Published in final edited form as:

Transp Res Rec. 2013 ; 2392: 22–30. doi:10.3141/2392-03.

The Language of Driving: Advantages and Applications of Symbolic Data Reduction for Analysis of Naturalistic Driving Data

Anthony D. McDonald,
Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Mechanical Engineering, 1513 University Avenue, Madison, WI 53706

John D. Lee,
Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Mechanical Engineering, 1513 University Avenue, Madison, WI 53706

Nazan S. Aksan,
Department of Neurology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 2006 Roy J. Carver Pavilion

Jeffrey D. Dawson,
Department of Biostatistics, College of Public Health, 105 River Street, S-145 College of Public Health Building, University of Iowa, Iowa City, IA 52242

Jon Tippin, and
Department of Neurology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 2006 Roy J. Carver Pavilion

Matthew Rizzo
Department of Neurology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, 2006 Roy J. Carver Pavilion

Abstract

Recent advances in onboard vehicle data recording devices have created an abundance of naturalistic driving data. The amount of data exceeds the resources available for analysis; this situation forces researchers to focus on analyses of critical events and to use simple heuristics to identify those events. Critical event analysis eliminates the context that can be critical in understanding driver behavior and can reduce the generalizability of the analysis. This work introduced a method of naturalistic driving data analysis that would allow researchers to examine entire data sets by reducing the sets by more than 90%. The method utilized a symbolic data reduction algorithm, symbolic aggregate approximation (SAX), which reduced time series data to a string of letters. SAX can be applied to any continuous measurement, and SAX output can be reintegrated into a data set to preserve categorical information. This work explored the application of SAX to speed and acceleration data from a naturalistic driving data set and demonstrated SAX's integration with other methods that could begin to tame the complexity of naturalistic data.

Corresponding author: A. D. McDonald, admcdonald@wisc.edu.

The recent developments in onboard vehicle data acquisition systems (DAS) and the increased interest in naturalistic driving studies have created an environment where the amount of data collected outpaces the available resources to analyze it. Naturalistic studies have generated terabytes if not petabytes of data, which have overwhelmed traditional approaches to data analysis that were developed for epidemiological analysis of crash data or analysis of simulator or on-road experiments. This limitation led researchers to focus their analyses on critical events and small time intervals surrounding them (1–3). These critical events were generally identified by applying simple heuristics to measures such as steering wheel velocity or lateral and longitudinal acceleration (4). Some recording systems also allow critical events to be defined by optical sensors or the drivers themselves (5). In all of these situations, the time intervals included in the analysis seldom extended beyond 30 s surrounding the event.

The critical event approach allowed researchers to identify the immediate causes of critical events that occurred during the surrounding time interval, but the approach limited researchers' ability to understand long-term driving behavior. Analyzing complete data sets may provide key insights into naturalistic data, such as why a small proportion of drivers cause a large proportion of crashes and near crashes (6). Such analyses would also allow researchers to better understand driver behavior in the context of environmental factors, such as weather and traffic, to improve the validity and generalizability of critical event analyses (4).

Analyzing complete data sets requires time series data reduction techniques that can reduce data size and facilitate comparisons between drivers, drives, and driving events. The interpretability and ease of output manipulation will also affect the adoption of data reduction techniques among the naturalistic driving research community and thus are desirable. Many common data reduction techniques, including sliding window reduction, Fast Fourier Transforms, and wavelet transforms, fail to meet all of these requirements. The shortcomings of these methods with respect to naturalistic driving have pushed researchers to explore innovative techniques. One of these techniques, chunking, has shown promise, but does not provide a complete solution (7).

This paper introduces a data reduction technique, symbolic aggregate approximation (SAX), which satisfies all of the aforementioned requirements and demonstrates the application of SAX to naturalistic driving data. SAX converts time series data into strings of symbols that represent ranges of values in the measurement. This conversion expedites critical event detection, facilitates new visualizations of drives, and enables comparisons between complete driver data sets and drives.

SAX Time Series Analysis

Reducing hundreds of megabytes of data into a more concise record represents an important challenge for research that uses naturalistic data. A technique for dimensionality reduction, SAX operates by binning segments of data into quantiles labeled with a letter (8). SAX has been successfully applied in several domains, including healthcare (9), weather prediction (10), and finance (11). The technique has not been applied to driving data.

SAX can be applied to any continuous measure in the data, such as speed, acceleration, lane position, and heading. A major advantage of SAX is that it reduces dimensionality while maintaining the original structure of the data. Another advantage of SAX is that the symbolic conversion enables the application of natural language algorithms for pattern matching, word comparison, segmentation, and other textual analyses. These algorithms can be used to identify instances of interest in the data set, such as drowsiness-related lane departures.

With SAX, time series data from a single variable are evaluated over a time window defined in seconds. Inside each window, the raw data are normed by the local mean and standard deviation derived from the data in the window or from a global mean and standard deviation derived from a representative sample of the entire data set. Next, the data are divided into equal-size groups called epochs. The quantiles of the normal distribution are used to bin the mean of the data inside each epoch and the bins are used to convert values to letters. Each quantile corresponds to a specific letter, as shown in Figure 1 along the *y*-axis. After the data have been reduced to letters, the letters are combined into a word, with each window generating one word. The size of the window, the percentage of overlap between windows, word length, and alphabet size are all adjustable parameters. Figure 1 demonstrates SAX applied to a single window with a word length of six and an alphabet of five letters. The final result of the conversion is the word "Deedbb."

The window size can range from the maximum resolution of the data to the duration of the entire data set. The amount of overlap between windows can vary between 0% and 100%. Word length must be between one and the total number of samples per window (sampling rate * window size). The amount of data in each epoch depends on the sampling rate and window size. The alphabet size defines the resolution of the binning. Generally an alphabet of five to eight letters is sufficient (8). Increasing the alphabet size increases the precision of the resulting data set but results in fewer similarly labeled epochs.

Application of SAX to Naturalistic Driving Data

This section describes the application of the SAX reduction method to speed and acceleration from a naturalistic driving data set that was collected to assess the efficacy of a sleep apnea treatment. All analyses reported here were completed with R 2.15.1 (12).

Naturalistic Driving Data

The data described in the following analysis originated from an ongoing National Institutes of Health study focused on evaluating the effects of positive airway pressure (PAP) therapy used to treat sleep apnea. Sixty-five participants started a data collection process that was to continue for 3.5 months. The 65 participants represented two groups: participants diagnosed with obstructive sleep apnea (OSA) and healthy control participants. The OSA-afflicted patients represented 45 of the 65 drivers, including 31 males with a mean age of 47 (SD = 7.36). The following criteria were used to match individual control participants to a participant with OSA: gender, age (within 5 years), education (within 2 years), and county of residence. The control participants represented 20 of the 65 total participants, including

10 males with a mean age of 45 (SD = 8.53). Drivers were compensated for their time and effort pending the completion of the data collection and surveys.

Participants had their personal vehicles equipped with an in-vehicle data acquisition system (IV-DAS) that recorded Global Positioning System (GPS) information, onboard diagnostics (OBD) speed, three-axis accelerometer, and accelerator input at 10 Hz. The IV-DAS also collected intermittent video of the driver's face and the road. Video was recorded during the first minute after ignition, for a 20-s interval every 15 min during a drive, and a 20-s interval when the lateral or longitudinal acceleration measures exceeded 0.35 *g*. Data were collected for a period of 2 weeks before OSA participants receiving PAP therapy and for 3 months following the start of PAP therapy. Data for control participants were recorded for 3.5 months without interruption. The data were partitioned into individual drives defined by ignition engagement and disengagement.

The analysis reported here focused on OBD speed and accelerometer data from only the 45 participants afflicted with OSA. Three subjects were removed from the analysis because of incomplete data (two participants) or unreliable data (one participant). Speed and acceleration data were selected because they are common across many naturalistic driving studies, generally seen as diagnostic of driving behavior, and the most robust data sources in the current data set. Acceleration data contained some measurement uncertainty because the accelerometer could rotate slightly and cause purely lateral movements to be interpreted as a combination of lateral and longitudinal measures. For the effects of this uncertainty to be understood, acceleration was analyzed in two ways: first, based on lateral acceleration measures and second, based on a vector sum of lateral and longitudinal acceleration. The combined acceleration captured all peak accelerations regardless of direction (i.e., a hard stop or an abrupt swerve). Lateral acceleration alone will only be sensitive to abrupt swerves and turns and may miss some of these events if the accelerometer is poorly positioned. The combined measurement was used for all analyses except for the critical event detection.

Applying SAX

The primary goal of this work was to assess the applicability of SAX to naturalistic driving data. The parameters were selected through a combination of educated intuition and heuristics provided by Lin et al. (8). This selection process led to the following global input parameters: window size, 1 s; window overlap, 0%; and word length, one letter.

SAX was applied separately to speed and acceleration (lateral and combined) data. In both cases, the data were normed by a global estimate (based on 50,000 random samples) of the mean and standard deviation, with zero values removed. The zero values occurred at a significantly higher rate than other values and were removed to avoid skewing the mean and standard deviation measurements. The alphabet size varied between speed (nine letters) and acceleration (four letters). These alphabet sizes were selected with the goal of highlighting differences between common speed limits (i.e., 10 mph, 25 mph, 45 mph, and 65 mph) and unusually high accelerations (>0.4 *g*). The speed alphabet size was odd because an additional letter was incorporated for periods when the vehicle was stopped for the entire window. The letters and bin definitions for speed and combined acceleration are shown on top of the histograms of the sample data in Figure 2.

The resultant SAX output can be combined with categorical and other non-SAX variables from the original data to form a new data structure with a letter for speed, combined acceleration, and lateral acceleration characterizing each second of data. An example of the SAX-reduced data set is shown in Table 1. The speed and acceleration letter columns have been augmented with real-world measures to demonstrate the connection to the original data.

The data can be further reduced by combining consecutive rows with the same speed and acceleration letters and assigning an additional column that records the duration of each state. For example, if a driver was stopped at a traffic signal for 10 s, the event could be represented with a single row of duration 10, with speed letter "i," and acceleration letter "d." The sample data from Table 1 represent all the measures collected by DAS in this experiment. SAX is not specific to this data set and could be applied to any data set, independent of the sampling rate and measures taken. Integrating categorical variables or continuous variables with SAX reduction is simply a matter of bookkeeping. Categorical variables, such as vehicle type, that do not change over a window can be simply appended to the data set. Categorical variables and continuous variables that may change in a single window, such as road type or GPS location, can be combined based on a majority vote or the median value over the window, respectively.

Results

The initial data set used in this analysis contained 15,953 drive files from 42 drivers, corresponding to 22.2 GB of file storage space. The SAX output was contained in a single file corresponding to 1.65 GB, a 92.6% reduction. This level of reduction makes it feasible to conduct global analyses of driver behavior while preserving the potential for critical event analyses. The following section provides details on a subset of new analyses made possible with SAX, in addition to applying SAX to critical event analysis.

Identifcation of Critical Events

The symbols that code speed and acceleration (or any SAX-reduced variable) provide an index of the original data. Therefore, critical event analyses that use bounds, such as acceleration greater than 0.4 *g*, can be extracted from SAX-reduced data by simply analyzing all data containing the letter that matches the bound, acceleration *a* in this case. Beyond this single-letter case, SAX data can be mined for patterns of any length. For example, all patterns with large changes in speed over short periods of time, which would correspond to the string *ai*, can be extracted. These longer phrases may be used to expand the definition of a critical event.

To explore critical event pattern detection, this study used lateral acceleration and speed, specifically peak lateral accelerations (greater than 0.4 *g*, acceleration letter *a*) occurring at high speeds (greater than 79 km/h, speed letter *a*). These bounds were selected because they suggest a potentially dangerous combination of speed and lateral acceleration, such as swerving after a microsleep episode. The analysis identified 501 events in the data set, with a mean of 11.93 events per participant. However, a small proportion of the participants (six) accounted for nearly 54% (269) of the events. Eleven (26%) of the participants had one or no event.

Alone these events do not imply an accident or near accident; however, when the events are combined with other adverse conditions, such as driving during a circadian trough, they may be indicative of drowsy driving or some other impairment. Figure 3 shows the frequency of these events by hour for each participant, normalized by the number of drives by each participant. One participant, OSA016 (outlined in Figure 3), had a high density of these events in the very early morning. This process quickly identifies segments of interest for video analysis and can identify more complex critical events that extend over time rather than simply events based on variables exceeding a threshold.

SAX Visualizations

Naturalistic driving studies typically examine data at two levels: micro level analysis of real-time videos and kinematic data streams and macro level statistics about driver performance throughout a study (1, 6, 13). Data between these two poles are difficult to visualize and, by extension, gaining an understanding of the data at the drive level with anything other than aggregate summary measures is challenging. The SAX data reduction technique addresses this challenge with a novel visualization of drives that uses a two-dimensional density plot. Figure 4 shows density plots for two drives. Each box represents a SAX letter labeled with its minimum value in original units and the hue of each square encodes the proportion of the drive spent at each level. The top plot shows a residential drive during which most of the time is spent stopped or driving 56 km/h (∼35 mph) or slower. The bottom plot shows a highway drive during which most of the time is spent at low acceleration and speeds over 79 km/h (∼50 mph). These plots allow analysts to quickly understand the general contents of a drive in a way that aggregate summary measures cannot. In addition, the plots may alert analysts to adverse events during a drive: for example, if the box corresponding to speeds over 79 km/h and accelerations greater than 0.4 *g* was dark, the plot would suggest that a drive contained several hard stops or swerves.

Clustering and Comparing Data Across Drivers

These density plots can be used to identify similar individual drives. For example, Keogh et al. used density plots of genomic data to visually pair similar organisms (14). In the case studied here, the drives were analogous to the organisms and the groups represented different types of drives, for example a highway drive or a residential drive. Grouping these types of drives allows analysts to remove the variance in driver behavior associated with drive type from future analyses. This grouping process can be formalized and automated through the use of clustering, such as hierarchical clustering (15).

To cluster the drives based on the SAX description, each drive was defined by the time spent at each speed and combined acceleration letter, normed by the duration of the drive. The process resulted in the identification of six clusters (drive types). Figure 5 and Figure 6 each show a sample density plot from the drives contained in each cluster for two participants. Both figures show a clear separation between highway (Cluster 4) and short drives (Cluster 6) with long stops. Clusters 1, 2, 3, and 5 have more subtle differences and represent different forms of local drives. Clusters 1 and 2 have higher density toward the lower end of the speed scale and likely represent shorter drives than Clusters 3 and 5. Cluster 2 includes much higher density at higher acceleration and speed compared with Cluster 1, which suggests that the drives in the cluster occur in an environment with frequent and unpredictable stops, such as a city. Clusters 3 and 5 both include a highway component with some proportion of the drive occurring at speeds greater than 79 km/h; however, similar to Cluster 1, Cluster 3 has substantially higher density at low speed values compared with Cluster 5.

Most important, the clusters are similar across both drivers. These similarities make it possible to extract clusters common to all drivers and analyze them together. Essentially the clusters can be used to add to the data an additional categorical variable that codes the type of drive. To expand this cluster analysis, other drive information can be used, such as the presence of a median or the duration of the entire drive, to get a more specific definition of drive type. These drive types define the critical contextual information needed for sensible aggregation of naturalistic data.

Segmentation

The most important insight of this analysis was a shift from understanding naturalistic driving data as a series of events and summary metrics to a hierarchy of behavioral elements. The highest level of the hierarchy consists of the general behavior of the subjects within a study. This behavior can be decomposed into drives or trips characterized by drive type. The drives themselves are composed of segments and the segments are comprised of events. The concept of a driving segment seems obvious on a qualitative level. For example, a highway drive can be separated intuitively into the following segments: exiting a driveway, driving through a residential area, negotiating an entrance ramp, and driving on a highway. Defining these segments mathematically is significantly more difficult. Many of methods and algorithms that can be used for segmentation come from natural language processing research and require some form of text as input (16). This requirement aligns with the symbolic output of SAX.

One such segmentation algorithm is voting experts (VOX). Developed by Cohen et al., VOX is an automated process for segmenting continuous strings of symbols into words (17). The method operates on the theory that entropy within a word is much lower than entropy between words in any corpus. For example, given the phrase “thecar,” VOX will output two segments, “the” and “car” because these two words are much more common in the English language than any other segmentation, that is, “th” “ecar” or “thec” “ar.” This concept can be directly applied to SAX output to form drive segments.

Figure 7 shows an example of VOX output applied to SAX-reduced data for a portion of a drive. Specifically, the segment contains data from a drive to the grocery store. In the portion shown in the figure, a driver takes a left turn onto a segment of rural highway, accelerates to speed, maintains a steady speed, slows near the exit for the store, and finally turns into the store's parking lot. The top plot in the figure shows each segment plotted by latitude and longitude, the middle plot shows speed and acceleration changes during and between each segment, and the bottom table shows the SAX output for each segment paired with a qualitative description. This type of segmentation does not simply divide the drive by changes in speed and acceleration; it identifies natural boundaries between segments relative to the whole data set.

The output of SAX and VOX (SAX-VOX) is a set of words corresponding to a segment of speed and acceleration data. Figure 7 shows that these driving words correspond to segments of interest. Thus, it may be possible to search and classify the data by driving words, similar to a dictionary with language. The difficulty with this metaphor is that small changes in driving words may result in different classifications of qualitatively similar events. For example, a highway driving segment lasting 10 min will be defined by a different word than one that lasts 10 min and 5 s. In a sense, these words will be synonyms, but it might be advantageous to classify them as the same word.

Conclusion

The data reduction and preliminary analysis presented here suggested a considerable shift in naturalistic driving data analysis. SAX time series reduction provides an efficient and comprehensive method to transform peta-size data into manageable structures without significant loss of information or descriptive power. Data mining tools and visualizations were used to create a new data structure that facilitated a hierarchical description of driving as well as classification and comparisons between drivers and individual drives. The input parameters are flexible and can be tailored to individual applications and while the application here focused on speed and acceleration, the approach could be extended to include any measure recorded over time.

In a broader sense, the SAX-VOX approach translates driving data from a series of numbers to symbols that can be segmented into words, phrases, and even paragraphs, which create a language of driving. Although far from defining this language, the current work has demonstrated that SAX time series analysis significantly supports, enhances, and extends current analysis. Perhaps most important, the level of data reduction achieved by SAX will allow analysts and researchers to share data efficiently and go beyond the event-based focus of most naturalistic data analysis.

Acknowledgments

This work was sponsored by a grant from the National Institutes of Health.

References

1. Dingus, TA.; Klauer, SG.; Neale, VL.; Petersen, A.; Lee, SE.; Sudweeks, J.; Perez, MA.; Hankey, J.; Ramsey, D.; Gupta, S.; Bucher, C.; Doerzaph, ZR.; Jermeland, J.; Knipling, RR. The 100-Car Naturalistic Driving Study, Phase II—Results of the 100-Car Field Experiment. U.S Department of Transportation; 2006.
2. Gordon, TJ.; Kostyniuk, LP.; Green, PE.; Barnes, MA.; Blower, D.; Blankespoor, AD.; Bogard, SE. Transportation Research Record: Journal of the Transportation Research Board, No 2237. Transportation Research Board of the National Academies; Washington, D.C.: 2011. Analysis of Crash Rates and Surrogate Events: Unified Approach; p. 1-9.
3. Guo, F.; Klauer, SG.; Hankey, JM.; Dingus, TA. Transportation Research Record: Journal of the Transportation Research Board, No 2147. Transportation Research Board of the National

Academies; Washington, D.C.: 2010. Near Crashes as Crash Surrogate for Naturalistic Driving Studies; p. 66-74.

4. Jovanis, PP.; Aguero-Valverde, J.; Wu, KF.; Shankar, V. Transportation Research Record: Journal of the Transportation Research Board, No 2236. Transportation Research Board of the National Academies; Washington, D.C.: 2011. Analysis of Naturalistic Driving Event Data: Omitted-Variable Bias and Multilevel Modeling Approaches; p. 49-57.
5. Dingus TA, Neale VL, Klauer SG, Petersen AD, Carroll RJ. The Development of a Naturalistic Data Collection System to Perform Critical Incident Analysis: An Investigation of Safety and Fatigue Issues in Long-Haul Trucking. Accident Analysis and Prevention. 2006; 38(6):1127–1136. [PubMed: 16806028]
6. Neale, VL.; Dingus, TA.; Klauer, SG.; Sudweeks, J.; Goodman, MJ. Proc, 19th International Technical Conference on the Enhanced Safety of Vehicles. FHWA, U.S. Department of Transportation; 2005. An Overview of the 100-Car Naturalistic Study and Findings.
7. Dozza M, Bargman J, Lee JD. Chunking: A Procedure to Improve Naturalistic Data Analysis. Accident Analysis and Prevention. 2013; 58:309–317. [PubMed: 23000042]
8. Lin J. Experiencing SAX: A Novel Symbolic Representation of Time Series. Data Mining and Knowledge Discovery. 2007; 15(2):107–144.
9. Duchene, F.; Garbay, C.; Rialle, V. Mining Heterogeneous Multivariate Time-Series for Learning Meaningful Patterns: Application to Home Health Telecare. IMAG; Grenoble, France: 2004.
10. McGovern, A.; Rosenthal, D.; Kruger, A.; Beaton, M.; Brown, R.; Droegemeier, K. Understanding the Formation of Tornadoes through Data Mining. Proc, 5th Conference on Artifcial Intelligence and Its Applications to Environmental Sciences at the American Meteorological Society Annual Conference; San Antonio, Tex.. 2007;
11. Lkhagva, B.; Kawagoe, K. 22nd International Conference on Data Engineering Workshops (ICDEW '06). IEEE; New York: 2006. New Time Series Data Representation ESAX for Financial Applications; p. 17-22.
12. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.
13. Klauer, SG.; Dingus, TA.; Neale, VL.; Sudweeks, JD.; Ramsey, DJ. Report DOT HS 810594. NHTSA; Washington, D.C.: 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data.
14. Keogh, E. Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems. Proc, 6th IEEE International Conference on Data Mining; Hong Kong, China. New York: IEEE; 2006.
15. Johnson S. Hierarchical Clustering Schemes. Psychometrika. 1967; 32(3):241–254. [PubMed: 5234703]
16. Beeferman D, Berger A, Lafferty J. Statistical Models for Text Segmentation. Machine Learning. 1999; 34(1):177–210.
17. Cohen PR, Adams N, Heeringa B. Voting Experts: An Unsupervised Algorithm for Segmenting Sequences. Intelligent Data Analytics. 2006; 11(6):607–625.

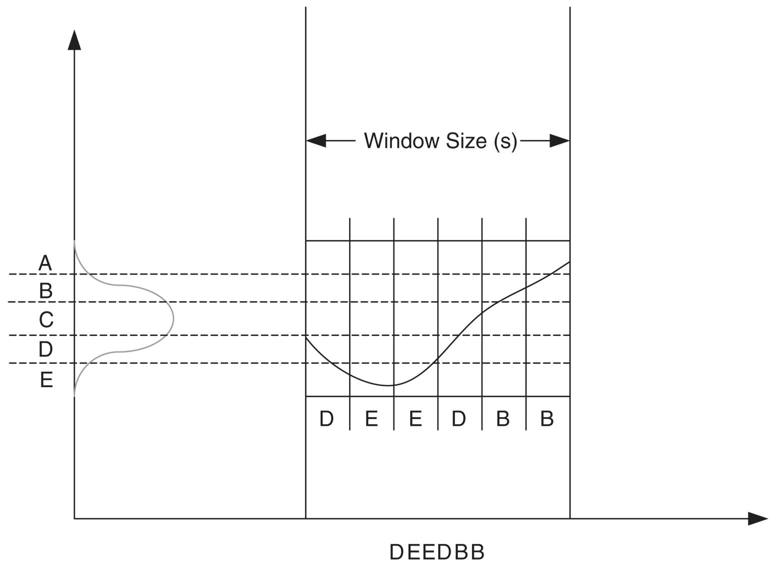


Figure 1. SAX reduction for single window of data with no overlap, letter size of five, and word length of six

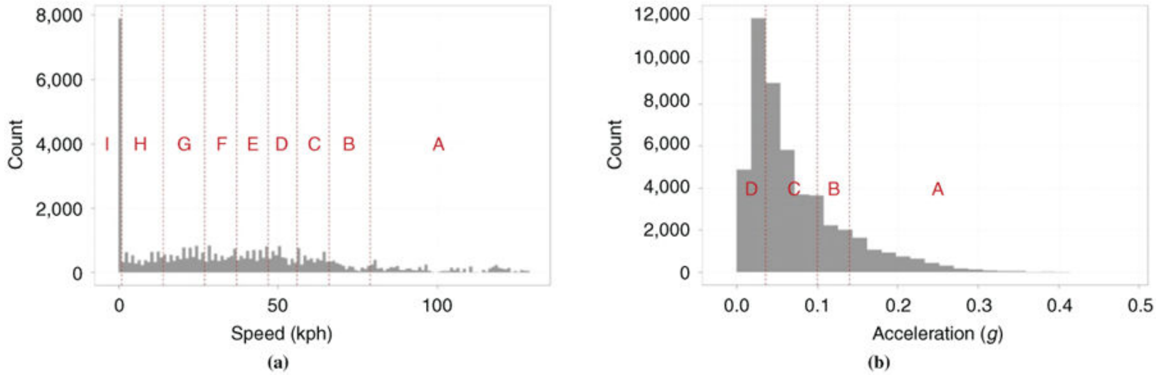


Figure 2. Histograms of sample data and letter assignments for (*a*) speed and (*b*) acceleration (kph = kilometers per hour)

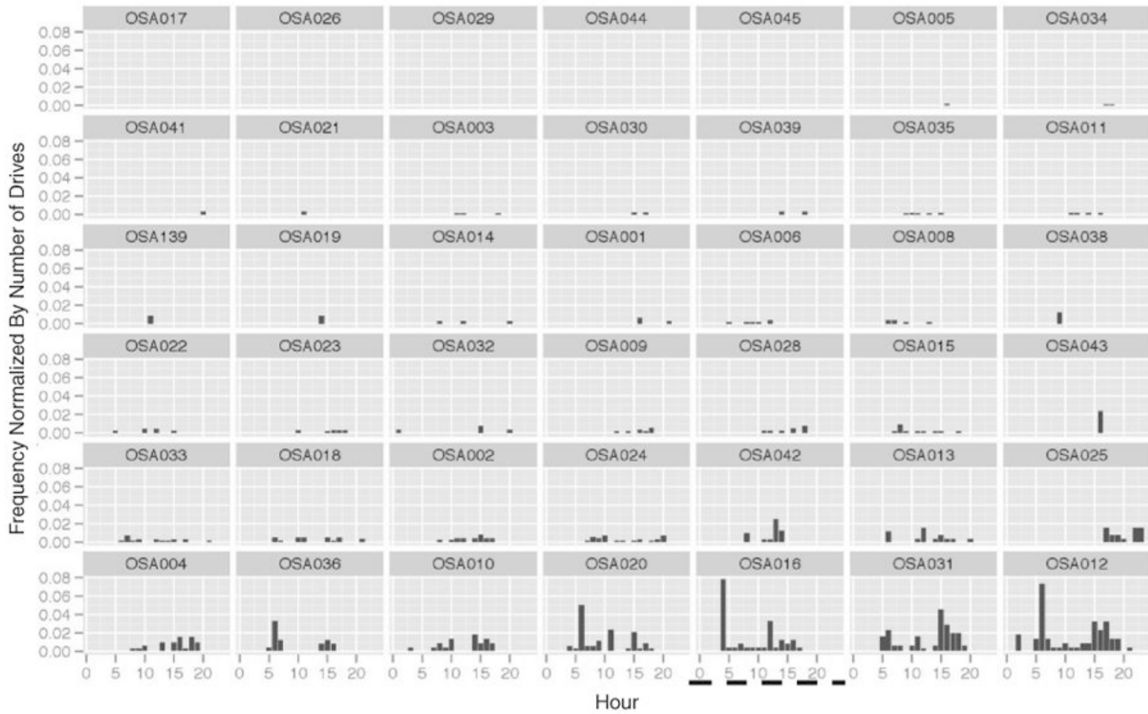


Figure 3. Histograms of peak lateral acceleration in high-speed events

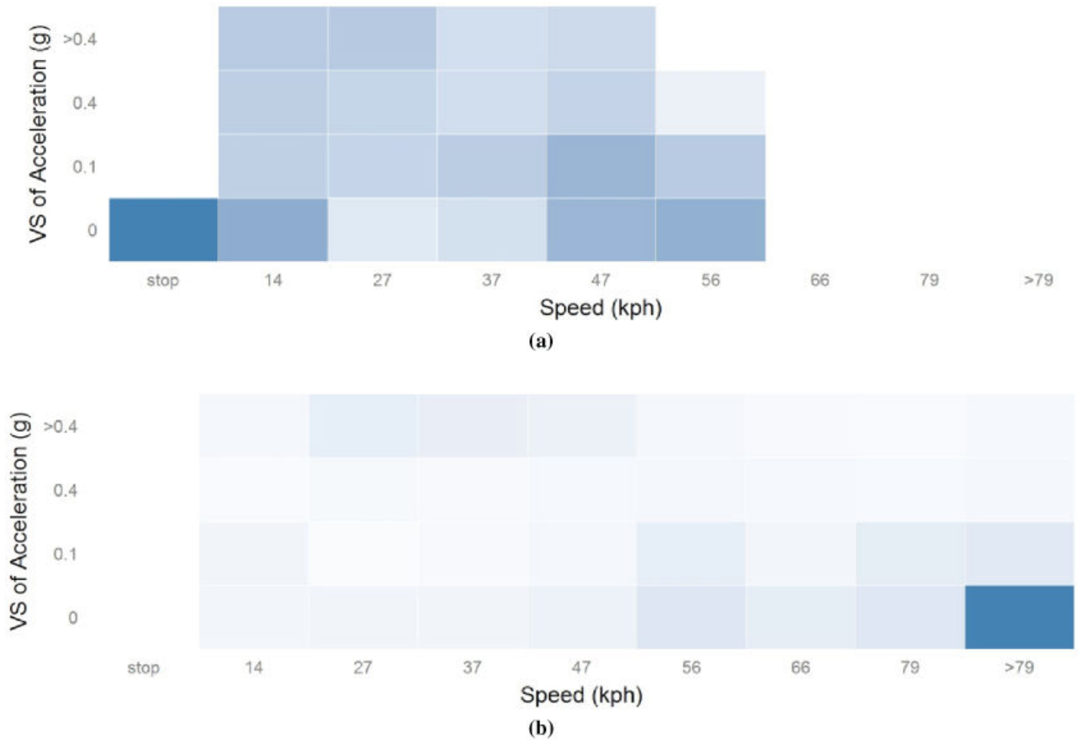


Figure 4. Two-dimensional density plot of speed and combined acceleration values for two full drives: (*a*) residential drive and (*b*) highway drive (VS = vector sum, the square root of the sum of squared lateral and longitudinal acceleration)

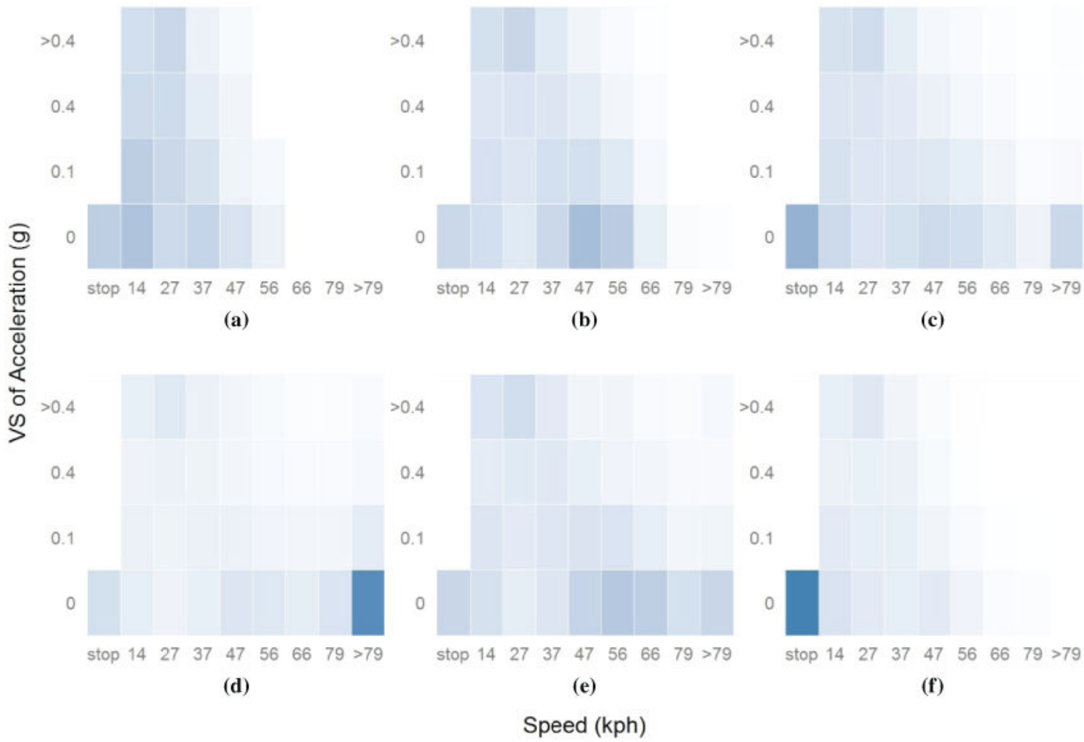


Figure 5. Two-dimensional density plots for each drive type (cluster) for Participant 1: (*a*) Cluster 1, (*b*) Cluster 2, (*c*) Cluster 3, (*d*) Cluster 4, (*e*) Cluster 5, and (*f*) Cluster 6

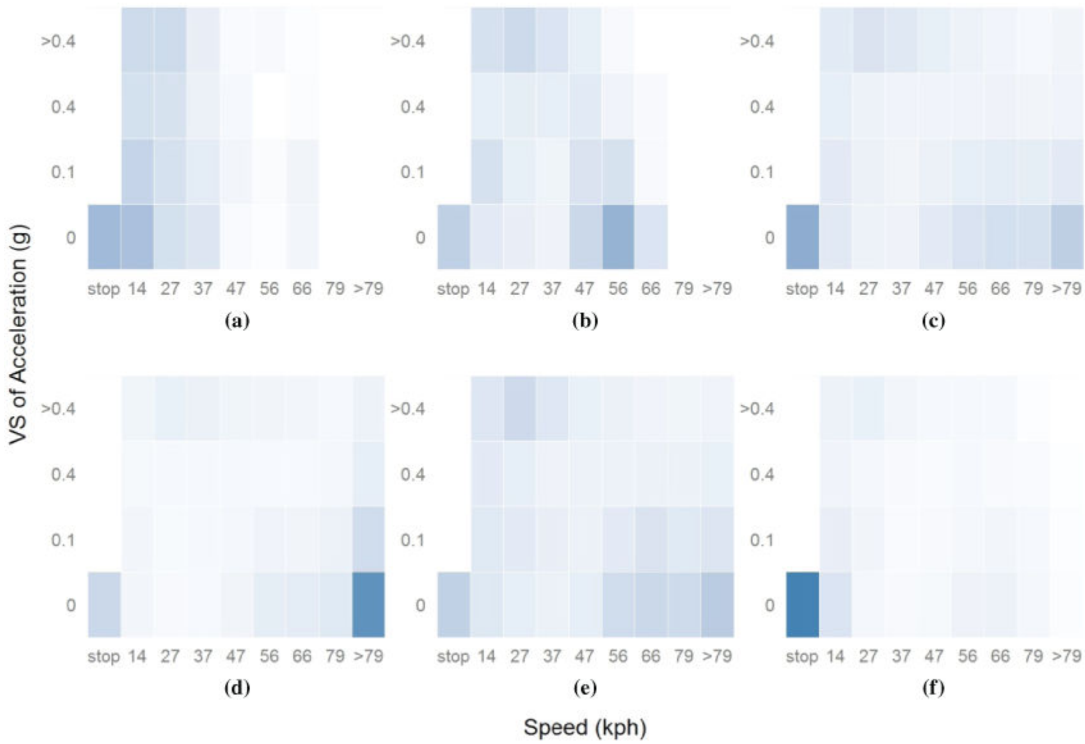
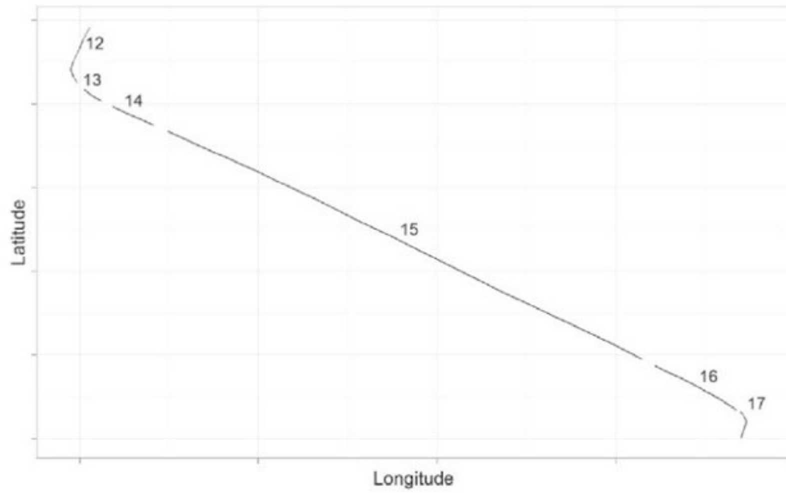
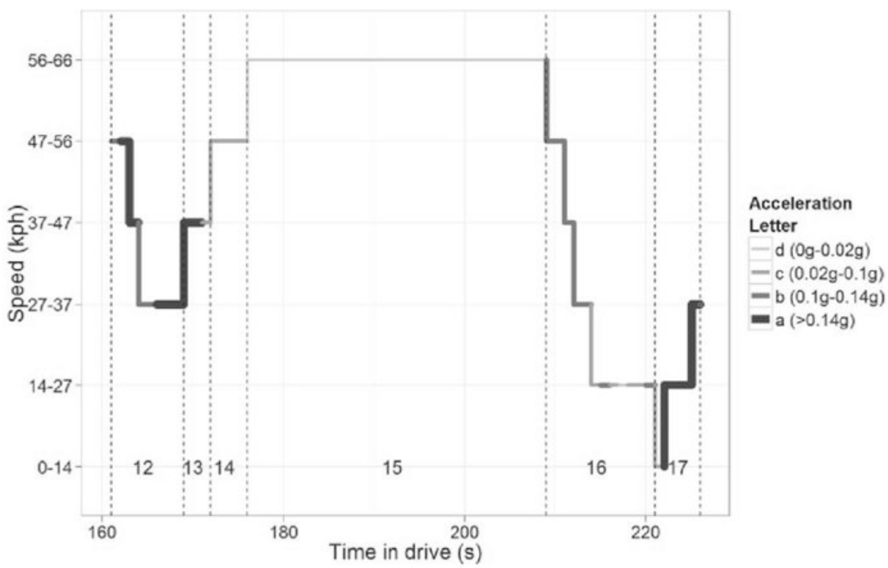


Figure 6. Two-dimensional density plots for each drive type (cluster) for Participant 10: (*a*) Cluster 1, (*b*) Cluster 2, (*c*) Cluster 3, (*d*) Cluster 4, (*e*) Cluster 5, and (*f*) Cluster 6


12
13
14
15
16
17
Latitude
Longitude

(a)


56-66
47-56
37-47
27-37
14-27
0-14
Speed (kph)
12
13
14
15
16
17
160
180
200
220
Time in drive (s)
Acceleration Letter
d (0g-0.02g)
c (0.02g-0.1g)
b (0.1g-0.14g)
a (>0.14g)

(b)

Segment	Speed Word Acceleration Word	Description
12	ddefffff baabbaaa	Left turn onto U.S. route
13	eee aac	Accelerate to speed
14	dddd cccc	Coast to speed limit
15	ccccccccccccccccccccccccccccccccc ddddddddddddddddddddddddddddddddd	Maintain steady straight drive
16	ddeffggggggg bbbbbcbcdccb	Decelerate in preparation for right turn
17	hgggff caaaab	Right turn into parking lot

(c)

Figure 7.

Visualizations and descriptions of VOX output for subset of data from single drive: (*a*) segments by GPS location and (*b*) segments by speed and acceleration.

(*c*) word output with paired real-world actions (segment numbers are maintained throughout three representations as reference points).

Table 1

Sample of SAX-Reduced Data Set from Single Subject and Drive

Time	Speed Letter	Combined Acceleration Letter	Lateral Acceleration Letter	Latitude	Longitude	Elevation	Heading	Throttle
57	d (47 km/h)	d (0 *g*)	d (0 *g*)	41.68	−91.60	213.2	301.56	23
58	c (56 km/h)	d (0 *g*)	d (0 *g*)	41.68	−91.60	213.1	304.85	20
59	c (56 km/h)	c (0.1 *g*)	d (0 *g*)	41.68	−91.60	213.1	305.07	17.2
60	d (47 km/h)	d (0 *g*)	d (0 *g*)	41.68	−91.60	213.1	305.23	18.5
61	c (56 km/h)	d (0 *g*)	d (0 *g*)	41.68	−91.60	213.2	305.28	19.8

Note: The data set is from Subject OSA001 on Drive 277. Speed and acceleration letters have been annotated with approximate real values.