MITIGATING DROWSINESS: LINKING DETECTION TO MITIGATION

Chris Schwarz

Timothy L. Brown

John Gaspar

Dawn Marshall

The University of Iowa

USA

John Lee

The University of Wisconsin

USA

Satoshi Kitazaki

The University of Iowa

USA

Julie Kang

National Highway Traffic Safety Administration

USA

Paper Number 15-0453

ABSTRACT

Drowsy driving contributes towards up to 24% of crashes and near crashes observed; 886 fatal crashes per year can be attributed to drowsy, fatigued or sleeping drivers. Drowsiness mitigation technology is composed of a detection algorithm and a mitigation component. This paper is primarily concerned with the latter, specifically for a driving simulation study about mitigating drowsy driving. The study is part of NHTSA's Driver Monitoring of Inattention and Impairment using Vehicle Equipment (DrIIVE) program. The detection algorithm incorporates time series probabilistic estimation using a Hidden Markov Model, so a drowsiness prediction at any time is dependent on a previous history of observations. Two mitigation methods are designed for testing in the simulation study. One is a three stage audio/visual alert that requires a driver response through a button press. The second is a binary haptic alert that uses a vibrating seat. Additionally, each mitigation will include three varying levels of sensitivity: a nominal model, an over-sensitive model, and an under-sensitive model. These variations will expose drivers to different numbers of false alarms while also potentially missing episodes of drowsiness. Various parameters in the detection algorithm were tested and the vote thresholds of two Random Forest models were selected for variation. It was observed how these parameters affected the output of the detection and mitigation system using previously collected drowsy driving data. Three specific levels were chosen as candidates for the experiment. It is hoped that the study will answer questions about how effective a mitigation system is at changing driving performance, whether drivers willfully ignore the mitigation, and how many alerts are too many.

INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) estimates that 83,000 crashes per year and 886 fatal crashes per year can be attributed to drowsy, fatigued, or sleeping drivers (NHTSA, 2011). The 100-car naturalistic driving study found that drowsy driving contributed to 22% to 24% of crashes and near-crashes observed (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Other studies suggest that despite known dangers many drivers continue to drive drowsy and fall asleep behind the wheel (MacLean, Davies, & Thiele, 2003; McCartt, Rohrbaugh, Hammer, & Fuller, 2000). Technology may be able to address some of these risks.

Drowsiness mitigation technology consists of two subsystems, a drowsiness detection system and a driver feedback system. The drowsiness detection system or algorithm collects data from the driver or vehicle, processes this data with a detection algorithm, and makes predictions about the alertness of the driver. The feedback system activates when the detection system predicts that the driver is drowsy and alerts the driver in order to prevent a drowsiness related crash. With some exceptions, research on drowsiness mitigation technology has largely focused on the detection algorithm. This piece of the system is critical because it strongly influences drivers' trust and reliance on the mitigation technology and constrains the design space of the feedback system (Balkin, Horrey, Graeber, Czeisler, & Dinges, 2011).

Research on drowsiness detection algorithms can be differentiated by the input data, prediction algorithm, and ground truth definition of drowsiness. Input data typically consists of camera-based eye measures (Dinges & Grace, 1998; Grace et al., 1996; Ji, Zhu, & Lan, 2004), electric potential measures from the brain (Lal, Craig, Boord, Kirkup, & Nguyen, 2003; Lin et al., 2005; Wali, Murugappan, & Ahmmad, 2013), or driver input to the vehicle such as steering wheel angle (Krajewski & Sommer, 2009; McDonald, Lee, Schwarz, & Brown, 2013a; Sayed & Eskandarian, 2001). Prediction algorithms vary from simple thresholds (Dinges & Grace, 1998), to more complex graphical models (Ji et al., 2004). The ground truth definitions also vary between studies and range from general levels of drowsiness associated with lack of sleep (Sayed & Eskandarian, 2001; J. H. Yang, Tijerina, Pilutti, Coughlin, & Feron, 2009), to more episodic measures of drowsiness such as drowsiness-related lane departures (McDonald et al., 2013a). Recent research primarily focuses on innovations in the prediction algorithm dimension. One prominent development in this dimension is a transition from static prediction algorithms to time-based prediction algorithms (Ji, Lan, & Looney, 2006; G. Yang, Lin, & Bhattacharya, 2010; J. H. Yang et al., 2009). These time-based prediction algorithms allow predictions to account for well-understood temporal effects of drowsiness: for example, a drowsy driver is likely to stay drowsy and an alert driver is likely to stay alert. Additionally, they can be built around previously non-temporal (or static) algorithms to improve predictions (Ji et al., 2006, 2004). The success of these algorithms and their strong basis in the theory of drowsy driving suggests that it could be helpful to enhance other non-temporal models by incorporating them into temporal frameworks.

Mitigation systems are the critical link between the detection system and influencing driver behavior. While the detection system aims to accurately assess driver state, the aim of the mitigation system is to present driver state information to the driver in a way that is likely to persuade the driver to make choices that improve safety. This process involves the translation of the raw detection system outputs for use by the mitigation system. These systems can theoretically take many forms, from a simple audible chime or visual icon to more complex displays that relay different levels of performance or instruction to the driver. Although the same algorithm might be used across systems, the type of the interface will dictate the required adaptation of the raw data.

The topic of this paper is the design of a mitigation system to provide feedback to the driver about the system's perception of their state of drowsiness. The mitigation system should help the driver become more aware of their drowsiness. In the short term, it may help them to improve their driving performance; however, the ultimate desired effect would be to cause them to pause their trip and take a rest.

There are several drowsiness alert systems on the market currently (see Figure 1). Many are binary alerts that display a coffee cup icon and play a chime when the alert is triggered. Some systems attempt to provide a more continuous, or at least multi-level discrete, scale of drowsiness to the driver. Some systems require the driver to press a button to acknowledge the alert.

Figure 1. Example Mitigation interfaces from Volvo, Mercedes-Benz and Ford

PRIOR WORK

The NHTSA DrIIVE program focuses on the detection and mitigation of driver impairment from drowsiness and distraction. Several models were generated in phase 1 of the DrIIVE program, including a Bayesian Network, a time-to-lane-crossing (TLC) model, and a Random Forest model based on steering wheel angle (McDonald, Lee, Schwarz, & Brown, 2013b). A Random Forest model that incorporates temporal steering information into a static algorithm was trained on drowsy lane departure data (2013a), (Brown, Lee, Schwarz, Fiorentino, & McDonald, 2014).

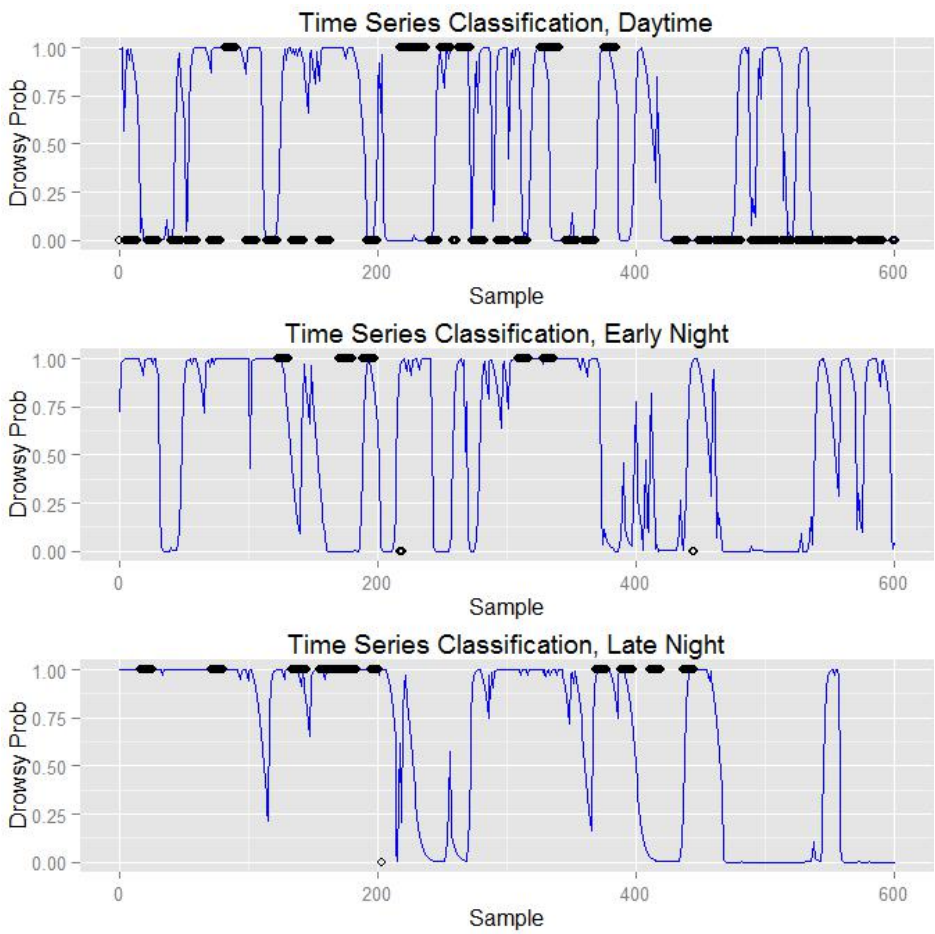

Figure 2. Example Time Series: ground truth markers and drowsiness HMM output

This initial algorithm was then extended by placing the static steering algorithm into a temporal prediction framework and exploring the effect of this approach on the timeliness of the detection algorithm (Schwarz,

McDonald, Lee, & Brown, *submitted*). The enhancements produced a set of Random Forest (RF) models that were fed into a Hidden Markov Model (HMM) capable of capturing the heuristic that an awake driver is more likely to remain awake in the near future, while a drowsy driver is likely to remain drowsy.

The Random Forest models were trained in the open source statistical software R (R Development Core Team, 2009) using the caret package (Kuhn, 2008). Normally, a classification is inferred using an RF model by running all the decision trees and using the majority vote as the output. However, if one keeps track of the vote count for each instance the model is run, then the vote count can be used as the continuous predictor in a Receiver Operator Characteristics (ROC) analysis. Then, an optimal threshold on the vote count may be computed from the ROC curve using Youden's Index (Powers, 2007). An optimal set of RF models was produced using vote thresholds of 162 votes for the steering RF model and 151 votes for the pedals RF model, where all RF models had 500 decision trees.

A Hidden Markov Model (HMM) was designed to include the effect of historical observations and accept inputs from the RF models, and was trained using the HMM library in R (Himmelmann, 2010). A regular time interval of six seconds is selected as the model frequency. Two pieces of evidence are provided, one from the steering RF classification, and the other from the pedal RF classification. The output of the HMM, shown in Figure 2 is compared to a threshold to classify each time sample as a drowsy or awake. The threshold value was selected from an ROC curve to be 0.74. The RF models along with the HMM complete the drowsiness detection algorithm.

METHOD

Two mitigation systems were designed for use as between subject conditions in a new drowsiness mitigation study. The first is a three stage audio/visual alert with driver interaction through a button. The second is a binary haptic alert that vibrates the driver's seat. Three levels of the drowsiness detection system are included in the experimental design as a between-subjects dependent variable. The three levels will include a nominal design, a design that is more sensitive, and one that is less sensitive. These levels will expose drivers to different numbers of false alerts, while perhaps also failing to detect the drowsiness in some cases.

MITIGATION DESCRIPTIONS

The audio-visual alert is a three stage warning. The threshold value used to trigger each stage is the same for each stage. If drowsiness is detected while in the nominal state of no mitigation, then a stage 1 warning is issued. This warning is a white coffee cup icon with an OK button for driver acknowledgement (Figure 3a) and an audio chime that plays when the icon appears. Once the driver presses the button, the icon is removed. The mitigation will remain in stage 1 for a minimum period of time; and during that time the detection algorithm may remain in a classification of drowsy state or return to an awake state. If the detection algorithm classification returns to awake, then the mitigation will abate after a fixed delay. However, if another drowsy episode is detected before the mitigation abates, or the drowsy state persists for 60 seconds, then the mitigation escalates to stage 2. On entry into stage 2, a stage 2 warning is issued using the visual icon in Figure 3b along with an audio beep. This icon is removed once the driver acknowledges the warning with a button press. Exactly the same logic is applied during stage 2 until the mitigation either abates back to stage 1 or escalates to stage 3. A stage 3 alert consists of the icon in Figure 3c, and a repeated audio beep. There cannot be any more escalations from stage 3, but the warning may be reissued if the drowsy state persists or soon repeats. Only incremental escalations and abatements are allowed. This mitigation has the chance to capture the driver's attention by varying the stimulus on repeated warnings; but it also has the potential to be a nuisance to a driver who is already self-aware or not drowsy.

The haptic alert is a binary alert system that provides a counterpoint to the three stage alert. It also differs in modality by providing a haptic alert through seat vibration, thus making it a more subtle, and potentially less annoying, cue. The same logic for stage escalation is applied in the binary alert to either trigger the initial alert or repeat it after 60 seconds. Once the drowsy detection expires, the mitigation naturally abates back to the nominal driving state.

MITIGATION SENSITIVITY MANIPULATIONS

A significant question addressed in this paper is: how can we vary the sensitivity of the algorithm / mitigation system? Random Forest models are especially opaque and little intuition about why a given parameter set works is available to the designer. Hidden Markov models are slightly easier to intuit, but are nonetheless complicated. There are several choices that were considered, most of which were either discarded or found to not have a significant effect upon the final outcome of the classification and mitigation performance. The Random Forest models are considered first.

(a) (b) (c)

Figure 3. Visual Mitigation Icons: (a) stage 1 interactive, (b) stage 2 interactive, (c) stage 3 interactive

Two Random Forest models are used in the drowsiness detection algorithm, one for steering and one for pedals. Each Random Forest is composed of 500 decision trees; and each decision tree may have on the order of 100 nodes. Therefore, it is not feasible to try to tune those parameters individually. One could think about a strategy of retraining new RF models with the intent of changing the sensitivity; but then there may be other performance differences between them that are confounded with sensitivity. The other parameter one can think about tuning is the voting threshold for output classification. Normally, RF models are majority rule, meaning that more than 250 trees in a 500 tree RF would have to agree to set a class output. This vote threshold number may be allowed to vary and we used it as a threshold variable in an ROC analysis in our prior modeling work. The lower the value of the vote threshold, the more trees potentially need to be run to gather up the required number of votes, thus potentially increasing the computational demand of the model by some small amount. For example, with a threshold of 100, one may have to evaluate as many of 400 of the trees to guarantee that there are not 100 votes for drowsiness.

Hidden Markov Models have fewer parameters than RF models and they are more intuitive than trying to tune a decision tree. The state transition probabilities set the probability of an HMM changing state from awake to drowsy or vice-versa at any time step. The probabilities in each direction can be set independently. The emission probabilities set the chances that any of the observed variables of the HMM, or combinations thereof, indicate the value of the state. The state transition model is the base of the HMM with the prior probabilities, while the emission model conditions the state transitions with the presence of evidence. We estimated values for the emissions probabilities by counting the presence of RF model classification and their likelihood of correlating with a drowsy driving ground truth state. The final parameter that could be varied is the threshold we apply to the posterior probability, the output of the HMM, to set a final classification for the detection algorithm.

We chose not to attempt to tune the emission probabilities, for essentially the same reason we did not tune each decision tree. It would change the characteristics of the model and defeat the purpose of the machine learning training routines that optimize model parameters. We experimented with varying the state transition probabilities, the effect of which is similar to that of changing a low pass filter cutoff frequency that is filtering the HMM output (Figure 2). A lower transition probability will produce a more filtered signal that has a longer rise (or fall) time. Ultimately, the effect of varying these probabilities, while measurable, did not effect a significant change in the algorithm output.

Similarly, we explored the final HMM threshold value. This value was obtained previously as the optimal operating point on an ROC curve obtained after a model optimization process conducted on the RF and HMM models. This parameter is the easiest to understand, effectively dividing Figure 2 into an upper and a lower region that corresponds to drowsy and awake predictions, respectively. Unfortunately, the variation of this last threshold has the least effect out of all the parameter tuning that was tried. This is likely because most state transitions changed the posterior probability all the way from zero to one and vice versa. The number of cases where a transition changed direction partway was fewer than one might have expected. In that situation, we can only shift the edges of the state transition by a few seconds by varying the output threshold.

The parameters that had the greatest effect on the drowsiness detection algorithm were the vote thresholds of the two RF models. An RF model with a higher vote threshold simply requires more of its constituent decision trees to agree on the output classification. Setting this threshold above the majority value may be problematic because it may then be that neither class gathers enough votes to meet the threshold. Ten levels of parameters for the RF vote thresholds were set. Values for the Steer RF model are: {162, 170, 180, 190, 200, 210, 220, 230, 240, 250}. Values for the Brake RF are: {151, 160, 170, 180, 190, 200, 210, 220, 230, 240}. The parameters are always varied together, not independently. Level one values correspond to the optimal threshold obtained in prior work to optimize the ROC curve indicators of model performance. The subsequent levels step up the values of each threshold until the steering RF value reaches majority rule. Notice that the relationship between the two values is essentially maintained through the levels such that the steer RF threshold is always greater than the brake RF.

This particular range of parameters fits nicely with the goals of our model variation exercise. The optimal values gave the best performance when compared to the awake and drowsy ground truth data points; however, the majority of time history samples are not associated with any ground truth because there was no lane departure. Therefore, the algorithm performance at these points is difficult to judge. We did observe however, that many of these in-between points are classified as drowsy and thus contribute to the overall number of mitigation warnings. We would therefore consider this parameter set as being on the sensitive side. To make the models less sensitive, we wish to make it harder for the RF models to issue drowsy classifications, which means requiring more models to agree on drowsiness. Therefore we increase the values of the vote thresholds up to the majority rule value, but no further.

The different levels of RF models were run on the DrIIVE Phase I drowsiness data with all other parts of the detection algorithm held constant. Some simple metrics were calculated on the detection / mitigation system in order to compare across levels. A mitigation was considered to be in a 'correct' stage at each ground truth data point if it was in stage 0 (no mitigation) at an awake point or in any stage of mitigation at a drowsy point.

The system was designed to operate at speeds greater than 40 mph, so the percentage of time that the vehicle was traveling faster than this limit was calculated as a reference for other measures. A variable, *timeAtSpeed*, was calculated as the amount of time in the drive that the car was traveling above this limit. A variable, *timeInMitigation*, was calculated as the amount of time that the mitigation system was in any mitigation stage. Then a normalized measure was calculated as

$$Time\ in\ Mitigation\ (\%) = \frac{timeInMitigation}{timeAtSpeed} \times 100$$

Confining ourselves to only those samples with ground truth data, we counted which of those points were in the 'correct' stage of mitigation, as described above. This may be expressed in an indicator variable, I_C, of zeros and ones of length N, where N is the number of ground truth points in a drive. The percentage of correctly mitigated ground truth points was then computed in each drive as

$$Accuracy\ (\%) = \frac{100}{N} \sum_{1}^{N} I_C$$

This coarse metric does not indicate whether a ground truth data point falls in the first or last part of a mitigation, nor which stage of mitigation is active, nor whether an incorrectly mitigated ground truth point falls just before or after a period of mitigation. The accuracy metric, together with the time-in-mitigation metric, provide an idea of how parametric variations affect the output of the detection / mitigation system, and create a tradeoff between sensitivity and accuracy.

RESULTS

Ten levels of vote thresholds for the Steer RF and Pedals RF model were tested on the DrIIVE Phase I drowsiness data, which was all collected in unmitigated conditions. Both the three stage audio-visual mitigation as well as the binary haptic mitigation were run on each drive in the three conditions of that study: Day, Early Night, and Late Night. Since it was not possible to provide human interaction with the button response, an automatic button response was programmed after one second; therefore, the simulations do not account for unresponsive drivers.

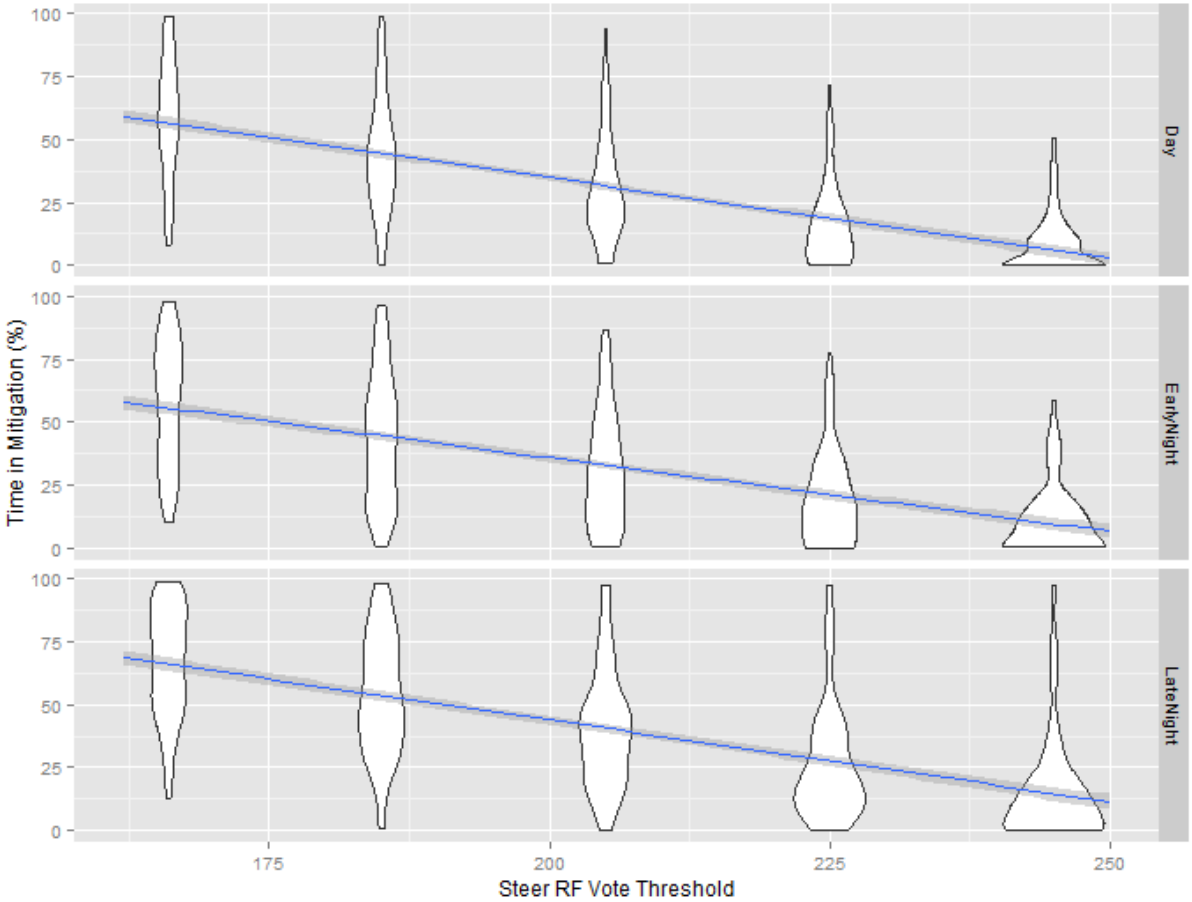

Figure 4. Violin plot and linear fit of the percentage of time in a drive that the mitigation is active for the audio/visual alert mode

The Time in Mitigation and Accuracy metrics are displayed for the three stage mitigation in Figure 4 and Figure 5, respectively. These figures show an overlay of a violin plot with a line fit, the latter with confidence intervals represented by a gray band. A violin plot shows information similar to a box plot, but shapes the sides of each 'box' according to the probability density of the sample points (Hintze & Nelson, 1998). The wider the shape is, the denser the points are at that location in the plot. The ggplot2 library (Wickhan, 2009) in R was used to generate the plots; and the plotting function was allowed to bin the horizontal axis from ten levels into just five bins, making the figure somewhat less dense and easier to comprehend.

The violin plot in Figure 4 shows the density of samples of the Time in Mitigation measure decreases as the vote threshold increases. This result holds across all three conditions and is completely intuitive. As more votes are required for the RF models to issue drowsy classifications, it becomes more difficult for the algorithm to transition into the drowsy state; and less time is spent in all stages of mitigation. The time in

mitigation at level ten for the Day, Early Night, and Late Night conditions is approximately 5%, 10%, and 12% respectively.

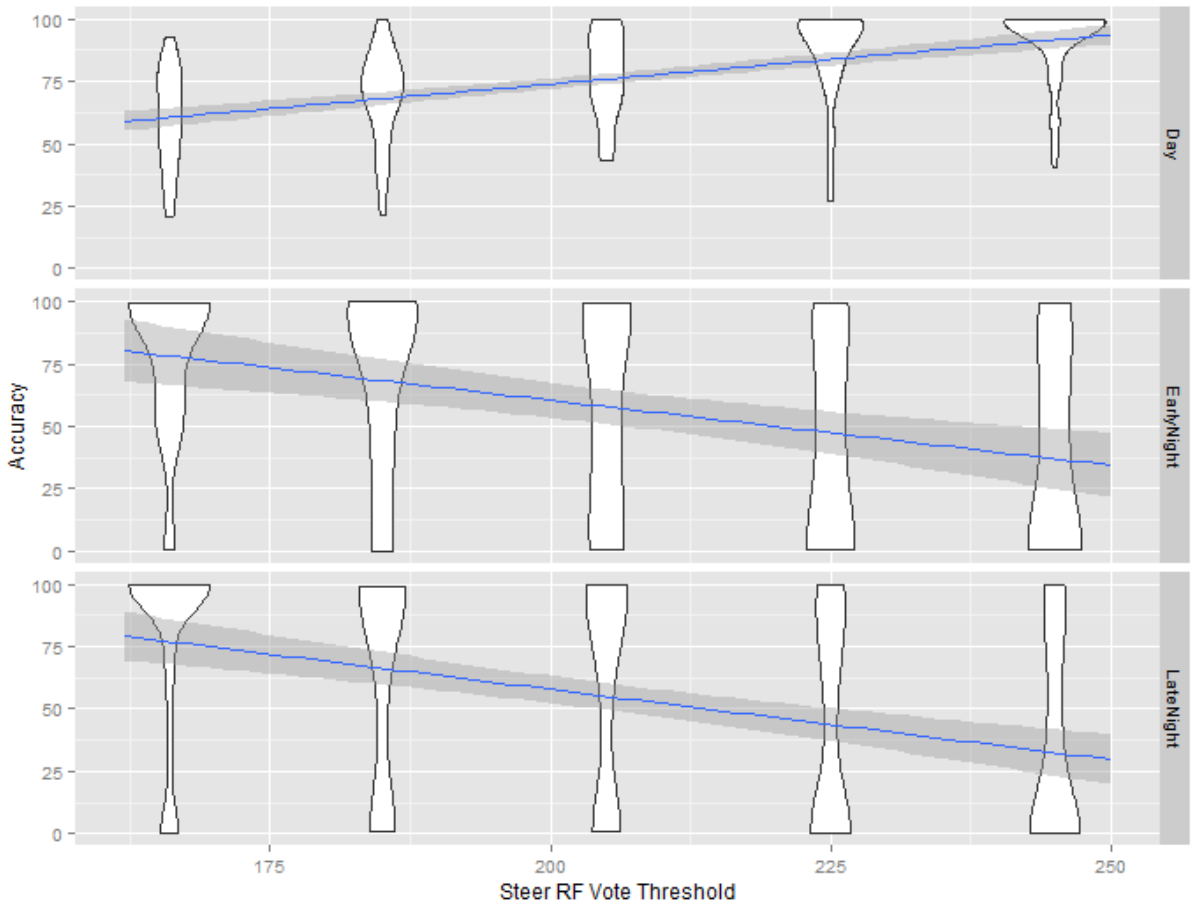

Figure 5. Violin plot and linear fit of the percentage of ground truth points that are correctly mitigated for the audio/visual alert mode

The violin plot in Figure 5 shows the density of samples of varying Accuracy as a function of the Steer RF vote threshold. The line fit serves to make clear the shift in the density of samples as the vote thresholds are increased. The accuracy in the Day condition actually rises as the RF models become more conservative. The explanation for this result is that almost all of the ground truth points in the Day conditions are awake points. Then it becomes clear that a simplistic approach of turning off the mitigation altogether would increase the accuracy in this condition to almost 100%. On the other hand, the accuracy is seen to drop for both night conditions as the vote threshold is increased, as expected. At the far end of the test, where vote thresholds for steering and pedals are 250 and 240 respectively, the estimated accuracy in the Day, Early Night, and Late Night conditions is 90%, 38%, and 30% respectively.

A similar pattern of results was obtained for the binary haptic mitigation, though the haptic system had smaller overall values for the Time in Mitigation metric. The different logic of the binary mitigation as compared to the more complex three-stage system result in less time spent in mitigation.

THREE SENSITIVITY LEVELS

The drowsiness mitigation study will have Early Night and Late Night conditions, but will not include Day drives. However three levels of sensitivity will be designed for each of the two mitigation types. Previous work resulted in trained models and an ROC curve evaluation of the models to optimize a drowsiness detection

algorithm. This 'optimal' model corresponds to the level one parameter set described in this paper. As discussed earlier, that optimization only considered ground truth points and classifications on in-between points were not part of the evaluation. In reality, the in-between samples make up a majority of the data in most drives and therefore contribute significantly to the number of mitigation alerts. All three conditions were in some stage of mitigation over half the time, which is especially surprising for the Day condition.

A commercial system would most likely condition the algorithm output on other factors such as time of day, driving style, traffic density, and perhaps other variables. Since we are not collecting additional Day drives, conditioning by day/night is not necessary at this time. Three mitigation models were selected for the two modalities with the purpose of obtaining a wide spread in the *timeInMitigation* and *accuracy* metrics. The starting point was to choose three target accuracy values. Those values then mapped to corresponding time in mitigation and RF model vote thresholds. The values selected for the three models are summarized in Table 1. These models are spaced far enough apart that they offer a clear distinction to the drivers who experience them.

Table 1. Three levels of mitigation selected using Late Night data for audio/visual and haptic modes

Level	Steer RF Votes		Pedal RF Votes		Time in Mitigation (%)		Accuracy (%)	
	Visual	Haptic	Visual	Haptic	Visual	Haptic	Visual	Haptic
Over sensitive	170	162	160	151	63	40	75	70
Nominal	190	175	180	165	50	38	63	63
Under-sensitive	215	195	205	185	38	30	50	50

CONCLUSIONS

Two mitigation systems, a three stage audio/visual and a binary haptic, were designed to use the output of a previously designed drowsiness detection algorithm. Additionally, three levels of each system were obtained to provide a good range of system sensitivity to drowsiness. In this way, a range of false alarm rates will be generated from the study and questions about the effectiveness of the mitigation might be differentiated from questions about the nuisance factor of the mitigation alerts.

The ultimate desired outcome for a drowsiness mitigation system is that the driver would realize their own impairment and pause the trip to rest. Such an outcome is not allowed for however in the protocol of the simulator experiment. On the other hand, a primary interest of the DrIIVE program is to study vehicle-based measures of impairment. Having determined that such measures are useful for classifying drowsiness, the data from the upcoming study may be used to test whether a mitigation system also causes detectable differences in driving performance as measured by vehicle-based sensors.

REFERENCES

Balkin, T. J., Horrey, W. J., Graeber, R. C., Czeisler, C. A., Dinges, D. F. (2011). The challenges and opportunities of technological approaches to fatigue management. *Accident Analysis & Prevention*, *43*(2), 565–72.

Brown, T., Lee, J., Schwarz, C., Fiorentino, D., McDonald, A. (2014). *Assessing the Feasibility of Vehicle-based Sensors to Detect Drowsy Driving* (Final Report No. DOT HS 811 886). Washington, DC: NHTSA.

Dinges, D. F., & Grace, R. (1998). *PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance*. *Federal Highway Administration*. Washington D.C.

Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J.-M., Gricourt, D., Davis, B. K., Carnahan, B. (1996). A drowsy driver detection system for heavy vehicles. In *17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings,* Vol. 2, pp. I36/1–I36/8.

Himmelmann, L. (2010). HMM: HMM - Hidden Markov Models (Version 1.0). Retrieved from http://cran.r-project.org/web/packages/HMM/index.html

Hintze, J. L., & Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, *52*(2), 181–184.

Ji, Q., Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *36*(5), 862–875.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue. *IEEE Transactions on Vehicular Technology*, *53*(4), 1052–1068.

Klauer, S., Dingus, T., Neale, V., Sudweeks, J., Ramsey, D. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. *Analysis*. (Report No. DOT HS 810 594). Washington D.C.: National Highway Traffic Safety Administration.

Krajewski, J., & Sommer, D. (2009). Steering wheel behavior based estimation of fatigue. In *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment and Design* (pp. 118–124).

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5).

Lal, S. K. L., Craig, A., Boord, P., Kirkup, L., Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research*, *34*(3), 321–328.

Lin, C., Wu, R., Liang, S., Chao, W., Chen, Y., Jung, T. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *52*(12), 2726–2738.

MacLean, A. W., Davies, D. R. ., Thiele, K. (2003). The hazards and prevention of driving while sleepy. *Sleep Medicine Reviews*, *7*(6), 507–521.

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C., Fuller, S. Z. (2000). Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis & Prevention*, *32*(4), 493–504.

McDonald, A. D., Lee, J. D., Schwarz, C., Brown, T. L. (2013a). Steering in a Random Forest: Ensemble Learning for Detecting Drowsiness-Related Lane Departures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

McDonald, A. D., Lee, J. D., Schwarz, C., Brown, T. L. (2013b). Steering in a Random Forest Ensemble Learning for Detecting Drowsiness-Related Lane Departures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

NHTSA. (2011). *Drowsy Driving* (Traffic Safety Facts No. DOT HS 811 449). Washington, DC: NHTSA.

Powers, D. M. W. (2007). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation* (Technical Report No. SIE-07-001). Adelaide, Australia: Flinders University.

R Development Core Team. (2009). *{R: A language and environment for statistical computing}*. Vienna, Austria: R Foundation for Statistical Computing.

Sayed, R., & Eskandarian, A. (2001). Unobtrusive drowsiness detection by neural network learning of driver steering. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *215*(9), 969–975.

Schwarz, C., McDonald, A., Lee, J. D., & Brown, T. L. (submitted). Time Series Classification of Drowsiness from Driving Input Signals. *IEEE Journal on Intelligent Transportation Systems*.

Wali, M., Murugappan, M., & Ahmmad, B. (2013). Wavelet Packet Transform Based Driver Distraction Level Classification Using EEG. *Mathematical Problems in Engineering*.

Wickhan, H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.

Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences*, *180*(10), 1942–1954.

Yang, J. H., Tijerina, L., Pilutti, T., Coughlin, J. F., Feron, E. (2009). Detection of Driver Fatigue Caused by Sleep Deprivation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *39*(4), 694–705.