Reducing Accelerometer Data from Instrumented Vehicles

Michael O. Bishop[1], Jeffrey D. Dawson[1], Jennifer Merickel[2]
Matthew Rizzo[2]
[1]University of Iowa College of Public Health, Department of Biostatistics,
145 N. Riverside Drive, Iowa City, IA 52242
[2]University of Nebraska Medical Center, Department of Neurological Sciences,
Mind and Brain Health Labs, Omaha, NE 68198

Abstract

In on-road driving behavior studies, vehicle acceleration is sampled at high frequencies and then reduced to meaningful metrics over short driving segments. We examined road test data from 65 subjects driving over a common route, as well as driving in naturalistic situations using their own vehicle. We isolated 24-second segments, then reduced the accelerometer data via two methods: 1) standard deviation (SD) within a segment, and 2) re-centering parameter from a time series model previously developed for driving simulator data. We analyzed the data via random effects models to ascertain the intraclass correlations (ICC's) of the metrics. With and without adjusting for speed, the ICC of SD within a segment tended to be much greater than the ICC of the re-centering parameter for the segment (range: 0-30% vs. 0-1%). Also, ICC's from the naturalistic driving data tended to be greater than the fixed-route data (range: 0-27% vs. 0-9%), which could reflect individuals exhibiting their more usual driving behavior in naturalistic environments. Findings illustrate the challenges of identifying meaningful driving metrics and comparing these across different epochs, road segments and research platforms.

Key Words: Lateral Acceleration, Longitudinal Acceleration, Reliability, Cognitive Ability, Driving Metrics, Naturalistic Driving

1. Introduction

With the ever-expanding world of big data and high performance computing, researchers are now able to address scientific questions in ways that are unprecedented. One such field is that of driver behavior and vehicle telemetry. Past focuses in the realm of driver behavior have been based on driving simulators and instrumented vehicles on fixed routes. Both of these situations can provide valuable insight into driver behavior, but lack the ability to truly replicate the variety of situations that drivers face on a day to day basis.

Naturalistic driving studies are conducted to provide insight into the behavior of drivers while also observing the vehicle and surroundings in a non-obtrusive manner (van Schagen et al, 2011). With innovations in technology allowing us to monitor attributes about vehicle performance, driver behavior, and road conditions at high frequencies (e.g., at the 1 to 30 Hz), there arises the need to develop statistical tools and methods to reduce these data into meaningful metrics so that analyses can be conducted in an efficient and intuitive manner.

For this report, we utilized data from the Predictions of Driver Safety in Advancing Age: Real-World Recorders project, and attempted to explore and assess reliability of data reduction measures in fixed-route and naturalistic driving scenarios. More specifically, we aimed to reduce high-frequency acceleration data into meaningful and reliable summary metrics over short driving segments. These summary metrics could potentially be used in later models as important metrics representing an individual's acceleration profile, either to be predicted by individual demographics and cognitive factors, or to be predictive of real world events such as crashes, near crashes, and driving cessation.

2. Methods

Drivers were all 65 years old or older, and were recruited from the Omaha, Nebraska area. Our key factors of interest within the study were: *Acceleration Type, Frame Rate, Drive Type, Outcome Metrics,* and *Vehicle Speed*.

Acceleration Type was split into two possible categories for this analysis: lateral acceleration and longitudinal acceleration. Lateral acceleration accounts for the rate of change in speed in a side-to-side direction while driving, and could be interpreted as the "steadiness" of an individual within a lane while driving. Longitudinal acceleration accounts for the rate of change in speed in a forward and backward direction while driving, and could be indicative of the "aggressiveness" of a driver. However, both of these acceleration types would be affected by a number of factors beyond the driver's control, such as speed limit, traffic density, road construction, stoplights and stop signs, and other drivers.

Frame Rate was also a factor of interest, as it is a measure of the rate in which data is being received from the subject vehicle. If one could show that a reduction in frame rate yielded no loss in acceleration metric viability, then one could reduce the overall data storage and computational burden of future analyses. Within this analysis, we compared data at the 10Hz level (10 frames per second) and data that had been averaged to the 1Hz level (1 frame per second).

Drive Type identified the source of the data. Within this study, data were collected on each individual in two different methods, in a fixed-route scenario, and in a naturalistic driving scenario.

Within the fixed-route driving scenario, all 65 of the drivers used in this report participated in a drive along a specified route in Omaha, NE (see Figure 1). The drive was conducted in an instrumented vehicle with a study administrator present. For the purpose of our analysis, six driving segments (each 24 seconds long) were extracted from the overall drive and analyzed for every individual driver. These six segments were selected by meeting the following eligibility criteria: segments had to start at the same geographic location for all drivers, and a segment had to be 24 contiguous seconds where velocity was at least 10 miles per hour. Segments were selected in this manner so as to reduce variability in the data, by ensuring that all drivers were driving under similar conditions (i.e. same locations and somewhat similar speeds). As a note, specific driving tasks were given to the drivers along the route to address certain aims of the overall study, and were tied to specific geographic locations along the drive. The segments we analyzed included sections both within and between such task sections, to get a reasonable breadth in the driving experience.

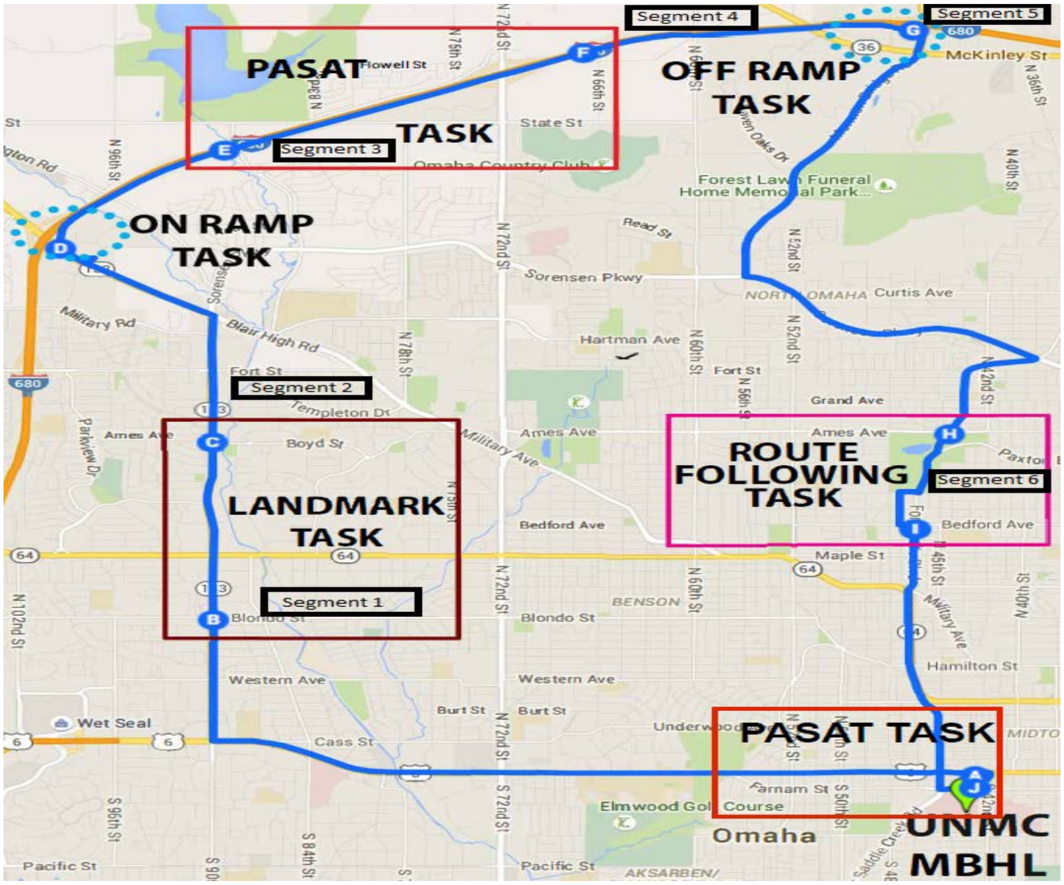

Figure 1: Identification of driving segments to be used in the fixed-route portion of the analysis. The route was located in Omaha, NE, and all drivers in the analysis participated.

Within the naturalistic driving scenario, again, all 65 analysis drivers participated, and were all based in the Omaha, NE area. Data were recorded on each individual's driving during a three month period, with subjects driving their own personal vehicles, which were outfitted with a non-intrusive set of cameras and accelerometers. Drives for this portion of the study ranged all over the continental United States. Multiple driving segments per subject were collected (Range: 2-627 per person), all of which met an eligibility criteria of having 24 contiguous seconds of velocity if at least 50 mph. This limiting criterion was imposed to make the drives as consistent as possible across and between drivers, since using a common geographic location of drives was not feasible. This was done with the hope that in individual driving behavior was not masked by differences in the environment.

Outcome Metrics were functional summaries of the acceleration sequences, we considered two metrics: Segment Standard Deviation (Segment SD) and a Time Series Centering Parameter. The Segment SD measures the variability of the acceleration within the specified driving segment. The Centering Parameter (denoted as the "γ_1" parameter in Dawson et al, 2010) measures the tendency of a variable to return to zero; hence, for the case of acceleration could be loosely thought of as the tendency for an individual to remain at a constant speed in a given direction. This metric was originally built for lane position data, and higher values suggest a greater tendency to return to zero (i.e. the tendency for a driver to return to a default lane position).

Vehicle Speed was considered within the analysis as a variable to further account for driving conditions that may exhibit differences in driving acceleration behavior. Speed was measured in miles per hour.

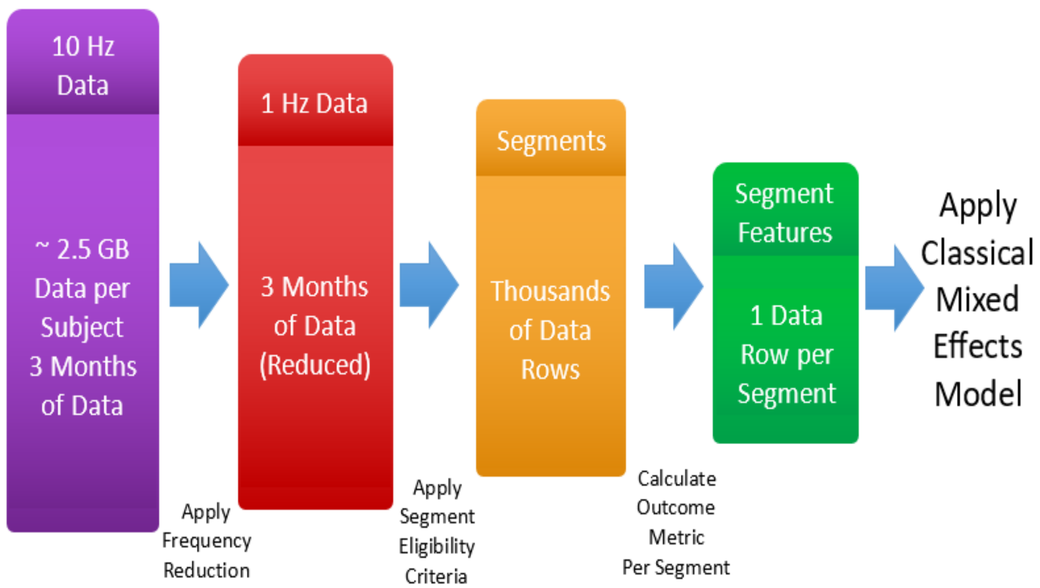

Figure 2: Description of the data flow during the analysis for an individual in the naturalistic driving setting.

Figure 2 illustrates the data flow of the analysis, with regards to filtering, eligibility criteria, and metrics to reduce data into one observation per segment per subject. After reducing the data in this manner, classical mixed effects models were applied. Outcome metrics, if describing individual driving behavior properly, would be expected to be "reliable" in the sense of being correlated within subjects. Therefore, the Intra-Class Correlation (ICC) estimates (Koch, 1982) obtained from the fitted mixed effects models were used to evaluate the success of the data reduction across the levels of the factors described above. Higher values of ICC were indicative of higher reliability, and metrics with high ICC are likely to be useful for future studies involving subject level data.

3. Results

Since we considered five different factors, (acceleration type, effect of speed adjustment, data frequency, drive type, and the data reduction outcome metric), there were $2^5=32$ different combinations to consider. Not all combinations have been reported here, however primary comparisons of interest may be found in tables one through three.

Tables 1-3 show ICC estimates that ranged between 0 and 27%. A red shading indicates values of ICC that were less than or equal to 3%, light green indicates an estimated ICC of between 3% and 10%, and a darker green is indicative of an estimated ICC greater than 10%.

Table 1 illustrates that for the segment SD metric in the fixed route, the ICC's from the 1Hz data tended to be higher than those of the 10Hz data (range: 0-10% vs. 0-2%). Speed played an inconsistent role in the estimation of ICC, as in some cases its effect seemed to increase the estimated ICC, whereas in other cases it seemed to decrease it. Additionally, the segment SD data reduction outcome metric tended to have much greater ICCs than those based on the centering parameter (range: 0-28% vs. 0-3%). We also note that the ICCs from the naturalistic driving data tended to be much greater than the fixed route data (range: 0-28% vs. 0-10%).

Table 1: Segment SD – Fixed Route 10 vs. 1 Hz

Acceleration Type	Speed	Estimated ICC 10 Hz	Estimated ICC 1 Hz
Lateral Acceleration	Raw	0.00%	6.50%
Lateral Acceleration	Adjusted	1.94%	0.00%
Longitudinal Acceleration	Raw	0.94%	7.15%
Longitudinal Acceleration	Adjusted	2.14%	9.92%

Table 2: Centering Parameter – Fixed Route 10 vs. 1 Hz

Acceleration Type	Speed	Estimated ICC 10 Hz	Estimated ICC 1 Hz
Lateral Acceleration	Raw	2.68%	1.23%
Lateral Acceleration	Adjusted	1.96%	0.08%
Longitudinal Acceleration	Raw	1.64%	0.99%
Longitudinal Acceleration	Adjusted	1.75%	1.08%

Table 3: Centering Parameter – Fixed Route 10 vs. 1 Hz

Acceleration Type	Speed	Estimated ICC Fixed Route	Estimated ICC Naturalistic
Lateral Acceleration	Raw	6.50%	14.92%
Lateral Acceleration	Adjusted	0.00%	14.59%
Longitudinal Acceleration	Raw	7.15%	27.54%
Longitudinal Acceleration	Adjusted	9.92%	27.19%

4. Discussion

Based upon the results of the analysis, three main conclusions may be drawn. First, the ICCs from the 1Hz data tend to be higher than those from the 10 Hz data, suggesting that taking averages of 10 frames may reduce the random noise for a clearer driver profile in the data frequency reduction. Such a result is helpful in that we may be able to save on data storage requirements and computation time if we focus on 1 Hz data. However, there may be other aspects of the data, such as the patterns of acceleration that occur around crashes or near crashes, which might benefit from having 10 Hz data. Further work would include identifying novel smoothing methods to reduce data frequency, and identifying an optimal frequency level for outcome metrics.

Since the ICC of SD within a segment tended to be much greater than the ICC of the centering parameter, it would suggest that segment SD is a more reliable metric for summarizing accelerometer data. This is reasonable, as the Centering Parameter was built for modeling lane position in a simulator rather than accelerometer data in actual vehicles. In future work, we hope to process our video data using computer vision algorithms, so that we can estimate lane position data for instrumented vehicles and for naturalistic driving, so that we can assess reliability of the centering parameter for lane position in the real world.

The ICCs from the naturalistic driving data tended to be much greater than the fixed route data, suggesting a possible change in driver behavior between the fixed route and naturalistic driving settings. Since the subjects were in their own vehicles driving as they would normally for the naturalistic setting, it is reasonable to assume that they might behave differently (possibly less uniformly) than they would in driving on a pre-determined fixed-route with a study administrator sitting in the passenger seat.

One limitation of our comparisons between naturalistic and fixed-route driving is that we applied different eligibility criteria between those two driving conditions. Further work is needed to fully assess the effects of the difference in speed criteria between the two data sources, as well as the effect of being able to map driving segments to specific geographic locations shared by all subjects only in the fixed route data.

In preliminary analyses, we did not find any of our data reduction metrics to be associated with demographics or cognitive factors measured by neuropsychological tests. Thus, even though we found some situations with modest reliability (ICC values close to 30%), this may still not be high enough for practical purposes.

In conclusion, we have found promising indicators that Segment SD may be used as a possible functional summary of acceleration sequence, and that its reliability may be increased by looking at data from a reduced frame rate. We also found reason for looking at naturalistic driving study data as a more reliable source of individual driving behavior than that of the more traditional fixed-route driving study.

Acknowledgements

We thank all members of our study team, past and present, as well as the study subjects. The work was supported the NIH/NIA award RO1 AG017177. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Dawson, J.D., Cavanaugh, J.E., Zamba, K.D. & Rizzo, M. (2010). Modeling Lateral Control in Driving Studies. *Accident; Analysis and Prevention*, 42(3), 891-897. http://doi.org/10.1016/j.aap.2009.04.022

Koch, G.G. (1982). Intraclass correlation coefficient. In Samuel Kotz and Norman L. Johnson. *Encyclopedia of Statistical Sciences*, 4. New York: John Wiley & Sons. pp. 213-217.

Van Schagen, I., Welsh, R., Backer-Grøndahl, A., Hoedemaeker, M., Lotan, T., Morris, A., Sagberg, F., Winkelbauer, M. (2011). *Towards a large-scale European Naturalistic Driving study:final report of PROLOGUE. PROLOGUE Deliverable D4.2*. SWOV Institute for Road Safety Research, Leidschendam, The Netherlands.