Is Now A Good Time? An Empirical Study of Vehicle-Driver Communication Timing

Rob Semmens
Naval Postgraduate School
Monterey, CA
robert.semmens@nps.edu

Nikolas Martelaro
Accenture Technology Labs
San Francisco, CA
nikolas.martelaro@accenture.com

Pushyami Kaveti
Northeastern University
Boston, MA
kaveti.p@husky.neu.edu

Simon Stent
Toyota Research Institute
Cambridge, MA
simon.stent@tri.global

Wendy Ju
Cornell Tech
New York, NY
wendyju@cornell.edu

Abstract

Advances in automotive sensing systems and speech interfaces provide new opportunities for smarter driving assistants or infotainment systems. For both safety and consumer satisfaction reasons, any new system which interacts with drivers must do so at appropriate times. We asked 63 drivers, ”Is now a good time?” to receive non-driving information during a 50-minute drive. We analyzed 2,734 responses and synchronized automotive and video data, and show that while the chances of choosing a good time can be determined with better success using easily accessible automotive data, certain nuances in the problem require a richer understanding of the driver and environment states in order to achieve higher performance. We illustrate several of these nuances with quantitative and qualitative analyses to contribute to the understanding of how to design a system that might simultaneously minimize the risk of interacting at a bad time while maximizing the window of allowable interruption.

Keywords

Driving, timing, verbal interaction, datasets, automotive, navigation

ACM Reference Format:
Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time? An Empirical Study of Vehicle-Driver Communication Timing. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300867

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300867

Figure 1: When is a good time for a car to speak to drivers? We created a labeled dataset to find the best predictors.

1 Introduction

Recent advancements are laying the groundwork for intelligent vehicles that can help drivers with a variety of vehicle-based activities. Many of these services require user input, response or awareness. Speech interaction is often the preferred mode for these services because it does not require visual attention or physical manipulation. Speech interfaces can allow the driver to receive non-critical information or information not directly related to the driving task such as music, messages, or news, and also reciprocally solicit the driver for input. Examples of in-vehicle information that others are developing include: warnings about the car maintenance, information about future road closings, ads for local businesses, reminders of tomorrow’s meetings. There is research demonstrating that many individuals perform non-driving tasks while driving, and will seek non-driving information on hand-held devices even if it is against the law [8]. To be clear: our motivation for this work is to make such services safer, not to argue for such services.

Currently, most in-car speech interfaces require the driver to initialize the conversation via a physical button or a wake word for non-driving related tasks. This "pull" paradigm prevents the driver from being alarmed or surprised by unexpected interactions, but greatly restricts interactive service offerings. To enable more proactive interactive offerings for in-vehicle services, it is desirable to model when are good times for vehicle-driver communication.

In this paper, we address the question of when are good and bad moments for speech-based interfaces to engage with a driver in non-critical in-vehicle tasks. The goal of our study is to predict appropriate moments for interaction *in-situ* with an understanding of the driver-related and contextual factors that may lead to better or worse moments for interaction. To accomplish this, we conducted a naturalistic driving study with 63 drivers who each answered the question “Is now a good time?” throughout a 25.8 km (16 mi) course through freeway, suburban, and urban driving settings.

Multi-channel video and vehicle telemetry from our instrumented car allow us to examine many factors which influence communication timing. The results of our analysis suggest that people prefer to be spoken to when they are driving straight at constant speed or when they are stopped. They particularly do not want to be spoken to when they are in the middle of a driving maneuver or when they are off-course. These situations are easy to determine from navigation or CAN bus data. On the other hand, our analysis also includes factors that might be difficult to predict without performing scene recognition of the car’s surroundings.

2 Related Work

Prior research in the domains of speech user interfaces, interruption modeling and automotive human-machine interaction have laid the groundwork for this project. Here, we review the relevant literature that informs the design and analysis of our project.

Speech Interfaces in Cars

Currently, speech interfaces in cars are used for navigation, music selection, and cell phone interaction [32], although there is emerging interest in “artificial intelligence” assistants such as Toyota Motor Corporation’s Yui, that “gauges your mood, indulges in personal chitchat and offers to drive if it senses you are sleepy or distracted” [34]. Nissan-Renault are working with Microsoft to make an agent to help with predictive tasks such as “recommending routes to appointments stored in a calendar or suggesting a time for the vehicle to be taken in for maintenance” [23]. These interfaces take advantage of the long periods of time people spend in their cars in daily commutes, and can help people take care of daily chores or improve their moods.

On the other hand, the use of such systems could be unsafe if communications are designed poorly. A 2006 literature survey on speech interfaces in cars by Baron and Green found that in general, across 15 different studies, people drove at least as well, if not better (less lane variation, steadier acceleration/deceleration), when interacting with speech interfaces compared to manual interfaces. However, when compared to driving with no interface interactions, their driving was worse [5]. Some of the interaction advantages are due to the auditory vs physical or visual modes. After all, auditory interfaces are less detrimental to the driving task than visual interfaces [43].

However, speech introduces additional cognitive and psychological complexity into the mix. Horrey and Wickens found that conversing on a mobile phone degrades driving performance, even with a hands-free device, and naturalistic conversations degrade performance more than synthetic tasks [24]. Kun, Paek and Medenica found that lower speech recognition accuracy in in-simulator speech interfaces lowered participants’ driving performance [30]. Nass et al. found that mismatches in the emotions felt by drivers and expressed by in-vehicle voices affected driving performance [37].

Designers of automotive HMI have addressed the need to make the UI models and guidelines to make interaction flexible, to accommodate the cognitive limitations of drivers, to allow natural variation in speech input, and to reinforce speech with visual feedback and memory aids [10, 14, 26]. Nevertheless, last generation UI paradigms are often built around "voice buttons", i.e. voice triggering of existing function calls. These do not resemble natural speech communication, nor do they address the interaction issues raised by current generation in-vehicle voice assistants, which have the ability to proactively offer services or start interactions. These older systems also do not address the car’s ability to sense and model the driver’s broader context, so that the car might time communications around driving-based tasks much as a passenger in the car would. Research suggests that drivers’ conversations with passengers in cars degrade performance less than conversations with people not in the car because passengers modulate conversation during more demanding urban driving in naturalistic interaction and driving situations [12, 17].

Modeling Interruptibility

Although speaking to people while they are driving a car is not actually interrupting—because people are frequently able to conduct both tasks simultaneously—the HCI research in the area of interruptibility is pertinent. McFarlane proposed a taxonomy of how interruptions could occur—immediately, negotiated (signaling the prospective interruption), mediated (requesting interruption based on user’s load), scheduled (breaking in only at planned intervals)—and found that people performed best on the primary (continuous) task and on the interruption task if the interruptions were negotiated or mediated [33].

Working from ideas of predicting interruptibility first posed by Miyata and Norman in 1986 [35], Czerwinski et al. found that timing of instant messages could lessen the performance impact of the interruptions on primary task activity [13]. Fogarty et al. found that the people's interruptibility could be predicted using Naive Bayes classifiers with sensors placed *in-situ* in the office environment [4, 19, 20, 27]. Chen and Vertegaal similarly used sensors, but mounted on the person rather than in the environment, so as to model interruptibility based on user cognitive or physiological state instead of task phase [11].

The advent of smartphones equipped with a wide variety of sensors has made it possible for interruptibility to be modeled on-line and in-situ. Fisher and Simmons used reinforcement learning on smartphones to create a personalized model of interruptibility for incoming phone calls [18]. Similarly, Anguita et al. used the inertial sensors on smartphones to perform human activity recognition using a Support Vector Machine [3].

Driver modeling and workload management

In the automotive sector, cellphones have primarily been the distraction and interruption that keep people from driving well. Numerous attempts have been made to design systems to identify driving situations where in-vehicle information, mostly audio-based automotive warning alerts, must be managed [2, 6, 36, 42, 45]. Most of these were designed in a top-down fashion, architected from models of the vehicle-driver system and then evaluated with people after the fact. In practice, systems mostly limited access to cellphones or in-vehicle entertainment functions when the car was in motion, and formed a priority queue so that emergency alert warnings in the car would not overlap or overwhelm the driver [9].

One of the limiting factors in previous workload management systems was the lack of ability to perform sophisticated recognition of the driving context and the driver state. More recently this type of sensing and recognition has become increasingly viable. Research to detect driver actions [7, 39], intent [16, 31], distraction [44], drowsiness [25], inebriation [38], and stress [41] using in-vehicle sensors could help with workload management.

Currently, the intelligent vehicle community is looking beyond model- and rules- based approaches towards more data-driven approaches with large datasets of real-world driving. For instance, Virginia Tech's 100-car naturalistic driving study was the first instrumented-vehicle study undertaken with the primary purpose of collecting large-scale, naturalistic driving data [15]. More recently, MIT's Autonomous Vehicle Technology study is gathering data from 25 instrumented vehicles collected over months and years, with the goal of modeling driver behavior and interaction with current-day automated driving features [21]. In the computer vision and machine learning community, there is interest in using these datasets and developing new ones in order to enable the vehicle to predict the drivers' maneuvers. Brain4Cars, for example, uses recurrent neural networks with long short-term memory units to try to predict driver actions 3.5 seconds before occurrence. [28]

While these naturalistic driving projects may prove useful for characterizing current day driving activity at scale, they do not translate directly into actionable designs. In the specific area of vehicle-driver communication timing, for instance, this type of research does not establish "ground truth" to determine when driver vehicle communication should occur. In the "Sensors know when to interrupt you in the car" project, Kim et. al. addressed this by specifically modeling interruptibility using various on-body sensors and vehicle telemetry [29]. This work identifies episodes of peripheral interaction, when drivers are adjusting the radio or eating, as moments where drivers were interruptible. However, it is difficult to know if the drivers themselves would agree with that assumption.

Within this landscape, our project, which identifies opportune moments for vehicle-driver communication by asking drivers themselves whether it is a good time to engage, is unique. Our human-centered approach employs a negotiation model for communication timing. It focuses on driver preference and self-assessment of availability as opposed to task category, inferred driver capability or inferred safety risk. The question of when and why drivers feel they are interruptible may yield different answers than previous work modeling workload or activity, but it is a critical question nonetheless. Our human-centered work focuses on the empirical experience of individual users, and examines what aspects of interruptibility can and cannot be easily recognized using existing modeling techniques.

3 Data Collection

A total of 63 drivers completed a 25.8 km (16 mi) loop through freeway, suburban, and urban driving settings. Each driver answered the randomly posed question "Is now a good time?" an average of 43 times (SD = 6.6). Contextual information was collected using an instrumented 2016 Toyota Prius V. The data capture system included multi-channel video of the cabin and road, vehicle telemetry, position, inertial forces, and driver physiological data. Data collection occurred between August 2017 – May 2018.

All drivers gave consent to be captured using audio and video recorders during the study and further gave consent for the recordings to be used for scientific purposes and to be part of a dataset available to the research community.

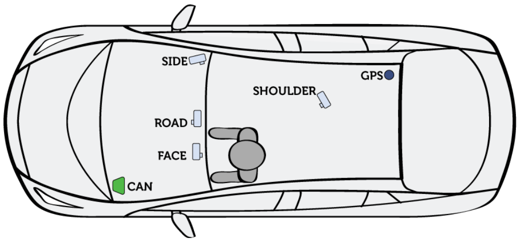

Figure 2: Instrumented 2016 Toyota Prius V.

Safety

On-road driving research can be distracting and dangerous, but field studies are critical for recreating ecologically valid findings. Indeed, the potential risk of vehicle-driver communications is one of the motivating factors for our research. To minimize risk to participants and other road users, we recruited drivers with valid driver's licenses as participants, vetted the course we intended the participants to take thoroughly, and instructed drivers that they could choose not to respond if they did not feel they were able to do so safely. We believe the level of risk and driving interference to be on the same order as verbal navigation instructions. This research was conducted under our institution's IRB protocol #41842 and with the purview of our institution's Office of Risk Management.

Participants

62 drivers (F = 23, M = 37, O = 1) aged between 21–96 years (M=40, SD=17) participated in our study. (Note that the demographic data for one driver is missing, so our analyses of the survey data show 61 participants. That driver's real-time responses are still part of the overall analyses.) Drivers reported having between 1–80 years (M=21, SD=18) of driving experience.

Data streams

The data used in this study is part of a larger project collecting naturalistic driving data to help understand when are good and bad moments for vehicle-driver communication.[1]

From the dataset, we extracted video, location, and automotive data collected using an instrumented 2016 Toyota Prius V. The following low-cost, consumer-grade sensors were used around the vehicle with their positions shown in Figure 2:

(1) 4× GoPro Hero 4 video cameras at 1920×1080px, 30 Hz with audio. These were mounted facing (i) the road ahead, (ii) the driver's face, (iii) the driver's body from the side, and (iv) over the driver's shoulder looking at the steering wheel.
(2) GlobalSat BU-353-S4 USB GPS - GPRMC sentences, data rate: 1 Hz
(3) 8devices USB2CAN reader - CAN (Controller Area Network) Bus data was collected including throttle position, brake sensor, speed, & steering angle, data rate: 100 Hz

Video from each camera was streamed via the live HDMI output to a video multi-viewer (Eazy2HD HDMI 4×1 Quad Multi-Viewer) where it was logged as a montaged 1280×720 px video. The montaged video was captured using H. 264 encoding using Apple QuickTime video capture software at an average effective framerate of 30 Hz.

The various sensor streams were logged using a laptop with a 2.5 GHz Intel i7 quad-core processor, 16 GB RAM, and a 1 TB solid-state disk. All incoming data was timestamped with UNIX epoch time using the Python time subsystem.

We captured interruption experience samples from drivers via speech. During the drive, an audio script prompted the driver with the question "Is now a good time?" every 30–120 seconds. The query time was logged with a UNIX epoch timestamp. The driver's response was recorded via the audio and video capture, see Figure 1.

Driving Route

The 25.8 km driving route included 5.2 km of highway, 19 km of suburban surface street, and 1.6 km of urban driving, shown in Figure 3. The route included 43 traffic lights, 15 stop signs, 6 unprotected left turns (no left turn light protecting from oncoming traffic), 5 protected left turns, and 12 right turns (all right turns occur at stop signs or traffic lights). The route included varying number of lanes and varying traffic conditions. The driving sessions occurred between 10:00 and 15:00 local time so that drivers avoided excessive rush-hour traffic. This ensured that each drive was approximately 50 minutes long.

Drivers were provided a mobile-phone based GPS navigation of the route using the inRoute GPS Navigation application. The application provided directions using speech and visual display. Due to issues with GPS signal and data coverage, the GPS navigation did not always function correctly and a paper map of the route including text directions was provided as a backup to drivers. Drivers mostly stayed on-route, however, some drivers deviated from the course, as shown in Figure 3.

Procedure

Before the Driving Session We recruited drivers from a university campus and the surrounding community. Upon arrival at our garage, participants read and agreed to an informed consent form and then completed a questionnaire asking about

[1]The data and the annotations associated with this study will be shared as part of a concurrent dataset submission, to encourage adoption of a data-driven approach to developing driver vehicle interactions.

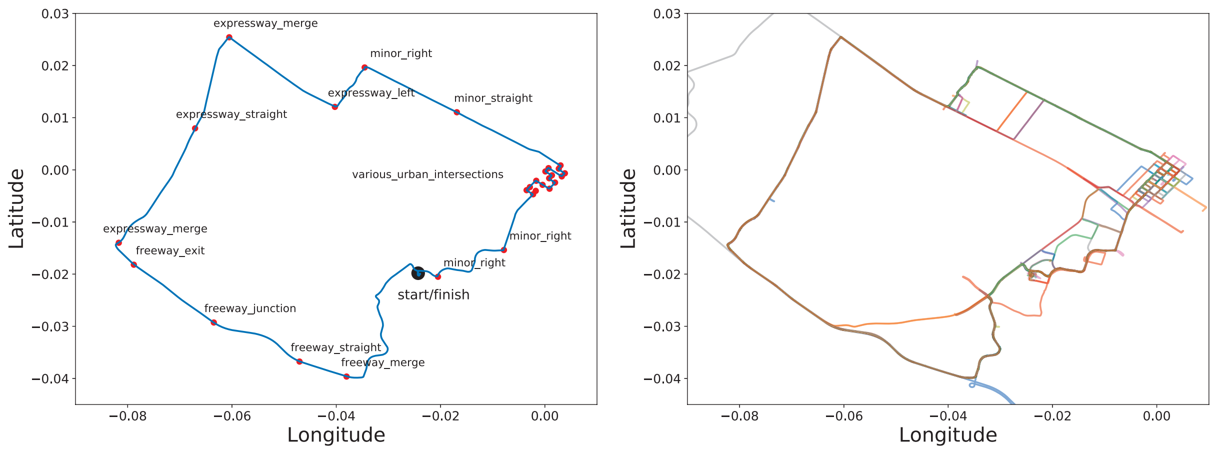

Figure 3: Left: the instructed driving route given to all subjects. This was a 25.75 km (16 mi) counterclockwise loop covering freeway, suburban and urban roads and a variety of protected and unprotected intersections. Right: actual GPS traces from the dataset, showing numerous deviations from the intended route, particularly in the urban area to the far right.

their demographic information and daily driving habits. We also checked that each participant had a valid driver's license.

An experimenter then explained the purpose of the study to the participant, telling them that they would be completing a loop around the surrounding area while answering the question "Is now a good time?" throughout the drive. Participants were told to interpret the question as if an intelligent car would provide them with non-critical information that they might like to know. We did not specify what content this information would contain and told participants to consider if it was a good time to receive information at all assuming it was not critical to the operation of the car or a safety warning. The drivers were instructed to answer the "Is now a good time?" query with a clear `''YES''` or `''NO''` unless they felt uncomfortable and preferred to stay silent. Drivers were told that they could also provide a rationale for their initial answer when they felt it was safe to do so. Aside from the interruption query task, drivers were told they could behave as they normally would while driving, including activities such as listening to the radio.

Before entering the car, the experimenter went over the route with the driver using a paper map. The experimenter also showed the driver the InRoute phone based navigation system and how to use it.

The experimenter helped the driver sit in the vehicle and adjust the seat and steering wheel. Once seated in the vehicle, the driver was instructed to relax and look forward while a five-minute physiological baseline was taken. During these five-minutes, the experimenter started the other data recording systems. Once the baseline was completed, the interruption query system was started and the experimenter let the driver hear a sample "Is now a good time?" query. The experimenter ensured that all drivers could hear the prompt. The experimenter told drivers that they could behave as they normally would while driving, including listening to the radio and using the climate controls. The experimenter reminded the participant to drive safely first and foremost. After any final questions, the driver left the garage and began the route.

During the Driving Session Drivers completed the drive in approximately 50 minutes. During the drive, participants followed the navigation directions for the route and answered the "Is now a good time?" query. Driver mostly stayed on the route, however, there were a number of instances where drivers went off-route. Due to signal issues, the GPS navigation did not always automatically reroute drivers. This occurred most frequently in the downtown and campus areas of the route.

After the Driving Session Upon returning the driver completed a short questionnaire about their experience. The experimenter conducted a five-minute interview asking the driver about what times were good, bad and generally how the driver felt during the drive.

4 Data Processing

Labeling

Drivers self-annotated their own communication preferences by answering the question "Is now a good time?" as they were driving. Drivers were instructed to provide a clear `YES/NO` after each query, or to not respond if they did not feel they were able to do so safely–these were coded as `NO-ANSWER` after the fact. The system did not perform any task or offer any services following the question. Drivers were also told

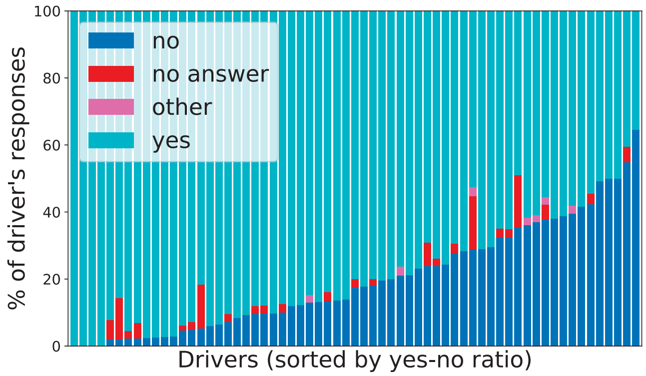

Figure 4: Distribution of responses by driver. The response ratio of NOs to YESes varied considerably between drivers, from 0% at one extreme to over 50% in the other.

that they could provide a rationale for their answer: for example, *"No. I am merging onto the freeway."*

We labeled the driver's responses by hand to ensure accuracy and handle when drivers answered with deviations from YES or NO such as *"sure"* or *"not really."* To do this, we created 15-second clips of each query. These clips included six seconds before the query and nine seconds after the query. We also transcribed any rationale that drivers provided. Further, we included a short description of what the driver was doing, such as "Driving straight. Low traffic ahead." or "Stopped at a traffic light. First car in queue." Because we were recording what the drivers said objectively, we did not require interrater reliability.

Descriptive Statistics

We collected 2,734 responses during the experiment, extracted from 59 hours of driving video. Table 1 shows the frequency and distribution of responses. In 77.9% of instances, drivers responded yes. We found no significant effects of gender (males said yes 78% of the time, females 77%), age, or driving experience. However, there was a large amount of individual variation, ranging from some participants answering yes all of the time to several who answered yes less than 50% of the time. See Figure 4 for the distribution of responses by individual.

5 Why is it a good time (or not)?

In this section, we focus on what drivers offer as first-person explanations of why they are saying YES or NO, as well as third person descriptions of in-vehicle activity performed by researchers who watched multi-perspective video clips of these moments.

Table 1: Responses to "Is Now a Good Time?".

Response	Number	Percentage
YES	2,130	77.9%
NO	545	19.9%
NO ANSWER	52	1.9%
OTHER	7	0.3%
Total	2,734	100.0%

Table 2: Relative frequency of words most used by drivers after responding YES.

Word	Good time freq.	Bad time freq.	Diff.
light	261	7	+254
stop/stopped	220	22	+198
traffic	211	33	+178
red	107	1	+106
just	118	16	+102
waiting	83	4	+79
driving	74	4	+70
straight	61	2	+59

What Drivers Said

As part of responding YES or NO when queried, participants were invited to provide a reason for why it was or was not a good time. Participants provided a reason 1,158 times. Tables 2 and 3 show the relative frequency of words used in good and bad reasons, respectively. Many of these words were used together in reasons. For example, a common good-time reason was, *"I'm just waiting at a stop light."* Another was, *"I'm driving straight, and traffic is light."* Very different words were used most frequently to describe bad times, as shown in Table 3. Drivers often responded with their intention, such as *"I'm trying to figure out the GPS,"* or *"I need to make this turn."* The verb turn is unusual, as it often appears when it is a good time as well. Several participants responded with *"I'm just waiting to turn,"* and that was a good time, whereas others said, *"I can't find my turn,"* and that was a bad time.

What Drivers Were Doing

Next we analyzed textual annotations of what drivers were doing when they answered the question. The relative frequencies of the words were not insightful in this instance–'driving' was the most frequent word for both good times and bad times. Table 4 shows works corresponding the highest probability following NO Several of the words relate to navigation, including "lost", "GPS", and "directions."

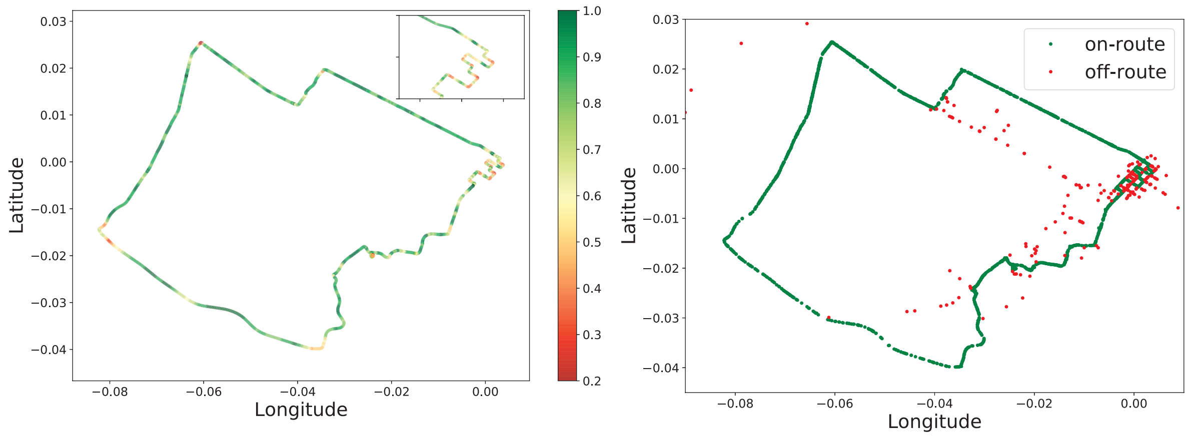

Figure 5: *Left:* estimated probability of a participant responding YES based on their location along the route. Several notable bad times are evident: these typically occur at medium– to high–speed intersections (such as merging on or off a major road), and around complex intersections such as in the urban area in the upper right. Intersections with stop-lights, such as the upper center are less likely to be considered bad times. *Right:* scatter plot of responses occurring when subjects were driving on or off the designated route. It was found that while driving off the route (i.e. lost or re-routing), a subject's chances of saying NO rose from 19% to 33%.

Table 3: Relative frequency of words most used by drivers after responding NO.

Word	Bad time freq.	Good time freq.	Diff.
trying	67	3	+64
turn/turning	64	31	+33
figure	30	0	+30
make	23	4	+19
GPS	19	4	+15
merging	15	3	+12
intersection	18	7	+11
left	18	7	+11
need	12	2	+10

Within the NO clips we reviewed, a few unusual situations stood out. For example, many NOs had to do with unique environmental factors; *"Traffic is getting more intense and a police car pulled someone over on the other side of the street," "There was a big truck," "Someone almost just backed into me,"* and *"Changing lanes in front of the Ferrari."* These examples suggest that there can be specific things in the environment that may make drivers nervous or more vigilant and thus reduce their availability for communication. None of these things are common, but in fact the uncommonness of the situation is what made it a NO moment.

The presence of vulnerable road users also makes for bad times to communicate. Even if drivers were stopped or slowing, we noticed that they would remain more vigilant around pedestrians or bikers. Some drivers noted, *"Trying to make a right on a red and not hit these pedestrians"* or *"There are bikes and people turning. So I'm going to be careful of the biker."*

Table 4: Probability of participant answering NO conditioned on words in the activity annotations.

Word	Good time freq.	Bad time freq.	Probability
lost	1	13	0.93
listening	3	16	0.84
GPS	6	24	0.80
directions	3	8	0.72
trying	3	7	0.70
unprotected	12	9	0.69
backing	5	7	0.58
lanes	8	11	0.58
make	27	35	0.56
exit/exiting	14	17	0.54
changing	8	8	0.50

Finally, we noticed some of the serial NO and NO-ANSWER moments came when people were *"groovin to the music,"* or were busy and wanting to concentrate.

From an HCI perspective, while these moments may be less frequent, they are still important to understand. They represent moments that cannot be easily divined using existing in-vehicle technology, but might be determined through connected vehicle networking technology, driver activity monitoring, or environmental monitoring.

Table 5: Frequency of YES/NO when driver was following or had strayed from the designated GPS route.

		Good time?	
		YES	NO
On the route?	YES	1,918 (81%)	440 (19%)
	NO	212 (67%)	105 (33%)

6 Predicting good times

The dataset collected offers numerous ways to predict good moments for vehicle-driver communication. Here, we focus on some hypotheses generated by the analyses of what the drivers mentioned and what we noticed while labeling: if people were currently trying to perform a particular driving maneuver, whether drivers were lost, and whether the driver is anticipating an upcoming event.

These are initial investigatory analyses, intended to explore specific aspects of the problem, and are not intended to represent optimal algorithms for predicting vehicle-driver communication timing.

Location

We asked participants to follow a prescribed route and provided them with turn-by-turn directions via a GPS. Nevertheless, for a variety of reasons (e.g. construction, ambiguous GPS instruction, unable to follow direction safely) participants frequently exited the prescribed course. In the third party coding, it appeared evident that these moments, when people were off course, were not good times for participants.

While on the correct driving route, there were 2,358 responses (88% of total YES/NO responses) of which 440 were negative. While driving off of the correct driving route (during which the GPS would have been recalculating), there were 317 responses (12% of total YES/NO responses) of which 105 were negative. Participants who were off route were 2.12 times more likely to respond negatively than those who were not. A chi-square test of independence was performed to examine the relation between being off route and whether it was a good time. The relation is significant, $\chi 2(1, N = 2675) = 35.16, p < .0001)$.

CAN Bus-only Analysis

CAN (Controller Area Network) has been a standard on US cars and light trucks since 1994 and became mandatory in Europe in 2001, and in the US in 2008. The benefit of being able to predict vehicle-driver timing using CAN bus data alone is that CAN bus is readily available in existing automobiles and hence could improve driver vehicle communication timing immediately.

Data pre-processing CAN data messages pertaining to the motion of the vehicle were decoded and synchronized with the video and response times. These messages included vehicle speed, accelerator pedal angle, steering wheel angle and brake oil pressure. A dataset of CAN bus data clips was created by extracting ±5 seconds of data around each of the 2,734 participant responses.

Probability heatmaps To explore the relationships between the extracted CAN data and the responses of subjects, we created pair-wise heat maps representing the empirical probability of responding YES in different parts of the data space using a K-nearest neighbors approach. A selection of the more revealing heat maps are shown in Figure 6. In the figure, the transparency of the heat maps are modulated such that they become more transparent as the density of observations declines.

Figure 6 (left panel) shows that the probability of responding YES is higher when the steering wheel is straight, and is higher still when the car is traveling faster. Also, the darker region to the left of -200 degrees wheel angle, as compared to the lighter region to the right of 200 degrees, indicates that the probability of people saying NO is higher when turning left for any speed, which is expected since left turns are often across traffic and require more concentration on average.

Figure 6 (center) plots the change in vehicle speed to change in brake pressure. A positive value of change in vehicle speed means that the car is accelerating. For change in braking force, zero indicates no change, whereas a positive number indicates the driver is applying the brake (as opposed to releasing it). The yellow patch at the center corresponds to no change in speed and no change in brake pressure. This indicates that there is a high probability of saying YES when the car is moving at a constant speed or is completely stopped. The upper left quadrant represents the car accelerating from a stop and the lower right quadrant represents active braking. From the plot it can be observed that people are more likely to say YES while accelerating than when braking.

In Figure 6 (right), we examine the change in braking force compared to vehicle speed. This heat map shows that the probability of saying yes is higher when the participant is releasing the brake versus applying the brake, regardless of the speed. It also shows that the probability of saying yes is highest when there is no change in brake pressure and the car is traveling faster. These support the observations made from center plot.

Feature extraction To try to capture these relationships in a simple model for the binary classification task of detecting good or bad times from CAN data, we extracted the following simple-to-compute features from each message stream, x, for each response at time t seconds:

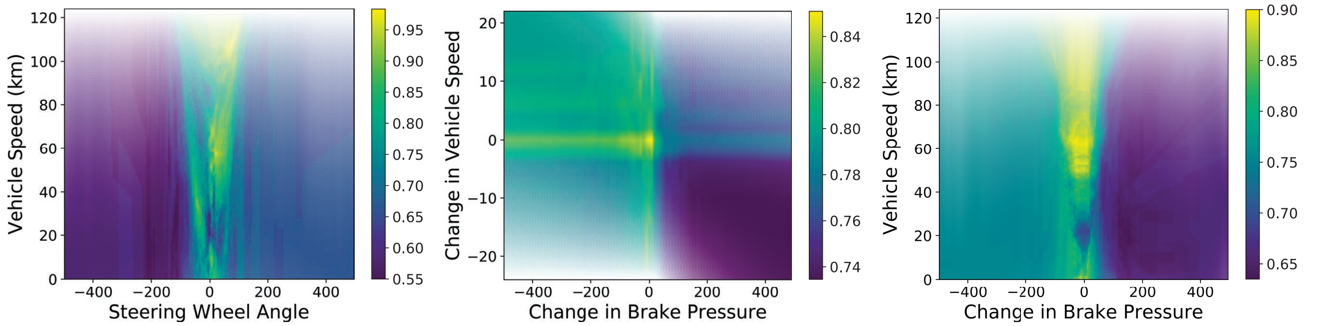

Figure 6: Pairwise heat maps of CAN data showing the empirical probability of drivers responding YES. Left: There is a high probability of saying YES while driving straight. Center: Subjects tended to say YES when the car was stopped or moving at a constant speed. Right: Subjects were more likely to say NO while they were actively braking versus releasing the brake.

(1) Mean value of x over the time windows $(t-5, t-2.5)$ and $(t-2.5, t)$
(2) Change in value of x from $t-5$ to $t-2.5$ and from $t-2.5$ to t
(3) Difference in features from (2)

where the message streams x included vehicle speed, brake pressure, steering wheel angle and gas pedal pressure. These features formed our "history" feature set.

In addition, we generated a second, augmented set of "history + lookahead" features, which used the same features calculated above but this time evaluated both over the historical range ($t-5$ to t) and the lookahead range of (t to $t+3$). The purpose of this was to understand the performance improvement of being able to perfectly predict the vehicles state (as expressed by these CAN signals) 3 seconds into the future. The time spans of 5 and 3 seconds were chosen empirically and all features were normalized to have zero mean and unit variance over the dataset.

Modeling We model the feature sets using support vector machines (SVMs) [22] as they are well-suited to binary classification problems, require relatively limited computation at test-time and are a good choice of model when the dataset is small and prone to over-fitting. The non-linearity of the decision boundary between good and bad times in the feature space can be addressed by choosing non-linear kernels such as polynomial functions, radial basis functions or sigmoid functions. As shown in Table 1, the classes in the dataset are imbalanced with the majority (77.9%) of samples belonging to YES. We used a stratified sampling approach to account for this imbalance. This maintains a constant ratio of samples belonging to each class in the training set and validation sets of all cross validation folds. To further account for the class imbalance issue, we weight the cost of misclassifying each class with a term which is inversely proportional to the class' frequency of occurrence. We tune the hyperparameters

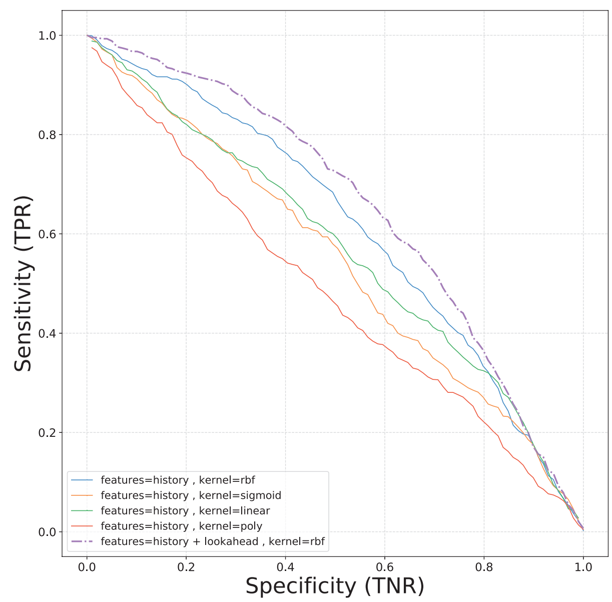

Figure 7: Sensitivity vs specificity for SVM models using various kernel functions and extracted CAN data features.

of the SVMs via grid search with 5-fold cross validation to obtain the best model for each kernel tested.

Analysis In Figure 7 we plot sensitivity vs. specificity curves for various SVM setups. These curves help in the assessment of performance of the model at different true positive rate and true negative rate thresholds (where we consider a true positive here to be a correctly identified bad time). The thick lines on the graph show that the radial basis function kernel was better at capturing the underlying shape of the feature space better than other, less flexible kernels. However, the overall performance achieved by the model using historical features leaves room for improvement: at 80% performance for correctly detecting bad times for interruption, the system would only preserve 35% of the overall driving window for correct interruption. This suggests that such features extracted from CAN data, though rich enough to draw some intuitive conclusions about driver responses (as shown in Figure 6) are insufficient to capture the many nuances behind whether a driver thinks it is a good time to interrupt.

The dashed line represents the "history + lookahead" feature set, where the model is given the benefit of knowing 3 seconds into the future about the state of the vehicle. While performance improves somewhat over the "history"-only feature set, the overall performance remains relatively low.

7 Discussion

Our goal was to explore specific aspects of the problem which emerged from what people said and what we noted people were doing. What did we learn about how to predict driver availability? First, simple CAN features alone (such as those we extracted in Section 6 do not provide sufficient predictive power. Second, being on or off a predesignated route is predictive. And third, monitoring driver state and intention would increase the accuracy of determining if it is a good time to interact.

CAN Data

Because of its widespread availability, it would be extremely useful if we were able to predict vehicle-driver communication timing using CAN bus data alone. The CAN data's predictions align with the reasons drivers gave for their responses. The heat maps in Figure 6 demonstrate that it is a better time to talk when the driver is driving straight and when there is little change in brake pressure. This describes both driving straight and waiting at lights, which were among the most frequent words used in reasons that it was a good time.

However, the features we extracted from the CAN data alone were not a strong predictor of good times. Figure 7 shows that if we would like the model to be 80% correct about knowing it's a bad time, it would only be able to interrupt during 35% of the good times. So, while the features we extracted from the CAN data might be useful in forming predictions for vehicle-driver timing, they are not very powerful on its own.

Location

Conversely, location relative to the route is a good predictor. Drivers were nearly twice as likely to find that it was not a good time for non-critical communication when they were off-route. This is also reflected in the frequency of words the drivers used in describing reasons for `NO` responses, as well as what the labelers found when annotating the videos.

Information about whether people are on- or off-route can easily be determined by in-car navigation systems. However, navigation information is not always readily available to vehicle systems because many drivers prefer the turn-by-turn directions provided by smart phone applications to factory-installed navigation [1]. Hence, the use of on-route vs off-route as a predictor of communication opportunity is currently more accessible for smartphone-based than vehicle-based applications. Nonetheless, this information is highly valuable towards predicting vehicle-driver communication timing.

Driver Monitoring and Intent Prediction

The reasons that the drivers said it was a bad time (Table 3) suggest that understanding the driver's future intent is a potential feature for determining when drivers are free to interact. Four reasons in particular, *trying*, *figure*, *make*, and *need* suggest that the driver is preparing to do something or is cognitively loaded and thinking about a future action. During these moments, drivers were often trying to find the next road to take or preparing to make a turn. These features may be determined using algorithms currently being developed to predict the driver's next maneuver [28] or estimate their cognitive load [40].

8 Conclusion

The in-car environment can be a rich context for human-machine interaction if the interactions can be well-timed. On the whole, drivers in our study indicated that they welcome non-critical speech-based interaction, and most moments in their drives were moments when people felt were a good time to talk. However, poorly timed interactions could cause stress and even danger to drivers. Hence, data, analysis, modeling and prediction of opportune moments for vehicle-driver communication are critically important to future interactions.

This work builds on the long history in CHI of figuring out how to adapt the machine to the needs, preferences and constraints of the human. The car is opportune space in which to use machine learning to improve interaction. We've identified that moments when people are off course and feeling lost are moments that they do not want to be interrupted by some non-critical interaction. Quantifying the correlation against other common-sense claims has real value in the life-or-death calculations of what should happen while people are driving. We believe this work breaks new ground in using machine learning to support human-machine interaction, and believe it points a path towards a future where machines are responsive to what people want to do and need as help to get there.

Acknowledgments

The Toyota Research Institute (TRI) provided funds to assist with this research, but this paper solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity. Rob Semmens, Nikolas Martelaro, and Wendy Ju began this this work while at Stanford University. Pushyami Kaveti began this work as an intern at TRI. This work was completed under Stanford University's IRB Protocol #30016. The authors thank Lawrence Domingo, Quillan Smith, and Angela Montemayor for their assistance during data collection.

References

[1] 2018. Automakers locked in battle with phone manufacturers for consumers' attention, j.d. power finds. (Aug 2018). https://www.jdpower.com/business/press-releases/jd-power-2018-us-tech-experience-index-txi-study

[2] Angelos Amditis, Luisa Andreone, Katia Pagle, Gustav Markkula, Enrica Deregibus, Maria Romera Rue, Francesco Bellotti, Andreas Engelsberg, Rino Brouwer, Björn Peters, et al. 2010. Towards the automotive HMI of the future: overview of the AIDE-integrated project results. *IEEE Transactions on Intelligent Transportation Systems* 11, 3 (2010), 567–578.

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. 2012. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *International workshop on ambient assisted living*. Springer, 216–223.

[4] Daniel Avrahami, James Fogarty, and Scott E Hudson. 2007. Biases in human estimation of interruptibility: effects and implications for practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 50–60.

[5] Adriana Barón and Paul Green. 2006. *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review.* Technical Report. University of Michigan, Transportation Research Institute.

[6] T Bellet, MP Bruyas, H Tattegrain-Veste, JF Forzy, A Simoes, J Carvalhais, P Lockwood, J Boudy, B Baligand, S Damiani, et al. 2002. "Real time" analysis of the driving situation in order to manage on-board information. In *Proceedings of the E-safety Congress and Exhibition–IT Solutions for Safety and Security in Intelligent Transport.*

[7] Asher Bender, Gabriel Agamennoni, James R Ward, Stewart Worrall, and Eduardo M Nebot. 2015. An unsupervised approach for inferring driver behavior from naturalistic driving data. *IEEE Transactions on Intelligent Transportation Systems* 16, 6 (2015), 3325–3336.

[8] Keli A. Braitman and Anne T. McCartt. 2010. National Reported Patterns of Driver Cell Phone Use in the United States. *Traffic Injury Prevention* 11, 6 (2010), 543–548. https://doi.org/10.1080/15389588.2010.504247 arXiv:https://doi.org/10.1080/15389588.2010.504247 PMID: 21128181.

[9] Robert Broström, Johan Engström, Anders Agnvall, and Gustav Markkula. 2006. Towards the next generation intelligent driver information system (IDIS): The Volvo car interaction manager concept. In *Proceedings of the 2006 ITS World Congress*. 32.

[10] Lawrence Cavedon, Fuliang Weng, Rohit Mishra, Harry Bratt, Badri Raghunathan, Hua Cheng, Hauke Schmidt, Danilo Mirkovic, Ben Bei, Heather Pon-Barry, et al. 2005. Developing a conversational in-car dialog system. In *Intelligent Transportation Systems*.

[11] Daniel Chen, Jamie Hart, and Roel Vertegaal. 2007. Towards a physiological model of user interruptability. In *IFIP Conference on Human-Computer Interaction*. Springer, 439–451.

[12] David Crundall, Manpreet Bains, Peter Chapman, and Geoffrey Underwood. 2005. Regulating conversation during driving: a problem for mobile telephones? *Transportation Research Part F: Traffic Psychology and Behaviour* 8, 3 (2005), 197–211.

[13] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. 2000. Instant messaging: Effects of relevance and timing. In *People and computers XIV: Proceedings of HCI*, Vol. 2. 71–76.

[14] Guido de Melo, Frank Honold, Michael Weber, Mark Poguntke, and André Berton. 2009. Towards a flexible UI model for automotive human-machine interaction. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 47–50.

[15] Thomas A Dingus, Sheila G Klauer, Vicki L Neale, Andy Petersen, Suzanne E Lee, JD Sudweeks, Miguel A Perez, Jonathan Hankey, DJ Ramsey, Santosh Gupta, et al. 2006. *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment.* Technical Report.

[16] Anup Doshi and Mohan Manubhai Trivedi. 2009. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems* 10, 3 (2009), 453–462.

[17] Frank A Drews, Monisha Pasupathi, and David L Strayer. 2008. Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology: Applied* 14, 4 (2008), 392.

[18] Robert Fisher and Reid Simmons. 2011. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, Vol. 1. IEEE, 436–441.

[19] James Fogarty, Scott E Hudson, Christopher G Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C Lee, and Jie Yang. 2005. Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 1 (2005), 119–146.

[20] James Fogarty, Andrew J Ko, Htet Htet Aung, Elspeth Golden, Karen P Tang, and Scott E Hudson. 2005. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 331–340.

[21] Lex Fridman, Daniel E Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Julia Kindelsberger, Li Ding, Sean Seaman, et al. 2017. MIT Autonomous Vehicle Technology Study: Large-Scale Deep Learning Based Analysis of Driver Behavior and Interaction with Automation. *arXiv preprint arXiv:1711.06976* (2017).

[22] Marti A. Hearst, Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. 1998. Support vector machines. *IEEE Intelligent Systems and their applications* 13, 4 (1998), 18–28.

[23] Tim Higgins. 2017. Auto Makers Showcase Cars Equipped With Virtual Assistants. *wsj.com* (Jan 2017). https://www.wsj.com/articles/auto-makers-showcase-cars-equipped-with-virtual-assistants-1483661702

[24] William J Horrey and Christopher D Wickens. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors* 48, 1 (2006), 196–205.

[25] Shuyan Hu and Gangtie Zheng. 2009. Driver drowsiness detection with eyelid related parameters by Support Vector Machine. *Expert Systems with Applications* 36, 4 (2009), 7651–7658.

[26] Zhang Hua and Wei Lieh Ng. 2010. Speech recognition interface design for in-vehicle system. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 29–33.

[27] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting human interruptibility with sensors: a Wizard of Oz feasibility study. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 257–264.

[28] Ashesh Jain, Hema S Koppula, Shane Soh, Bharad Raghavan, Avi Singh, and Ashutosh Saxena. 2016. Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture. *arXiv preprint arXiv:1601.00740* (2016).

[29] SeungJun Kim, Jaemin Chun, and Anind K. Dey. 2015. Sensors Know When to Interrupt You in the Car: Detecting Driver Interruptibility Through Monitoring of Peripheral Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 487–496. https://doi.org/10.1145/2702123.2702409

[30] Andrew Kun, Tim Paek, and Zeljko Medenica. 2007. The effect of speech interface accuracy on driving performance. In *Eighth Annual Conference of the International Speech Communication Association*.

[31] Firas Lethaus, Martin RK Baumann, Frank Köster, and Karsten Lemmer. 2013. A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing* 121 (2013), 108–130.

[32] Victor Ei-Wen Lo and Paul A Green. 2013. Development and evaluation of automotive speech interfaces: useful information from the human factors and the related literature. *International Journal of Vehicular Technology* 2013 (2013).

[33] Daniel C McFarlane. 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction* 17, 1 (2002), 63–139.

[34] Sean McLain. 2017. Toyota's Talking Car Wants to Be Your Clingy BFF. *wsj.com* (Oct 2017). https://www.wsj.com/articles/toyotas-talking-car-wants-to-be-your-clingy-bff-1509359421

[35] Yoshiro Miyata and Donald A Norman. 1986. Psychological issues in support of multiple activities. *User centered system design: New perspectives on human-computer interaction* (1986), 265–284.

[36] R Montanari, G Wenzel, S Mattes, F Kuhn, F BELLOTTI, and D Morreale. 2002. COMUNICAR: Integrated on-vehicle human machine interface designed to avoid driver information overload. In *9th World Congress on Intelligent Transport SystemsITS America, ITS Japan, ERTICO (Intelligent Transport Systems and Services-Europe)*.

[37] Clifford Nass, Ing-Marie Jonsson, Helen Harris, Ben Reaves, Jack Endo, Scott Brave, and Leila Takayama. 2005. Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI'05 extended abstracts on Human factors in computing systems*. ACM, 1973–1976.

[38] Victor-Emil Neagoe and Serban-Vasile Carata. 2017. Drunkenness diagnosis using a Neural Network-based approach for analysis of facial images in the thermal infrared spectrum. In *E-Health and Bioengineering Conference (EHB), 2017*. IEEE, 165–168.

[39] Nuria Oliver and Alex P Pentland. 2000. Graphical models for driver behavior recognition in a smartcar. In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*. IEEE, 7–12.

[40] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, 141–144.

[41] Pablo E Paredes, Francisco Ordonez, Wendy Ju, and James A Landay. 2018. Fast & Furious: Detecting Stress with a Car Steering Wheel. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 665.

[42] Walter Piechulla, Christoph Mayser, Helmar Gehrke, and Winfried König. 2003. Reducing drivers' mental workload by means of an adaptive man–machine interface. *Transportation Research Part F: Traffic Psychology and Behaviour* 6, 4 (2003), 233–248.

[43] Jaka Sodnik, Christina Dicke, Sašo Tomažič, and Mark Billinghurst. 2008. A user study of auditory versus visual interfaces for use while driving. *International journal of human-computer studies* 66, 5 (2008), 318–332.

[44] Fabio Tango and Marco Botta. 2013. Real-time detection system of driver distraction using machine learning. *IEEE Transactions on Intelligent Transportation Systems* 14, 2 (2013), 894–905.

[45] Yuji Uchiyama, Shin-ichi Kojima, Takero Hongo, Ryuta Terashima, and Toshihiro Wakita. 2002. Voice information system adapted to driver's mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 46. SAGE Publications Sage CA: Los Angeles, CA, 1871–1875.