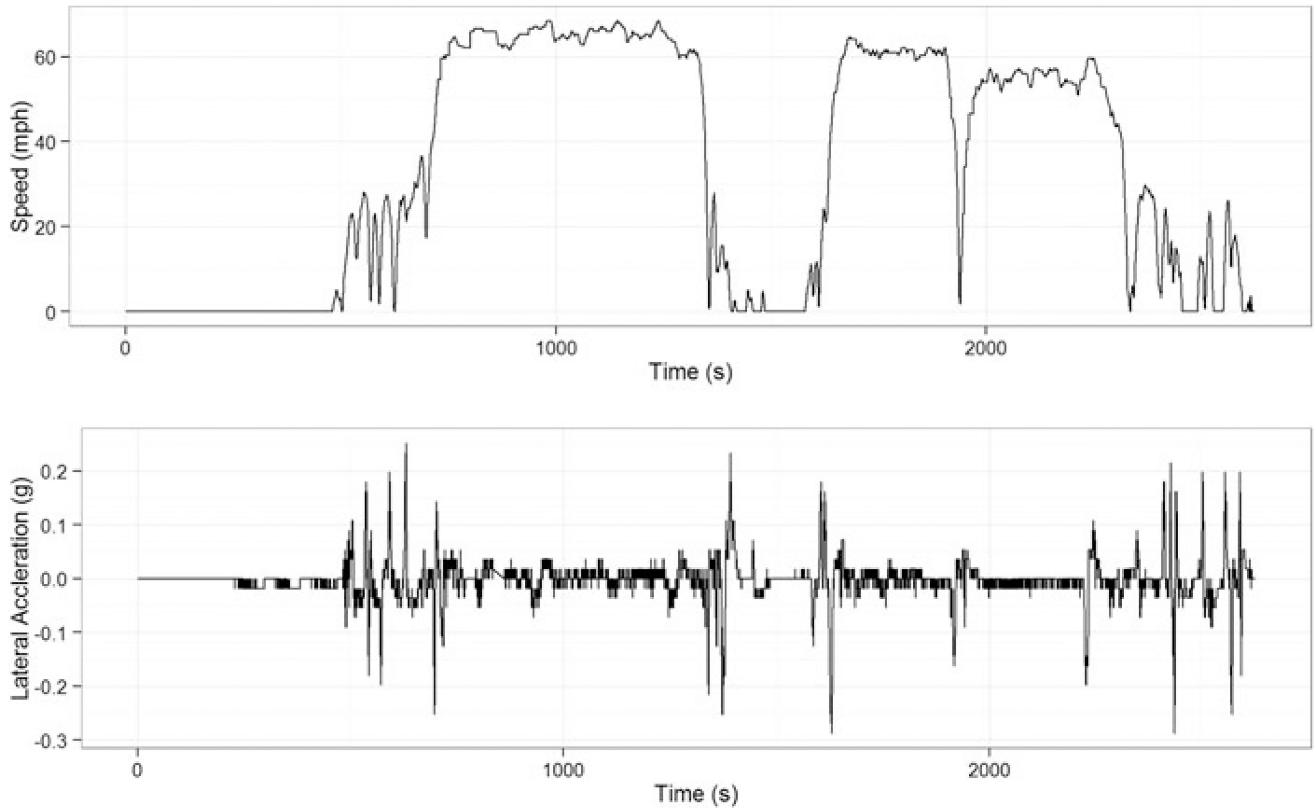


Figure 2.
Samples of speed and lateral acceleration data for a single drive.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

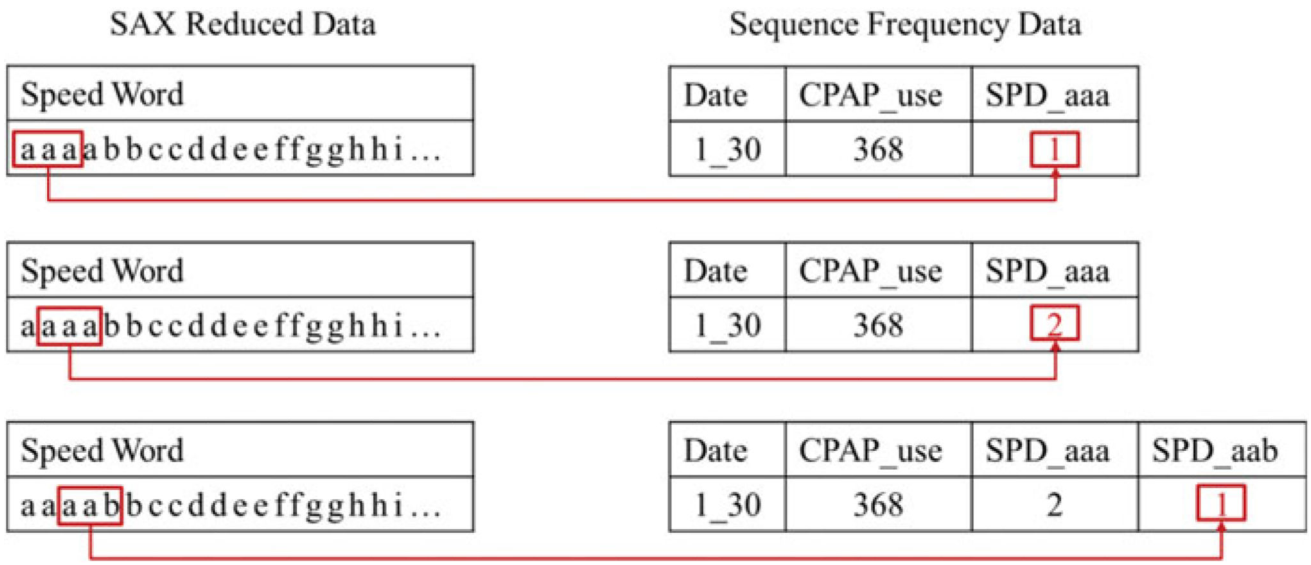
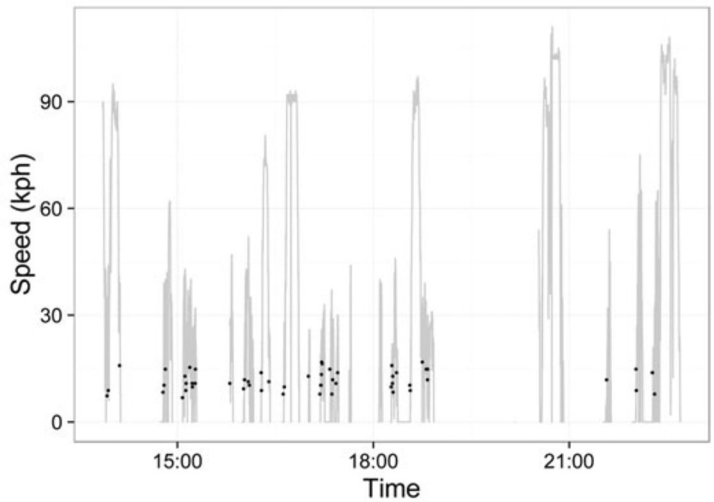


Figure 3.
Demonstration of three steps of the sequence frequency dataset creation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Drive	SAX Speed Letters
1	a a a a a a a a a a b b b b b c c c d d d e f g g g h h i i i i i i i i i i j **l l k** k j j i i h h h h h h i i h i i i m j j j i i i j k k l k j i i i i i i i i j j j k k k k l l l **l l k** k j j i i h h h h h h k m l k j j k l **l l k** k j j i i i i i h h g g f f f e d d c c c c c c d d e g h h g f f e e d d d c c c c c c c c b b b b b b b b a a a a a a a a a a a a a a a a a a a a a a b b b b b b b b a a a a a a a a a a a b b b b b b b b a a a a b b b b b b b b b b b b b b b b b b b b b b b b b b b b b b b a a b b b b b b b a a a a a a a b b b b b b b b c c c d d d e e f g i k k j i h h h h i i i i i i i i i h h h i i h i i j k l l l l l l m m m m
2	m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m l l l **l l k** j i i h i j l l m m m m m m m m m **l l k** k j j i i i i i i h h h j k k k k k k l **l l k** k k k l m m l k i i i j k k j i h h h h i i k m m m m m m m m m m m m m m m m l k j h g g f f f e e e e e e e e e e e e d e e e e e e e d d e e e e e e f g i l k k k k j j k k k j i i i i i j j j i i j j j j j k k l k l l m m m m m m
3	m m m m m l l l l **l l k** k j j i i i i i i i i i h h h h h i i i i j l m l k j i i h h h h h h h h h h h h h k **l l k** k k k k k k k k k k k k k l **l l k** k k **l l k** k l k j j k m m m l k k j j i j j i i i i i j k k l l j i j j k k j j j j j i i i j j j j j k l k k k k k k l k j i i i i i j j j **l l k** j h h i i i i j k j i j l k k k k k k k l l m m m m m m m m m m m m m l **l l k** k j k l **l l k** k k k k k k k k k k k j j k k l m m m m m m l l m l k k j j j j j j j k k k j j j j j k k k k l **l l k** l l l **l l k** k l m m m m
4	j j j j j j j j j k l m **l l k** k k j j j j j j j j j j i i h h g g g g g g h h j k l k k j j j j j j j j k l l l l l m m m m m m
5	m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m l l m m m **l l k** k k k k k k j j j k k k **l l k** j i h h h h h h h h h h h h h h h h h h h j k l k k j j i i i j j j i i i h h h h h h h h h h h h h h h h h h h h g g f f g h h h i i j **l l k** j i i j k m k j j k l m m m m m **l l k** j i i l m m m m l k i i i i j l m m m m l m m m m m l k k j j l m m l k k k l l l m l k l m m m l k k l m m m m l l l l l m m m m m m m
...	...

Subject	Adhered	CPAP use	Date	acc_ddb	acc_ddd	acc_dddb	speed_llk
OSA003	false	0	2011-05-26	3	308	3	3
OSA003	false	0	2011-06-05	11	1276	9	11
OSA003	false	0	2011-06-09	11	257	9	10
OSA003	false	0	2011-06-11	23	1341	20	9
OSA003	false	0	2011-06-12	15	1212	12	7
OSA003	true	411	2011-06-17	15	1152	11	9
OSA003	false	0	2011-06-28	14	686	11	13
OSA003	**false**	**0**	**2011-08-06**	**50**	**3118**	**38**	**53**
OSA003	true	439	2011-08-11	14	1047	12	10
OSA003	true	409	2011-08-20	7	883	7	8
OSA003	true	366	2011-08-29	5	346	5	4

Figure 4.

Depiction of the data reduction process for a single feature (speed_llk), subject, and day. The top right plot shows the unprocessed 60 Hz drives on a street map, occurrences of the feature are highlighted in black points. The top left plot shows the 1 Hz reduced speed data for the drives occurring over the day with the feature occurrences highlighted by black points. The bottom left table shows the SAX conversion of the speed data for each drive with occurrences of the feature highlighted. The bottom right table shows the row in the final dataset for the date and the total count of occurrences of the speed_llk feature, 53.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

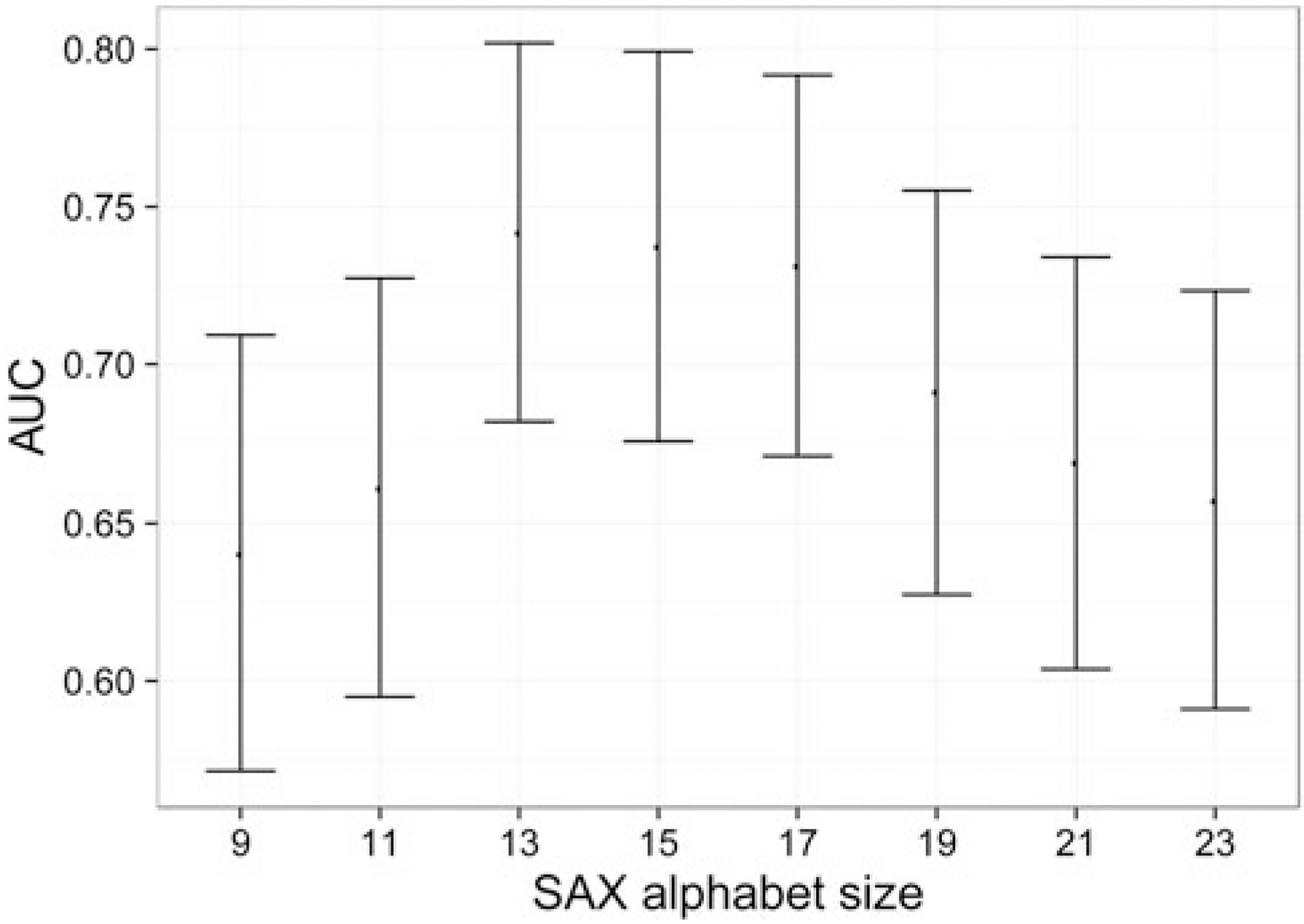


Figure 5.
Area under the curve and 95% bootstrapped confidence intervals for the SAX alphabet sizes evaluated.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

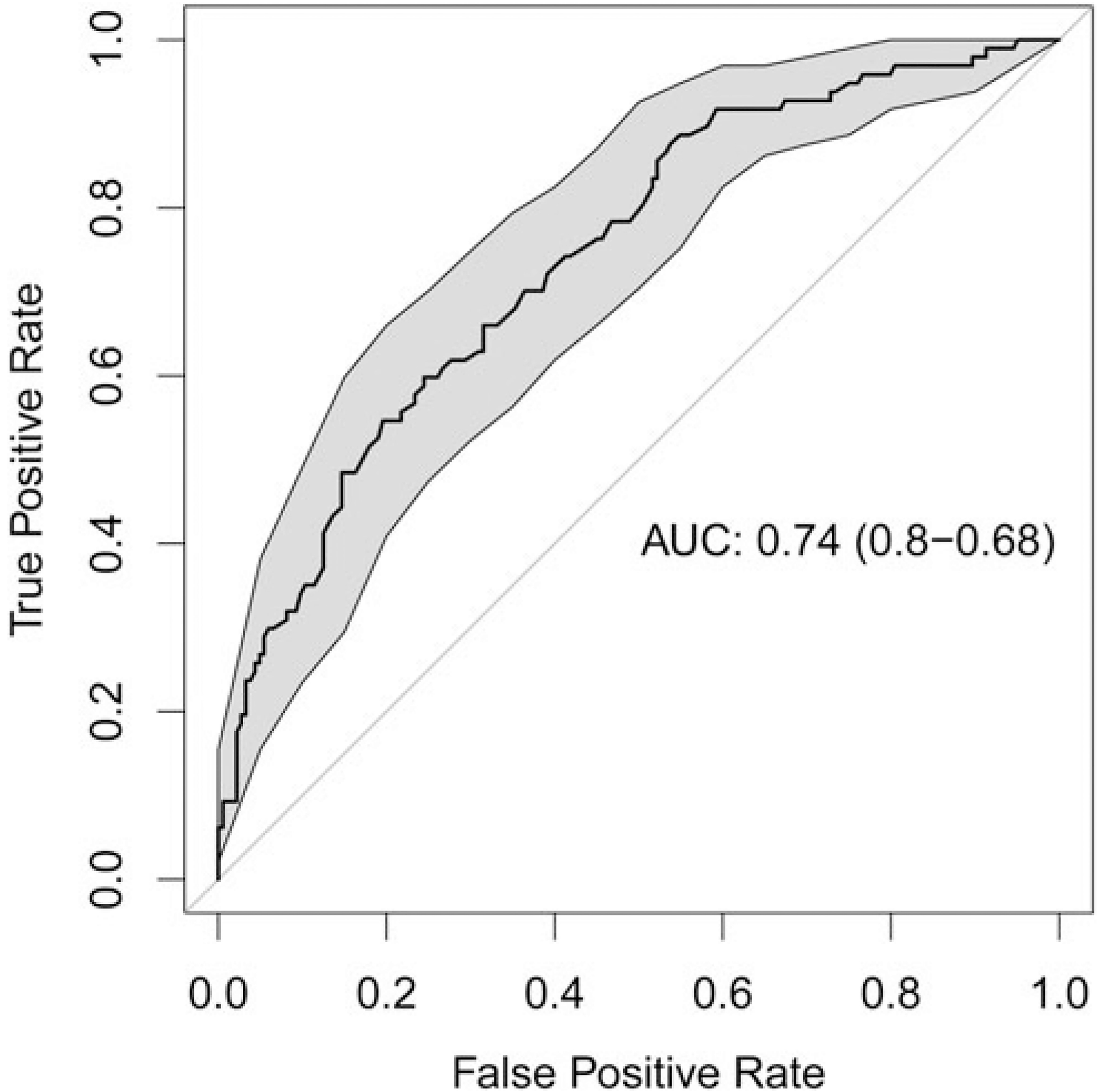


Figure 6.

ROC curve for the random forest model trained on SAX pattern frequency data. The grey band represents a bootstrapped 95% confidence interval based on 2,000 replicates. The label shows the AUC and the bounds of the bootstrapped confidence interval.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

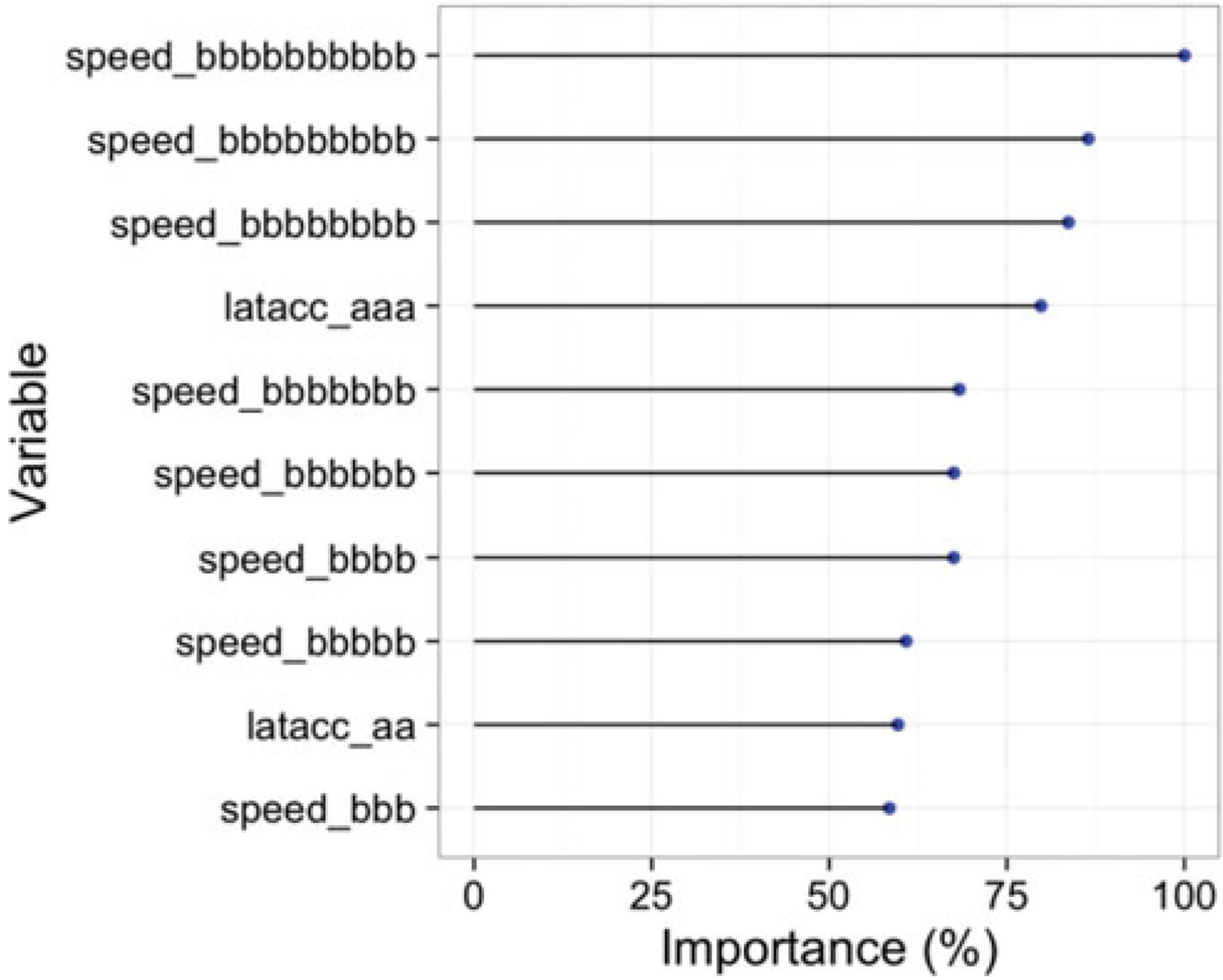


Figure 7.
Variable importance pin plot for the ten most important variables in the random forest CPAP prediction algorithm. Note that the raw variable importance has been scaled from 0 to 100 relative to the most important variable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

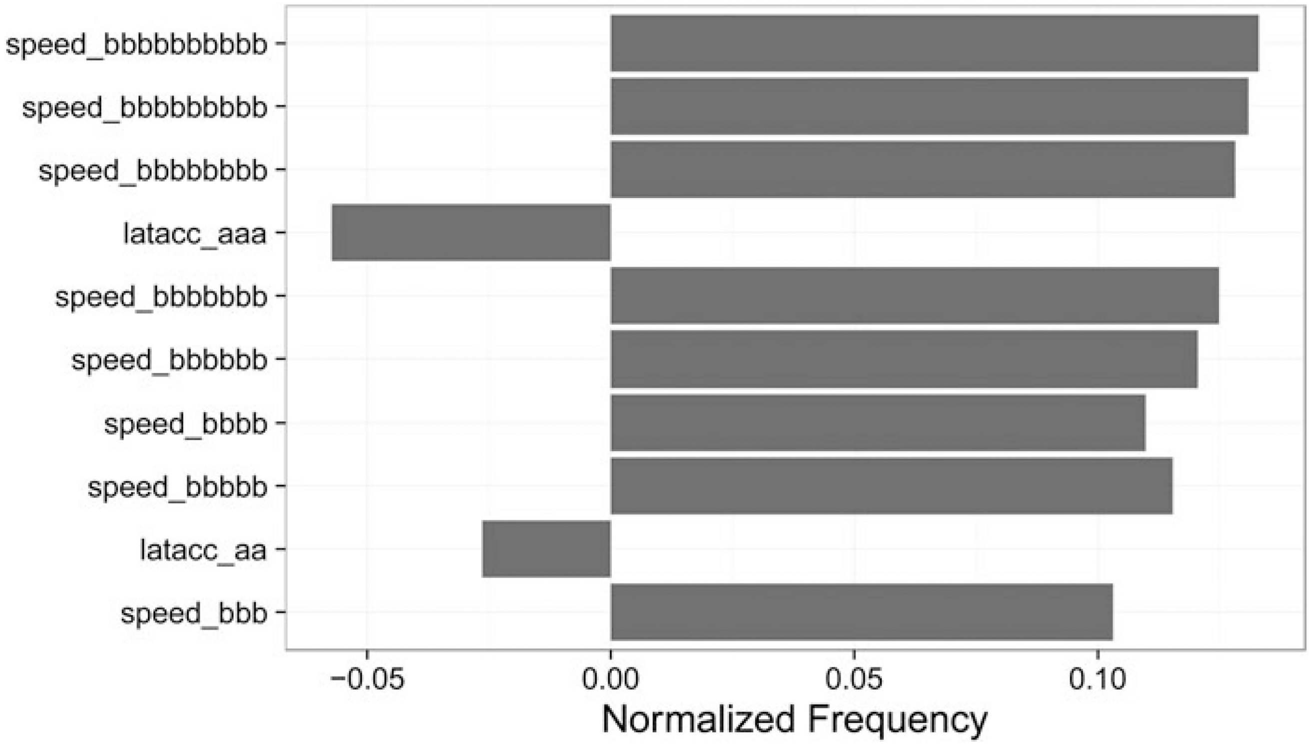


Figure 8.

Bar chart of normalized frequencies for the most important speed variables in the training and test data. Negative values indicate the variables are more common than expected in treated patients, positive values indicate the variable is more common in untreated patients.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

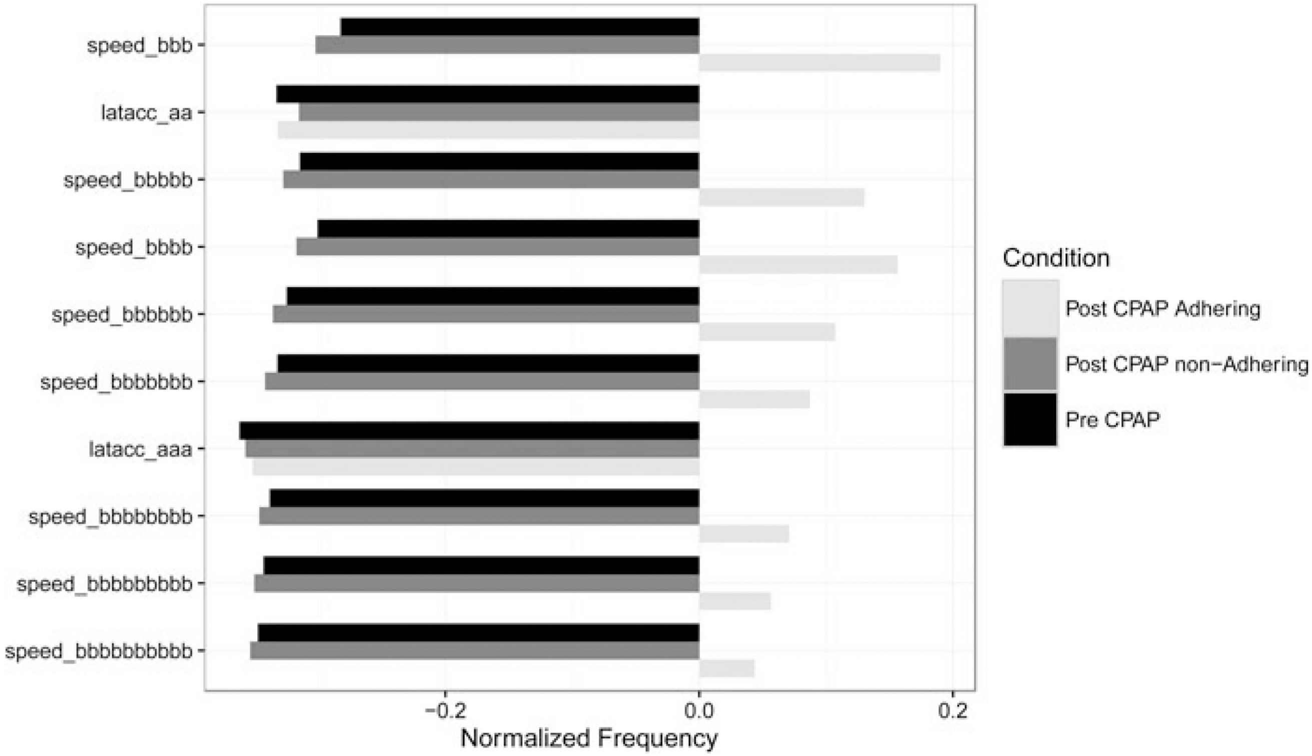


Figure 9.

Normalized frequency of speed and acceleration patterns broken down bypre-CPAP, post-CPAP with adherence, post-CPAP with non-adherence. Data are normalized relative to the number of drives in each area.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographic and anthropometric data for the naturalistic driving study participants.

Variable	OSA participants	Control participants
Number of participants	69	35
Age (years)	46 (7.7)	45 (8.1)
Gender	47 males, 22 Females	17 Males, 18 Females
Apnea–hypopnea index pre-GPAP (AHI)	33.55 (34.23)	
Body mass index (BMI)	37.28 (9.01)	

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Speed and acceleration bounds for each SAX letter. Speed is in kph and acceleration is in g.

Letter	Speed bounds (kph)	Acceleration bounds (g)
a	$x > 87.64$	$x > 0.14$
b	$87.64 < x < 76.07$	$0.14 < x < 0.10$
c	$76.07 < x < 67.99$	$0.10 < x < 0.06$
d	$67.99 < x < 61.32$	$0.06 < x < 0.00$
e	$61.32 < x < 55.36$	
f	$55.36 < x < 49.75$	
g	$49.75 < x < 44.25$	
h	$44.25 < x < 38.64$	
i	$38.64 < x < 32.68$	
j	$32.68 < x < 26.01$	
k	$26.01 < x < 17.93$	
l	$17.93 < x < 6.36$	
m	$6.36 < x < 0.00$	

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 3

Prediction results for the random forest CPAP adherence detection algorithm with the data from the healthy control participants.

Total predictions	Correct predictions	Accuracy
2,292	1,866	81.41%