

EVENT-BASED DRIVER PERFORMANCE ASSESSMENT

Erwin R. Boer
Erwin R. Boer Consulting
5434 Renaissance Ave.
San Diego, CA 92122
E-mail: erwinboer@san.rr.com

Nicholas J. Ward
HumanFIRST Program
ITS Institute, University of Minnesota
111 Church Street, SE
Minneapolis, MN 55455
E-mail: nicw@me.umn.edu

Summary: Driving is an intermittent control task during which drivers manage their work across a number of driving and non-driving sub-tasks. The multitasking nature forces drivers to adopt situated safety margins (tolerances) in terms of time-headway (THW), time-to-collision (TTC), time-to-line-crossing (TLC), etc. Unacceptable situations (i.e. violations of safety margins defined here as *events*) are characterized by corrective maneuvers or actions. The frequency with which corrective actions are necessary (bandwidth), the hazard level of the situation that inspired these actions (performance), and the rate and magnitude of the responses to these situations (effort) provide a rich signature of how drivers manage their task(s). We hypothesize that drivers' perception of performance and effort are founded in the characteristics of experienced events. This is explored by comparing driving characteristics of bus drivers who drive on the shoulder of a highway with and without the support of a haptic lateral support system (LSS). Subjective performance and effort scores extracted from a usability questionnaire and objective ones from our event-based analysis show highly significant correspondence when comparing supported versus unsupported driving. This provides validating support for the adopted event-based approach. The proposed approach offers not only sensitive metrics of driver performance and effort to evaluate ITS applications but also explanatory power by exposing the various strategically different ways drivers are affected by these systems. This method of quantifying and analyzing driver data affords new opportunities to evaluate driver responses to ITS applications.

MOTIVATION & GOAL

Exiting methodologies for gaining an understanding of the impact of ITS applications on driving performance often lack strong correlations with subjective metrics and insufficient richness to characterize and understand individual differences. Objective metrics of driving performance (P) and effort (E) that correlate highly with subjective scores are desirable because this demonstrates mutual validity. Aggregate metrics often lack differentiating power because linear time averaging washes out the important and generally infrequent events that signify drivers' limitations in maintaining safety margins. This shortcoming is recognized by many researchers

and is often treated by deliberately packing a driving session with demanding activities and boosting environmental disturbances. Rather than filling the trip with events, can one simply focus on the events that naturally occur in the first place? It is well known that memory is most readily created in association with an emotionally charged event (e.g., one that violates safety margins and requires a significant response). Given that memory of a trip is needed in order to provide subjective P and E ratings, driving should be evaluated along the dimension of event time rather than linear time. To demonstrate these principles, an event-based approach to driving assessment is presented.

ANALYSIS CONTEXT

The driving context assessed here is driving an 8.5-foot-wide empty city bus on a 9-foot-wide right shoulder of a two-lane highway with and without a lateral support system (Ward et al., 2003). Although the force magnitude produced by the LSS is such that the bus could drive autonomously even without hands on the steering wheel, the force can be over-ridden by the driver, and the system is designed and implemented as an assisting rather than a control system. Moreover, many naturally occurring events make unsupervised autonomous control undesirable (e.g., dips in the shoulder where drains are located, guardrails hugging the shoulder tightly, system errors, other vehicles encroaching into the shoulder). Therefore, all drivers kept their hands on the wheel. Each time drivers completed a trip down the same stretch of road, under very similar traffic conditions, they filled out a usability questionnaire (Ward et al., 2003). In a counterbalanced order, they made this drive with and without the LSS.

EVENT DEFINITION AND PERFORMANCE PLUS EFFORT METRICS

An event is defined as a situation that requires a corrective response (i.e., a violated safety margin). Here, an event is defined by a change in steering rate (i.e., a peak in steering amplitude motivated by the fact that existing objective methods, that are also sensitive to peaks in the steering profile, such as steering entropy (Nakayama et al., 1999) and steering reversal rates (MacDonalds et al., 1980) appear most successful in quantifying driver workload). The event time interval associated with an event ranges from the previous to the next event mark. The following three *effort* measures are extracted from each event:

- ***DeltaTime***: Time since last event. The inverse, or bandwidth, is used because a drop in bandwidth corresponds to a drop in effort (i.e., consistent with following two measures).
- ***SteerRate***: Maximum steering rate within the event time interval.
- ***DeltaSteer***: Maximum of the two amplitudes between two consecutive peaks in the steering profile over the event interval. Note that event interval spans three peaks.

These response measures are used to infer the amount of effort associated with the lane-keeping sub-task. Besides response characterization, it is also necessary to characterize the event in terms of hazard level. The following two *performance* measures are computed over the event interval; they are used to infer the safety margins that drivers are willing to adopt:

- ***InvTLC***: Minimum absolute time to line crossing (TLC). A virtual line half a bus width outside the shoulder was adopted instead of the true shoulder boundaries to avoid

numerical singularities and inconsistent changes in the sign of InvTLC in case the bus crosses the boundary (an event that happens quite frequently).

- **PredLatPos:** Maximum predicted lateral position. The prediction interval is 1s.¹

Each trip was quantified by the 75-percentile values of each of these five metrics. The 75% was used because strong events are expected to correlate higher with subjective metrics than minor ones. Furthermore, an event has to exceed the noticeability threshold to be tallied into subjective P and E scores (i.e., be associated with an emotion). Higher percentiles than 75% (i.e., even more severe events) are most likely correlated with strong external disturbances, such as cars cutting into the shoulder, which would occur at different rates for different subjects and would thus unjustly bias the scores.

USABILITY SCORES

The adopted usability questionnaire includes 16 questions, which subjects score on a scale of 1 (completely disagree) to 5 (completely agree).² Each question was assigned as either P- or E-related, and scores were changed to signed scores (-2 to 2). The raw subjective E score was simply the sum of all eight E questions, and the P score was similarly figured (see Figure 1 inserts).

RESULTS AND DISCUSSION

Event based metrics for each subject were placed in a radar diagram (Fig.1). A complete data set was available for nine subjects. Only data for six (2,4,6,7,11,12) are shown. The other three subjects (3,5,10) are not featured in these figures because their P and E changes were less pronounced or similar to those featured. To demonstrate the mutually explanatory power between objective metrics and subjective ratings, subject 4 was selected because he showed the least change in objective metrics (a small decrease in E paired with a slight increase in P—the ideal situation, which he “recognized” as reflected in his subjective scores).

The objective values along each radial axis in Figure 1 are obtained through a 5-step process: (i) normalize each metric by dividing it by the group median across both conditions (i.e., LSS on/off), (ii) subtract 1 from each normalized metric to obtain deviations from group “average,” (iii) per condition normalize the deviations in each metric by dividing by the group median deviation across both conditions, (iv) scale these normalized deviations such that the deviations for subjects 2,3,6,7,11 and 12 (i.e., those in radar plots) range between 0 and 1, and (v) add one to each scaled normalized deviation. Thus, a value of 1 along each of the radial axes indicates the median P or E value. A value greater than 1 indicates that P for that subject was worse or E greater than that for the “median subject.” In Figure 2, the objective dP value is simply the difference in sum of the two P scores divided by 2 between supported and unsupported driving.

¹ The haptic lateral support provided by the system was based on the same information as was used to compute the PredLatPos (i.e. the LSS assessed the state of the vehicle about 1s ahead to compute its steering torque). The 1s prediction / preview interval was chosen because it takes about 1s for a steering input to be noticed in a yaw change of the bus.

² Based on their experience with the system they were asked the following. “I view a system that supports lane keeping as: (1) A system to improve safety (P), (2) A system to enhance performance (P), (3) A source of confusion or distraction (E), (4) Useful in urban areas (P), (5) Useful in rural areas (P), (6) Useful on highways (P), (7) Increasing mental (and visual) effort (E), (8) Increasing driver comfort (P), (9) Creating difficulties on curves (E), (10) Encouraging faster than normal speeds (P), (11) Making the driver less vigilant (E), (12) Making the driver less stressed (E), (13) Making the passengers less stressed (P), (14) Unreliable in its operation (E), (16) Requires specialized training (E).

Thus, a positive dP indicates an increase in P with support. The change in effort dE is computed similarly based on the three effort metrics. The subjective (dE,dP) points in Figure 2 were obtained by scaling the raw subjective dE and dP such the subjective (dE,dP) point (squares in Figure 2) and the final objective (dE,dP) point (diamonds in Figure 2) for subject 6 (upper right in Figure 2) fall in each other's vicinity.

To facilitate relating the unit radial distance in the radar plots to meaningful units, the median values for each metric in each condition are shown in the following table. The last row shows the median values over both conditions (i.e., the one that exactly matches unity in the radar plots).

LSS	Effort Metrics			Performance Metrics	
	Median 75% of Delta Steer [Deg.]	Median 75% of Bandwidth [1/s]	Median 75% of Steering Rate [Deg./s]	Median 75% of Pred. Lat. Pos. [m]	Median 75% of Inverse TLC [1/s]
OFF	16.0517	0.645161	18.9127	0.2878	0.0952
ON	19.808	0.689655	21.1438	0.2298	0.0842
ON+OFF	16.605	0.678161	19.28225	0.2874	0.0907

Based on the Wilcoxon signed rank test of equality of medians, a marginally significant increase in Steering Rate ($p = 0.0547$) and Delta Steer ($p = 0.0742$) with lateral support is observed. Subjects produced faster and stronger steering corrections with the system. This may have been beneficial if initiation was guided by the system. This is supported by the fact that the bandwidth with the LSS is very similar for most subjects, suggesting that subjects adopted the natural bandwidth of the LSS steering torques.

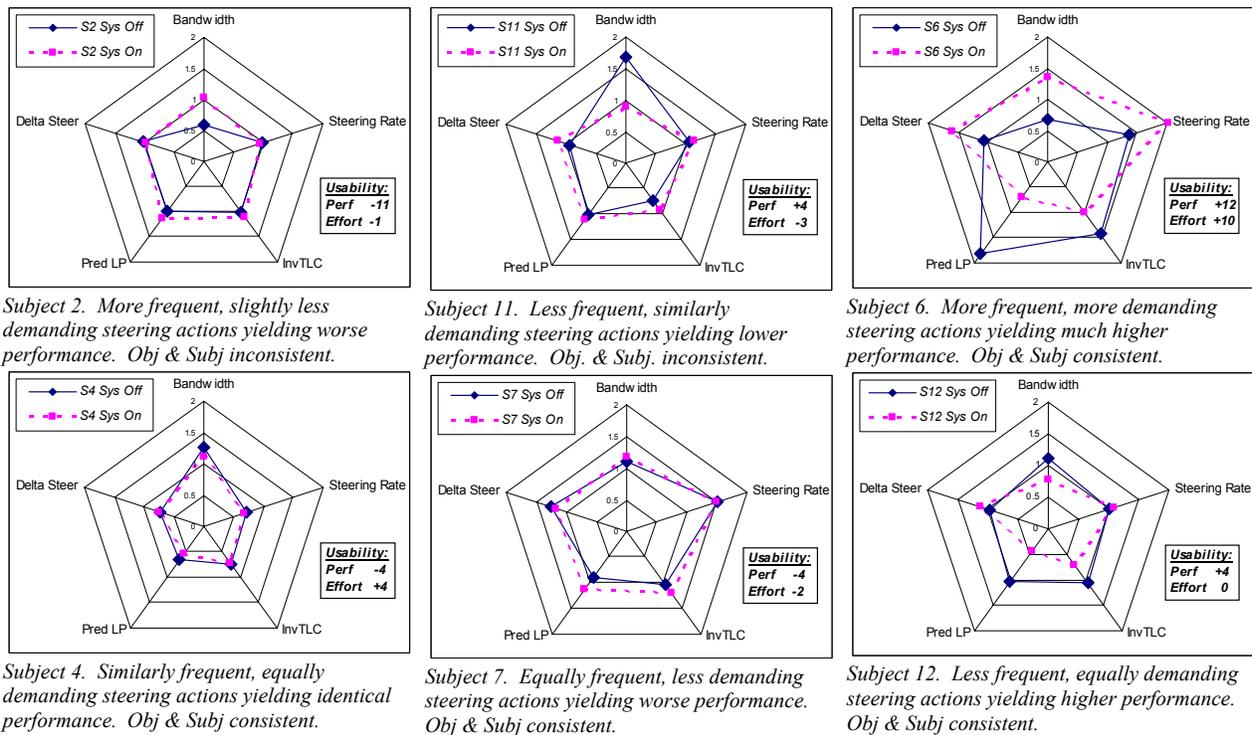


Figure 1. Group normalized plots of the event-based P (2 at base) and E (3 at top) metrics

The fact that the radar graphs show all possible changes in size, as well as shifts up or down, clearly demonstrates its power to differentiate the many ways in which support systems affect drivers' P and E. Four fundamental changes from solid lines with diamonds (without LSS) to

dashed lines with squares (with LSS) are interpreted as follows. The LSS results in: (i) contraction (with less E greater P is achieved – Ss 4 and 12), (ii) expansion (with more E worse P is achieved – S 2), (iii) down shift (with less E lower P is achieved – S 7), and (iv) up shift (with more E greater P is achieved – S 6). Other subtleties are also exposed such as a shift from highly frequent small correction to less frequent larger corrections (e.g., S 11). This demonstrates the importance of the adopted event-based approach, namely that no single measure should be used to assess effort (bandwidth and magnitude are equally important in delineating response strategies to support systems). Space restrictions limit further exposition of the way radar plots expose the intricacies of drivers’ interactions with support systems.

Significance in the degree of correspondence between objective and subjective metrics (i.e., change-vectors in the P-E space) is computed through the following process: (i) we can only compare the signs of dE and dP and not the magnitudes because they are incomparable³ (i.e., the only meaningful piece of information is the quadrant in which the change-vector points), (ii) the Null Hypothesis states that the subjective change-vectors are independent of the objective ones, (iii) the probability that a subjective change-vector points in the same quadrant as the objective one is 1/4, (iv) compute the probability that only two or fewer subjective change-vectors point to a different quadrant than the corresponding objective ones (i.e., Ss 2 and 11 as indicated by dashed lines between the objective and subjective change points in Fig 2), (v) if this probability is less than 0.05 we can reject the Null Hypothesis. The probability is computed with the

following formula $p = \sum_{k=0}^2 \binom{9}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{9-k}$ resulting in $p = 0.001343$ which indicates high

significance; thus the subjective and objective metrics share a common source. This source is hypothesized to be the frequency and magnitude of substantial (i.e., the adopted 75-percentiles) events. The number of events that exceed expectations is expected to determine the perceived effort and performance.

The adopted event-based approach clearly demonstrates that the LSS produces a full spectrum of effects on driver performance and effort as indicated by the fact that all four

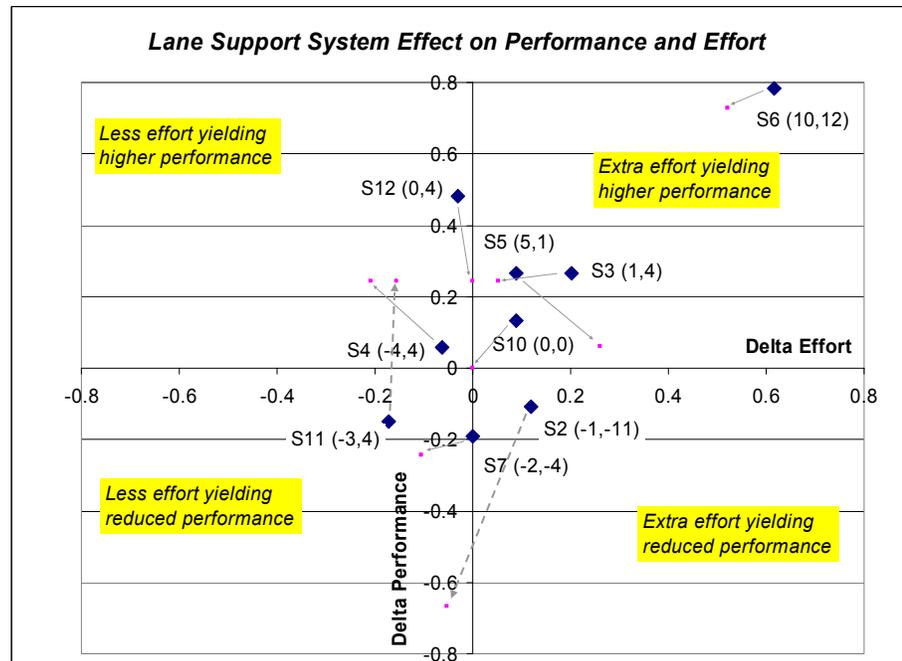


Figure 2. Correspondence between relative (i.e., w & w/o LSS) objective and subjective performance and effort metrics.

³ The objective and subjective measures for change in performance and effort were obtained by combining multiple scores or measure. The weights used were unity for lack of support to adopt a different weighting scheme.

quadrants in Figure 2 hold one or more drivers (this variety in ways the LSS influences drivers is consistent with the fact that bandwidth, PredLP, and InvTLC alone failed to reach significance across all subjects. In further analyses, the motivation for why subjects ended up in their respective quadrants will be investigated by taking a closer look at what questions in the usability questionnaire received what answers. The fact that the usability scores and the objective scores for each driver compare so well strongly supports the interpretation of these objective metrics and the strategy of looking at each individual driver. Many existing methods force group-based analyses because of a lack of strong correlation between objective and subjective metrics, thereby invalidating the subjective interpretations of objective results.

REFERENCES

- Nakayama, O, Futami, T., Nakamura, T., and Boer, E.R. (1999). *Development of a steering entropy method for evaluating driver workload*. SAE technical paper series (1999-01-0892). Warrendale, PA: SAE International.
- MacDonalds, W.A. and Hoffmann, E.R. (1980). Review of relationship between steering wheel reversal rate and driving task demand. *Human Factors*, 22(6), pp. 733-739.
- Ward, N.J., Shankwitz, C., Gorgestani, A., Donath, M., Boer, E., & De Waard, D. (2003). *Bus Rapid Transit Lane Assist Technology Systems Vol 2: A Pilot Study of a Lane Support System for VRT Operations – Bus Driver Stress in Dedicated Bus Shoulders*. ITS Institute, University of Minnesota.