Modeling of Driver Behavior in Real World Scenarios using Multiple Noninvasive Sensors

Nanxiang Li, Jinesh J. Jain and Carlos Busso, *Member, IEEE*

***Abstract*—With the development of new in-vehicle technology, drivers are exposed to more sources of distraction, which can lead to an unintentional accident. Monitoring the driver attention level has become a relevant research problem. This is the precise aim of this study. A database containing 20 drivers was collected in real-driving scenarios. The drivers were asked to perform common secondary tasks such as operating the radio, phone and a navigation system. The collected database comprises of various noninvasive sensors including the controller area network-bus (CAN-Bus), video cameras and microphone arrays. The study analyzes the effects in driver behaviors induced by secondary tasks. The corpus is analyzed to identify multimodal features that can be used to discriminate between normal and task driving conditions. Separate binary classifiers are trained to distinguish between normal and each of the secondary tasks, achieving an average accuracy of 77.2%. When a joint, multi-class classifier is trained, the system achieved accuracies of 40.8%, which is significantly higher than chances (12.5%). We observed that the classifiers' accuracy varies across secondary tasks, suggesting that certain tasks are more distracting than others. Motivated by these results, the study builds statistical models in the form of *Gaussian Mixture Models* (GMMs) to quantify the actual deviations in driver behaviors from the expected normal driving patterns. The study includes task independent and task dependent models. Building upon these results, a regression model is proposed to obtain a metric that characterizes the attention level of the driver. This metric can be used to signal alarms, preventing collision and improving the overall driving experience.**

***Index Terms*—Driver behavior, multimodal feature analysis, subjective evaluation of distraction, Gaussian mixture models.**

I. Introduction

DETECTING driver distraction has been an important research topic over the past few years. While there are many common reasons for vehicle crashes, driver distraction and inattention are very prominent causes. Previous studies have shown the impact that inattentions can have on driving behavior, which can lead to many crashes and fatalities. The study reported by *The National Highway Traffic Safety Administration* (NHTSA) indicated that over 25% of police-reported crashes involved inattentive drivers [1]. The 100-car Naturalistic Study concluded that over 65% of near crashes and 78% of crashes included inattention [2]. These high percentages are not surprising, since Ranney estimated that about 30% of the time drivers are in a moving vehicle, they are engaged in secondary tasks that are potentially distracting [3]. These numbers are estimated to increase as the usage of in-vehicle technologies for navigation, communication and infotainment, and the number of cars on the roads are expected to exponentially increase in the next years [4]. A driver monitoring system that is able to sense inattentive drivers can play an important role in reducing the number of accidents, preventing fatalities and increasing the safety on roads.

Different studies have addressed the problem of understanding driver behaviors under distraction, using various features and approaches. Driver behaviors can be studied directly from cameras or other sensors. For example, the attention field of the driver can be estimated from his/her head pose and eye gaze information [5]. Recent studies have also considered the use of monocular, *infrared* (IR) and stereo cameras to track driver distractions [6]. *Electroencephalography* (EEG), *Electrocardiograph* (ECG), *Electrooculography* (EOG) and other similar invasive sensors have been considered to estimate relevant biometric signals associated to distraction [7]–[9]. Another important source of information can be provided by the car activities [10]. Driver behavior directly affects how the vehicle performs, which can be analyzed using the *Controller Area Network-Bus* (CAN-Bus) data. The common available information includes the vehicle speed, steering wheel angle and brake value. Some studies have proposed multimodal solutions, by considering multiple sensors [11]. These studies have been conducted in either real [12]–[15] or simulated conditions [11], [16], [17].

This paper presents a multimodal approach to track distraction in real driving scenarios, by using noninvasive sensors. The study aims to build statistical models for determining driver distraction from real-world data. A multimodal database is recorded with real driving conditions using the UTDrive platform [18]. Among other sensors, this car is equipped with a front facing video camera to capture the driver's face and a microphone array to capture the audio. It also provides CAN-Bus data describing the vehicle activity. In this study, driver distraction is defined as the voluntary or involuntary diversion of attention from the primary task of driving due to involvement in secondary tasks [19]. The distraction reduces the driver's decision making and situational awareness. This definition does not include distractions or impairments produced by alcohol, fatigue or drugs [20]–[22]. While these types of distraction are important, the secondary tasks considered in this study correspond to activities that are commonly performed by individuals while driving. The tasks are operating a radio (*Radio*), operating and following a navigation system (GPS) (*GPS - Operating* and *GPS* -

This work was done when the authors were with Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX, 75080, USA.

C. Busso is with the Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX, 75080 USA E-mail: busso@utdallas.edu.

J. Jain is with Belkin International, Inc., Playa Vista, CA, 90094.

Following), operating and talking on a cellphone (*Phone - Operating* and *Phone - Talking*), describing pictures (*Pictures*) and taking to a fellow passenger (*Conversation*).

The first goal of the paper is to analyze the effects in driver behaviors induced by secondary tasks. We present statistical analysis on multimodal features to identify significant differences in patterns observed during normal and each of the secondary task conditions [23]. We observe consistent changes in features automatically extracted from the frontal camera, microphone array and CAN-Bus signal when the driver is engaged in secondary tasks. The analysis is validated with binary and multiclass classification experiments. Binary classifiers are built to distinguish between normal and each of the secondary task conditions. The average accuracy achieved by these binary classifiers is 77.2% across tasks. Then, a single multi-class classifier is trained to recognize among normal and the seven task conditions (eight-class problem). This classifier achieves an accuracy of 40.8%, which is significantly higher than chances (12.5%). We observed that the classifiers' accuracy varies across secondary tasks, suggesting that certain tasks are more distracting than others (e.g., *GPS - Operating* versus *Conversation*). Motivated by this result, the second goal of this study is to build statistical models that quantify the actual deviations from the expected normal driving patterns. For this purpose, we propose the use of *Gaussian mixture models* (GMMs). The study presents task dependent and task independent GMMs. The *receiver operating characteristic* (ROC) for these models reveals that it is possible to quantify the deviations from normal driving behaviors. Our third and final goal is to leverage the results from the feature and model analysis to estimate a metric describing the distraction level of the drivers. Given the differences in distraction induced by secondary tasks, external evaluators are asked to annotate the perceived distraction level of short videos from the corpus. These subjective scores are used to build a multimodal regression framework. The results from the regression model highly correlate with the perceived driver distraction scores provided by subjective evaluations. The proposed methods and algorithms to address these three aims represent important contributions in the area of automatic detection of distracted behaviors.

The paper is organized as follows: Section II presents the state-of-the-art in the field of driver distraction analysis. It suggests the open challenges currently existing in this research area. Section III describes the methodology behind the data collection, the UTDrive platform and the protocol used to record the database. Section IV presents the data modalities, feature extraction procedure and corresponding preprocessing steps. Section V studies the effects in driver behaviors induced by secondary tasks, including statistical analysis of multimodal features, and discriminative analysis between normal and task driving conditions (*aim 1*). Section VI presents the statistical models to quantify the actual deviations from the expected normal driving patterns (*aim 2*). Section VII describes the multimodal regression analysis to predict the driver' distraction level (*aim 3*). Section VIII concludes the paper with final remarks, limitations of the study, and our future research directions.

II. Related Work

The area of monitoring driver behaviors has received a growing attention. Previous work has considered various modalities [11], [15], [18], [24] and different cognitive and visual distractions [6], [10], [23], [25], [26]. Some studies have considered driving simulators [11], [16], [17] or real car equipped with multiple sensors [6], [18], [27]. This section summarizes the current state-of-the-art on detecting driver distraction. For detailed surveys in this area, the readers are referred to Ahlström and K. Kircher [28], Bach et al. [29], Dong et al. [5], and Wu [30].

A. Relevant Modalities for Detecting Driver Distraction

Previous studies have proposed different sensing technologies including video cameras facing the driver [5], [10], [24], *Controller Area Network-Bus* (CAN-Bus) data [6], [15], [18], microphones [15] and invasive sensors to capture biometric signals [7], [9], [11].

Frontal cameras can be useful to assess the distraction level of the driver [13], [31]. Relevant visual features include head pose, gaze range and eyelid movements [6], [10], [17], [24], [32]. Liang et al. [10] showed that eye movements and driving performance measures were useful for detecting cognitive distraction. Su et al. [24] presented an approach to monitor visual distractions using a low cost camera. The study relied on eyelid movements and face orientation to predict driver's fatigue and distraction. Azman et al. [32] used eye and lip movements to predict cognitive distractions in simulated environment. Kutila et al. [6], [25] extracted gaze angle, head rotation and lane position for cognitive distraction detection. Bergasa et al. [27] proposed to predict fatigue with *percent eye closure* (PERCLOS), eye closure duration, blink frequency, face position, fixed gaze and nodding frequency. They used IR-illuminator to mitigate the changes in illumination. A similar approach was presented by Zhu and Ji [33]. Other studies have considered cameras for capturing and modeling foot gestures for brake assistance systems [34], [35].

Car information provides valuable features about the driver behaviors [10], [16], [17], [36], [37]. Ersal et al. [16] proposed a neural network model that uses the pedal position to predict driver behaviors. Tango and Botta [17] conducted their experiments in a driving simulator using steering wheel, vehicle speed, and lateral position to study the reaction time of drivers as an indicator of driver attention level. Sathyanarayana et al. [36] built driver-dependent GMM using basic driving actions, such as turns, lane changes and stops. The features were derived from CAN-Bus signals including wheel angle, gas and brake pressure.

Studies have considered physiological signals to infer cognitive load, attention and fatigue [7], [9], [11]. Among all physiological signals, *electroencephalography* (EEG) is the predominant and most used modality [29]. Damousis and Tzovaras [9] proposed a fuzzy fusion system using *electrooculogram* (EOG) for detecting drowsy driving behaviors. Putze et al. [11] measure multiple biosignals such as EEG, respiration and pulse. They analyzed visual and cognitive tasks in driving simulations. The main drawback of using physiological signal is that invasive sensors are usually needed, which are not convenient for real-world driving scenarios.

B. Inducing Visual and Cognitive Distractions

Secondary tasks deviate the driver's attention from the primary driving task [16]. Various activities have been proposed to induce cognitive and/or visual distractions. For cognitive distractions, common approaches include solving math problems [6], [11], [25], [38], talking to another passenger [6], [36], and focusing on other activities such as following the stock market [10]. Common secondary tasks for visual distraction are "look and find" tasks [10], [11], [17], operating devices such as a touchscreen [16], or a cellphone [23], and reading sequences of numbers [6]. While these cognitive and visual tasks clearly affect the driver, some of them may not represent the common distractions observed in real scenarios.

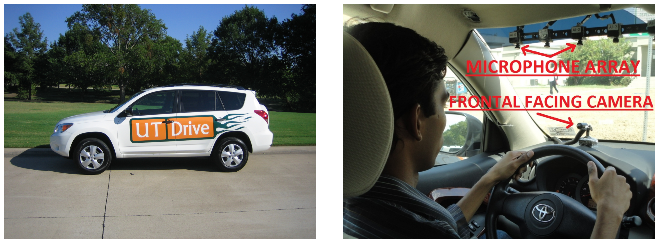
C. Driving Platforms

While most of the studies on driver behaviors rely on simulators [7], [9], [11], [16], [17], some studies have considered recordings in cars equipped with multiple sensors [6], [12], [13], [27], [36], [39]. Perez et al. [12] presented the "Argos" system for data collection. Murphy-Chutorian and Trivedi [13] reported results on data recorded in the LISA-P experimental testbed. The car has video and motion cameras with near-IR illuminator. They have used computer visual algorithms to automatically extract visual information, achieving promising results towards detecting driver distraction. Another data collection vehicle was designed by Takeda et al. [40]. The car is equipped with cameras and microphones, laser scanners (front, back), pedal pressure and physiological sensors. A similar car was designed by Abut et al. [41] called UYANIK. The UTDrive is another car platform, which will be used in this study (details are given in Sec. III-A) [15], [18]. These cars provide more realistic data to study driver behaviors.

Our approach derives its novelty from using noninvasive sensors to capture audio, video and CAN-Bus features from driving recordings in real-traffic situations. The analysis considers situations when the driver is performing common everyday tasks such as tuning the radio, operating and following a GPS, and operating a mobile device. In these realistic conditions, we proposed novel frameworks to quantify the distraction level of the drivers.

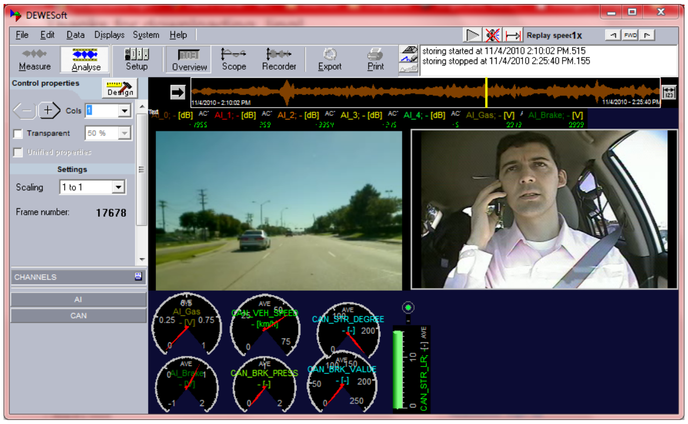
III. Methodology

The goal of the paper is to conduct a study on a real-world platform with drivers in normal and task conditions. A real-world driving study inherently involves numerous uncontrollable variables such as traffic, weather and traffic lights which are not easily replicated in a simulated environment. The proposed analysis aims to identify relevant features that are directly affected by the driver's behaviors, and to use these features to quantify the distraction level of the drivers.



(a) UTDrive Car (b) Setup

Fig. 1. UTDrive Car and Setup



storing started at 11/4/2010 2:10:02 PM.515
storing stopped at 11/4/2010 2:25:40 PM.155

11/4/2010 - 2:10:02 PM
11/4/2010 - 2:25:40 PM

Cols 1
□ Transparent 50 %
□ Unified properties

Settings
Scaling 1 to 1
Frame number: 17678

Fig. 2. Dewesoft interface that is used for recording and exporting the multimodal data.

A. UTDrive

The UTDrive (Fig. 1(a)) is a car platform belonging to The Center for Robust Speech Systems (CRSS) at The University of Texas at Dallas (UT Dallas) [15]. It is a 2006 Toyota RAV4 which has been custom fit with data acquisition systems with various sensors. It can extract and record various CAN-Bus signals, such as vehicle speed, steering wheel angle, brake value, and RPM acceleration. A pressure sensor on the gas pedal provides data for the gas pedal pressure and is extracted separately. It has a frontal facing video camera (PBC-700H), which is a key component of this work. The camera is mounted just behind the steering wheel on the dashboard facing the driver, as shown in Fig 1(b). The placement and small size of the camera are suitable to record frontal video of the driver's face without obstructing his/her field of vision. The resolution of the camera is set to 320×240 pixels and is being recorded at 30 fps. Another camera is placed facing the road, which records at 15 fps with the same resolution. All the modalities are being simultaneously recorded into a Dewetron computer which is placed behind the driver's seat. Fig. 2 shows the interface of the Dewesoft software, which is used for recording the data in the vehicle and for extracting the raw signal for each modality.

A GPS is mounted on the front windshield in the middle and adjusted as per the convenience of the driver before the recording. The radio is in its standard place, on the right side of the driver. For further details about the car and its unique features, the readers are referred to Angkititrakul et al. [18].

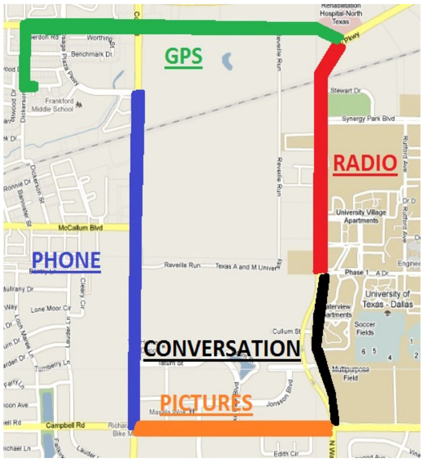


Fig. 3. Route used for the study. The subjects were asked to drive this route twice (5.6 miles long). During the first run, the driver performed a series of secondary tasks starting with *Radio* and ending with *Conversation*. During the second run, the subjects drove normally without any secondary task.

B. Database

The driving study included 20 subjects who were required to have a valid US Driving License and be at least 18 year olds. The average and standard deviation of the participants' age are 25.4 and 7.03, respectively. All of them were either university employees or students. The study was evenly distributed among 10 male and 10 female participants. The recordings were conducted during dry days with good light conditions to reduce the impact of the environment variables (e.g., reduced average speed as a result of wet roads). Notice that studies have reported that crashes related to distraction are more likely to occur with good light conditions and less traffic density, which validates our approach [42]. The subjects are advised to take their time while performing the tasks. The safety of the passengers was the most important priority.

C. Protocol

A 5.6 mile route, starting and ending at the university premises was selected for the test (Fig. 3). The route includes many traffic signals, stop signs, heavy and low traffic zones, residential areas and also a school zone. Although the current route does not include highways, we are planning to complement this corpus with new routes. The subjects took 13 to 17 minutes to complete the route.

Each subject was asked to drive this route twice. During the first run, the participants were asked to perform a series of secondary tasks. These tasks were selected to span different common activities performed by drivers that can lead to distraction. Some dangerous tasks such as text messaging were not included in the study to prevent accidents. The description of the selected tasks is given below:

- Operating the in-built car radio (Fig. 3, red route): The driver is asked to tune the radio to some predetermined stations.
- Operating and following instruction from the GPS (Fig. 3, green route): A pre-decided address is given to the driver to input into the GPS. Then, they are asked to follow the GPS instructions to reach the destination. This task is subdivided into *GPS - Operating* and *GPS - Following* (preliminary results suggested that driver behaviors are different for these two activities [23]).
- Operating and talking on the phone (Fig. 3, navy blue route): The driver is asked to call an airline automatic flight information system (toll-free) to retrieve flight information between any two given US cities, using a cellphone. This task is also subdivided into *Phone - Operating* and *Phone - Talking* for similar reasons as above. Notice that at the moment of the recordings, the State of Texas allowed drivers to use cellphones while driving.
- Describing pictures (Fig. 3, orange route): This task requires the driver to look and describe randomly selected pictures which are held out by a passenger seated beside the driver. The pictures are printed out in color on A4 size paper to avoid making this a difficult task. The purpose of this task is to simulate the task of looking at objects outside the car, such as billboards, sign boards and shops.
- Conversation with a passenger (Fig. 3, black route): The last task is a spontaneous conversation between the driver and a second passenger in the car. The driver is asked a few general questions in an attempt to get the driver involved in a conversation.

By splitting the phone and GPS tasks, seven tasks are considered. Tasks like *Radio*, *GPS - Operating*, *GPS - Following*, *Phone - Operating*, *Pictures* and *Conversation* are visually intensive at varying levels. *Phone - Talking* is a more cognitively intensive task.

The second lap involves normal driving without any task. The data collected from this lap is used as normal reference. The analysis is less dependent on the selected road, as the same route is used to compare normal and task conditions. Previous studies have followed a similar protocol to record driving behaviors in real roads, consisting in collecting data over a predefined route during which secondary tasks are performed in sequential order [15], [40], [41], [43]. By fixing the order of the tasks over predefined route segments, we can collect recordings that serve as reliable baseline for normal driving behavior, in which most of the other variables are kept fixed (e.g., route, traffic, and street signals). With this controlled recording we can study the differences in driving behaviors during tasks and normal conditions. The observed differences can be mainly associated with the behaviors induced by secondary tasks.

IV. Data Modalities

The collected data consists of three modalities: CAN-Bus features, visual features extracted from the frontal camera, and acoustic features extracted from the microphones. Table I summarizes all the features. The Dewesoft software is used to extract these streams of data – the low level features. This section presents these modalities and the preprocessing steps.

A. CAN-Bus Information

The CAN-Bus information consists of steering wheel angle in degrees, the vehicle speed in kilometers per hour (km/h),

TABLE I
FEATURE SUMMARY. LOW LEVEL FEATURES ARE TIME SERIES SIGNALS OVER WHICH WE ESTIMATE STATISTICS. HIGH LEVEL FEATURES ARE SINGLE VALUES DERIVED FROM THE ANALYSIS WINDOW (5 SEC).

Low Level Feature		Statistics
Video	Head Yaw Angle (Yaw)	Mean
	Head Pitch Angle (Pitch)	Standard Deviation (STD)
	Head Roll Angle (Roll)	Maximum (Max)
	Eye Closure Rate (Eye)	Minimum (Min)
Audio	Energy	Range
Can-Bus	Vehicle Speed (Speed)	Inter-Quartile Range (IQR)
	Steering Wheel Angle (Steering)	Skewness
	Brake Pressure (Brake)	Kurtosis
	Steering Wheel Jitter (Jitter)	
High Level Features (5 sec.)		
Eye Blink Frequency (Blink Freq.)		
Eyes-Off-Road Duration (EOR Dur.)		
Eyes-Off-Road Frequency (EOR Freq.)		

the brake value, the acceleration in revolutions per minute (rpm) and the brake and gas pedal pressures. Among these modalities, *steering wheel angle*, *vehicle speed*, and *brake pressure* are used to estimate the vehicle activity. It is observed in our experiments that drivers tend to reduce the vehicle speed while performing secondary tasks, either due to the distraction caused by the secondary task, or the driver's intention to perform both driving and secondary tasks safely. Therefore, these features are expected to be useful.

In addition to the exact value of the steering wheel angle, the *steering wheel jitter* is considered as a feature. It is calculated as the sequence of variance over 5 sec windows. It is hypothesized that the steering wheel jitter is directly affected by the driver behavior. When the driver is involved in secondary tasks, small corrections in the steering wheel will be frequently made to compensate drifts caused by the distraction. Therefore, the steering wheel jitter is expected to increase. During normal driving, the jitter is expected to be smoother. Table I summarizes the low level features derived from the CAN-Bus.

B. Frontal Facing Video Information

The video obtained from the camera facing the driver can provide valuable information about his/her behaviors. This study considers facial features describing head rotation and eye movement. These features are directly estimated from the video using the *computer expression recognition toolbox* (CERT) [44]. This toolkit was developed at the University of California San Diego as an end-to-end system for fully automated facial expression recognition. Notice that this toolkit has been used to detect fatigue during driving simulations [21], [22].

The head pose is parameterized with the pitch, yaw and roll angles. These angles are estimated with the algorithm included in CERT, which was developed by Whitehill et al. [45]. The algorithm is shown to be robust for large data sets and for varied illumination conditions, which is crucial for this study. Due to the limitation of the CERT software, information is lost when the head is rotated beyond a certain degree or when the face is occluded by the driver's hands. The algorithm produces empty data in those cases. However, one of the primary advantages of CERT is that the estimation is done frame by frame. This feature is important as the errors do not propagate across frames.

The eye movement information is directly estimated with CERT. The toolkit provides a numerical value describing the opening of the eyes (high values when the eye is close; low values when the eye is open). This value is referred to as *eye closure rate*, and it is used as feature. In addition, this study considers the *eye blink frequency*, which is derived from the eye closure rate. This variable is related to the percentage of eye closure or PERCLOSE. The driver is considered to be blinking when the eye closure rate provided by CERT is above a threshold. Notice that an adult blinks in average every 6 seconds. Each blink lasts approximately 200ms [46]. Therefore, we expect the driver to be blinking 3% of the time. The selected threshold considers this empirical result. We set the threshold as the mean of the eye closure rate plus two standard deviations. Assuming that the eye closure rate follows a Gaussian distribution, this threshold will select 2.5% of the frames. The mean and standard deviations of the eye closure rate are separately estimated for each driver using his/her recordings under normal driving condition.

Since the distractions induced by the selected secondary tasks are mostly visual, this study considers the *eyes-off-the-road duration* and *eyes-off-the-road frequency*. Studies have shown that when the eyes-off-the-road duration is greater than 2 seconds, the chances of accidents increase [29], [47]. Therefore, these features are important. We consider head yaw and pitch information for eyes-off-the-road detection (eyes can be off-the-road by turning head either horizontally or vertically). The head yaw and pitch values are numerical measures describing the relative horizontal and vertical rotations between the head and the camera, respectively. Since drivers have different height, their relative head positions with respect to the camera are different. Therefore, the study considers driver dependent thresholds which are estimated using the normal driving data. The head-off-the-road thresholds are set as the mean ± 1.5 standard deviation of the head yaw and pitch (statistics derived per driver). It is assumed that the drivers are glancing when either head yaw or pitch values are beyond these thresholds. This approach is consistent with approaches used in previous work to define the relevant field of view while driving [28]. Notice that the size of the square defined by these thresholds covers approximately 16 degrees. This value matches the threshold used to estimate the *percent road center* (PRC), which is defined as the percentage of time within 1 minute that the gaze falls in the 8 degree radius circle centered at the center of the road [48].

Table I summarizes the low level features derived from the frontal video camera. Notice that eye blink frequency, eyes off-road duration and eyes off-road frequency are high level features estimated over 5 sec windows (one value per window).

C. Audio Signal Information

The acoustic information is recorded using the microphone array. The channels are extracted using the Dewesoft software into separate audio files. The average speech energy is estimated from one of the microphones. This acoustic feature is relevant for secondary tasks characterized by sound or voice activity such as *GPS - Following*, *Phone - Talking*, *Pictures* and *Conversation*. Table I lists the low level features derived from the microphones.

D. Preprocessing

During real-world driving conditions, a driver has to stop or slow down due to traffic congestion or traffic signs (e.g., traffic lights and stop signs). During those stops, the driver may produce behaviors that are not related to the driving task. Since the focus of the paper is to analyze the behaviors observed when the car is moving, the study neglects segments in which the speed of the car is below 5km/h, as provided by the CAN-Bus data.

The CERT software requires close frontal views of the subject to extract reliable features. Although it tolerates varied head poses and moderate out-of-plane head motion, it fails to provide information for head rotations beyond ±15° or partial occlusion [44]. This issue is consistently observed during vehicle turns. Therefore, the analysis only considers the segments in which the car is moving straight. Data segments are neglected when the steering wheel angle is above 20° (empirically chosen), as provided by the CAN-Bus information. Any remaining gap in the data due to face rotation or hand obstruction is interpolated. These preprocessing steps are performed for each recording, including data during normal and task conditions.

V. Driver Distraction Analysis

The first goal of this study is to identify a set of distinctive features which are sufficiently representative of the variability observed across drivers while performing the secondary tasks. This problem is investigated with feature analysis (Sec. V-A), binary classifications between normal and secondary tasks (Sec. V-B), and multiclass classification across driving conditions (Sec. V-C).

A. Feature Analysis

Table I summarizes the nine low level features considered in this study. The data is segmented into 5 sec windows. For each of these segments, eight statistics are estimated from the low level features (see column *Statistics* in Table I). In addition, the three high level features described in section IV-B are included in the analysis. Altogether, a multimodal feature vector with 75 values is computed to describe the driver behaviors over each 5 sec window (9 low level features × 8 statistics + 3 high level features). The proposed analysis consists in analyzing the differences in the feature space between each secondary task and its corresponding normal condition (seven binary comparisons). Since streets have different characteristics, we compare the data under task conditions with the normal data collected during the corresponding route segments. This approach eliminates the route conditions factor introduced in the analysis (e.g., route segments have different speed limits).

A matched-pairs hypothesis test was performed to assess whether the differences in the features between each task and the corresponding normal condition are significant [49]. The matched variable in the analysis is the driver. A t-test is calculated, since the database consists of only 20 participants. Fig. 4 shows the results for some of the most relevant features considered in this study across each of the secondary tasks. The figure highlights in gray the features that are found significantly different (*p-value* = 0.05). It also provides the percentage of the features that are found significantly different per task (numbers after the tasks). The figure shows that the features *eyes off-road duration*, *audio mean* and *audio inter-quartile range* present significant differences across the different tasks. The figure also shows that there are tasks such as *GPS - Following* (28%) and *Phone - Operating* (21%), in which few of the selected features present significant differences. This result suggests that either the behaviors of the driver may not be significantly affected by these secondary tasks or that the selected features do not capture these differences. Likewise, there are secondary tasks such as *GPS - Operating* (40%) and *Pictures* (55%) that significantly change the values of the features.

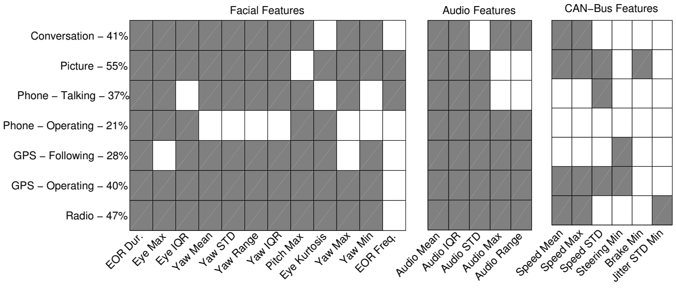
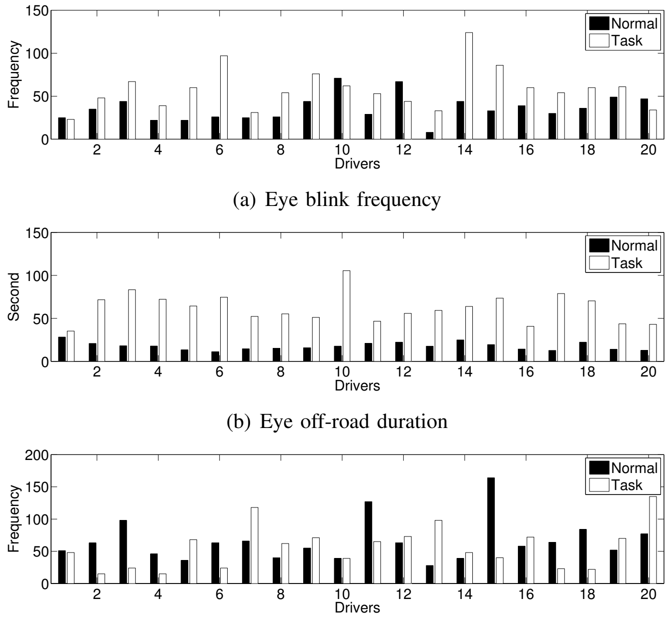


Fig. 4. Results of the matched pairs t-test: Features vs Tasks. For a particular task, gray regions indicate features that are found to be significantly different from normal conditions (*p-value* = 0.05). The figure also provides the percentage of the features that are found significantly different per task. The nomenclature of the features is given in Table I.

Fig. 5 provides further insights about the proposed high level features automatically derived from the video. The figure reports aggregate results across secondary tasks per driver (normal versus task conditions). Fig. 5(a) gives the total number of eye blinks during task and normal conditions per driver. Although there are some differences across drivers, the figure shows that the subjects blinked more when they were asked to perform secondary tasks. Fig. 5(b) gives the total number of seconds used for glancing. The figure shows that drivers spent more time with the eyes-off-the-road during task conditions. This result indicates the importance of this feature in the analysis of driver behaviors. Fig. 5(c) shows the number of times that the drivers glanced. The results from this feature are not conclusive, since the differences between both conditions are not significant (see Fig. 4). Notice that primary driving tasks such as checking the mirrors can be detected as an eyes-off-road action. Therefore, normal conditions can generate similar number of eyes-off-road instances as task conditions.



(a) Eye blink frequency

(b) Eye off-road duration

(c) Eye off-road frequency

Fig. 5. High level features extracted from 20 drivers for normal and task driving conditions.

B. Binary Classification: Normal versus Task Conditions

This section analyzes the discriminative power of the features by conducting separate binary classification experiments to recognize between normal and each of the task conditions (e.g., normal versus *Phone - Operating*). The features from normal condition are extracted during the corresponding route segments associated with the tasks (route-matched experiments). The goal is to identify the best feature set to characterize the driving behaviors observed when the subjects are engaged in secondary tasks.

The database is divided into 5 sec windows without overlap over which we estimate the 75-dimension feature vector (see Table I). For each of the seven binary problems, the dimension of the feature vector is reduced using *sequential floating forward selection* (SFFS). Starting from an empty set, SFFS add one feature at a time. After each forward step, the algorithm determines whether excluding a selected feature improves the objective function. The proposed criterion is to maximize the inter/intra class distance ratio. The goal is to select a feature set that preserves low intra class distance and high inter-class distance. Since the SFFS does not maximize the performance of a classifier, the feature set is independent of any particular machine learning algorithm (the focus of the analysis is on the features rather than the classifiers). The study considers *K-Nearest Neighbor* (k-NN) algorithm ($k = 5$ empirically chosen) and *support vector machine* (SVM). For SVM, we use linear and second order polynomial kernels. The classification experiments are implemented using "leave-one-out" cross validation approach. In each fold, we use a driver-independent partition with data from 19 drivers for training and data from one driver for testing. SFFS is used to select the feature set using the training partition. Then, the aforementioned classifiers are trained and evaluated using the selected feature set. The reported results correspond to the average across the 20 folds. We balance the number of samples from normal and task conditions during testing and training (chance is 50%).

TABLE II
AVERAGE ACCURACIES FOR BINARY (SEC. V-B) AND MULTICLASS (SEC. V-C) RECOGNITION PROBLEMS.

	Binary	**Multiclass**	
Algorithm	**Average Accuracy**	**#Feat.**	**Accuracy**
KNN	0.733	15	0.365
Linear SVM	0.772	11	0.361
Degree-2 SVM	0.760	9	0.408

Table II shows the average performance of the machine learning algorithms. The accuracies are averaged over the 20 folds and over the seven binary classification tasks. SVM with linear kernel provides the best result with 77.2% accuracy. Table III gives the detailed performance of the binary classifiers trained with SVM with linear kernel. The table reports the accuracy per task achieved with only video, audio and CAN-Bus features. The table also gives the accuracy when all the modalities are fused at the feature level. The results show that the features extracted from the video are the most discriminative features across all the secondary tasks with an average accuracy of 74.5%. The average accuracies for classifiers trained with acoustic and CAN-Bus features are 60.7% and 63.1%, respectively. However, the performance increases about 3% (absolute) when all the modalities are considered. This result is expected since the three modalities provide complementary information on various aspects of the distracted driving behaviors. CAN-Bus signal captures the direct effects on the vehicle caused by distractions. Therefore, it provides cues for various distraction types (e.g. visual distraction, cognitive distraction). Audio signal can be very useful for detecting sound-related distractions such as radio and passenger talking, which tends to increase the cognitive load of the drivers [25]. Features from the frontal camera provide valuable information about facial expression and head movement, which can signal the mental state and situation awareness of the drivers. By selecting features across modalities, the binary classifiers can identify task specific distractions. Another interesting observation is that even when one modality is used, the classifiers achieve performance above chance. This is particularly important when features from one modality are not available (e.g., videos with adverse illumination).

An interesting result is that secondary tasks are not equally recognized. Visually intensive tasks such as *Radio*, *GPS - Operating*, and *Picture* achieve accuracies over 80%. In our previous work, *Phone - Talking* was the most challenging task to recognize (59.1% – see [23]). By considering better features, we increase the accuracy to 73.2%. The task *GPS - Following* achieves the worst performance with 65.7%. This result is expected since only 28% of the considered features presented significant deviations from normal driver behaviors (see Fig. 4).

An important aspect of this evaluation is to identify discriminative features. Since the SFFS is estimated for each fold, the feature set may be different across folds. Table IV

TABLE III
AVERAGE ACCURACIES OF BINARY CLASSIFICATION USING DIFFERENT MODALITIES (SVM – LINEAR KERNEL).

Task	Video	Audio	CAN-Bus	All Features
Radio	0.793	0.606	0.667	0.807
GPS - Operating	0.773	0.620	0.760	0.831
GPS - Following	0.662	0.548	0.556	0.668
Phone - Operating	0.741	0.612	0.655	0.759
Phone - Talking	0.729	0.615	0.555	0.737
Picture	0.877	0.607	0.602	0.871
Conversation	0.637	0.646	0.558	0.729
Mean Across Tasks	0.745	0.607	0.631	0.772

lists the most frequently selected features used by the binary classifiers (SVM with linear kernel). The number of features corresponds to the dimension that maximizes the performance for that binary classification task (column *#Feat.*). The feature eye-off-road duration is chosen by most of the binary classification tasks, supporting the analysis presented in Section V-B. Features related to head roll angle, which was not used in our previous studies [23], [26], [50], are selected in three of the binary problems. This feature set is important for secondary tasks that force the drivers to tilt their head such as *Phone - Talking*. Notice that each of the binary classification tasks uses features derived from the three modalities (microphone, camera, CAN-Bus).

C. Multiclass Classification

A multiclass classifier is implemented to further explore the differences among the driving behaviors while performing secondary tasks. This classifier is trained to recognize between the seven secondary tasks and normal condition (8 classes). This multiclass problem can allow an active safety system to infer if the driver is engaged in a particular secondary task. It will also provide insights about relevant features across tasks.

We follow a similar procedure as the one used for binary classifiers (Sec. V-B). For each of the 20 folds, we randomly select an equal number of 5 sec windows per class, producing a balanced 8-class problem (chances is 12.5%). We use SFFS to train k-NN (k=8), and SVM with linear and second order polynomial kernels. The parameter k in k-NN is set to 8, which maximizes the recognition rate of the classifier. We separately select the dimension of the feature set to maximize the performance of the classifiers. Table II gives the results. SVM with second order polynomial kernel provides the best accuracy (40.8%), which is significantly higher than chances. Table IV lists the most selected features across the folds for this 8-class classifier (the feature set may be different across folds). The set includes features extracted from the modalities CAN-bus and camera.

VI. QUANTIFYING DEVIATION FROM NORMAL BEHAVIOR

The evaluation results in sections V-B and V-C show that it is possible to detect whether the driver is engaged in secondary tasks. The performances of the classifiers vary across the tasks. This result suggests that the multimodal features are not equally affected by these activities. Our second goal is to build models that can quantify the actual deviations from the expected normal driving patterns. These models will be valuable tools in the design of active safety systems that alert the drivers when their behaviors deviate from normal patterns beyond an acceptable threshold. For this purpose, we propose the use of *Gaussian mixtures models* (GMMs). GMM is a popular framework to capture the complex distribution of multimodal data. Equation 1 describes the probability distribution of an observation vector $X = x$ given a GMM parametrized by Θ:

$$P(X = \mathbf{x}|\Theta) = \sum_{j=1}^{K} \alpha_j \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_j|^{\frac{1}{2}}} e^{\frac{-1}{2}\left((\mathbf{x}-\mu_\mathbf{j})^T\Sigma_j^{-1}(\mathbf{x}-\mu_\mathbf{j})\right)} \tag{1}$$

where D is the dimension of the feature set and,

$$\Theta = \{\alpha_j, \mu_\mathbf{j}, \Sigma_j\}_{j=1}^{K} \qquad \alpha_j > 0 \;\; j = 1 \ldots K, \quad \sum_{j=1}^{K} \alpha_j = 1$$

where μ_j, Σ_j and α_j are the mean vector, covariance matrix and weighting coefficient of mixture j. We propose to train two reference GMMs. The first model Θ_n describes normal driving behaviors. It is trained with multimodal features extracted from the recordings when the drivers did not perform any secondary task. These GMMs capture the expected feature variability associated with the primary driving task. The second model Θ_t describes the behaviors observed when the drivers were engaged in secondary tasks. Given these two GMMs, we propose to quantify the deviation from normal driving behavior with the ratio $R(X|\Theta_n, \Theta_t)$ defined in Equation 2. If the driver displays normal behaviors, the likelihood $P(X = x|\Theta_n)$ will be higher than $P(X = x|\Theta_t)$ and $R(X|\Theta_n, \Theta_t)$ will be high. As the driver behaviors deviate from the expected normal patterns, the value of $R(X|\Theta_n, \Theta_t)$ will decrease.

$$R(X|\Theta_n, \Theta_t) = \frac{P(X = x|\Theta_n)}{P(X = x|\Theta_t)} \tag{2}$$

The *expectation maximization* (EM) algorithm is used to estimate the parameters (Θ). The maximum number of iterations was set to 200. The features correspond to statistics derived from the multimodal features during 5 sec windows. The normal and task GMMs are trained and tested with the same number of samples. Similar to the scheme used in Sections V-B and V-C, a 20-fold cross validation approach is implemented to maximize the usage of the database, while keeping the results driver independent.

A. GMM approach for classification

First, we demonstrate the approach by considering a constrained scenario in which separate GMMs are built for each task (i.e., $\Theta_n^{Radio}/\Theta_t^{Radio}$, $\Theta_n^{Conv.}/\Theta_t^{Conv.}$). The seven pairs of GMMs ($\Theta_n^{Task}/\Theta_t^{Task}$) are separately trained with the best 10, 15 and 20 features selected by the SFFS for the corresponding task in the binary evaluation (Sec. V-B). We report average results across folds.

We evaluate the performance of the proposed approach as a binary classification problem (normal versus task conditions).

TABLE IV
LIST OF MOST SELECTED FEATURES FOR BINARY (SEC. V-B) AND MULTICLASS (SEC. V-C) RECOGNITION PROBLEMS. SVM WITH LINEAR KERNEL IS USED AS CLASSIFIER FOR BINARY CLASSIFICATION AND SVM THE SECOND ORDER POLYNOMIAL KERNEL IS USED AS CLASSIFIER FOR MULTICLASS CLASSIFICATION (SEE TABLE I FOR THE NOMENCLATURE OF THE FEATURES).

Binary Tasks	**#Feat.**	**Selected Features**
Radio	23	EOR Dur.; Eye IQR; Yaw Mean; Speed Max; Speed Mean; Speed IQR; Steering Max; Yaw Kurtosis; Pitch Skewness; Brake IQR; Audio Mean; Audio Min; Audio IQR; Pitch Max; Roll Mean; EOR Freq.; Brake Mean; Brake Kurtosis; Jitter STD Min; Brake Min; Roll Min; Jitter STD Kurtosis; Roll Range
GPS - Operating	13	EOR Dur.; Speed Max; Audio IQR; Yaw Mean; Roll Min; Steering Min; Steering Skewness; Steering Kurtosis; Pitch Max; Speed STD; Eye Skewness; Speed IQR; Pitch Mean
GPS - Following	23	EOR Dur.; Blink Freq.; Steering Skewness; Audio STD; Pitch Mean; Audio Mean; Roll Min; Yaw Mean; Brake IQR; Jitter STD Min; Audio IQR; Audio Kurtosis; Eye Max; Yaw Min; Audio Max; Roll Kurtosis; Jitter STD Mean; Brake Max; Audio Skewness; Speed IQR; Roll Mean; Brake Range; Jitter STD Max
Phone - Operating	24	Eye STD; Pitch IQR; Audio Max; Audio Range; Eye Range; EOR Dur.; Jitter STD Skewness; Speed STD; Speed IQR; Audio IQR; Brake Range; Brake Kurtosis; Audio Mean; Audio Kurtosis; Roll Max; Roll IQR; Steering IQR; Audio Skewness; Eye Kurtosis; Pitch STD; Pitch Skewness; Roll Min; Roll Kurtosis; Speed Mean
Phone - Talking	16	Roll Mean; EOR Dur.; Audio Mean; Eye Mean; Yaw Min; Audio STD; Pitch Min; Brake Max; Roll Min; Yaw STD; Roll Range; Jitter STD Kurtosis; Speed STD; Steering Skewness; Blink Freq.; Speed Range
Picture	15	Audio Max; Yaw Min; Audio IQR; EOR Dur.; Audio Range; Eye STD; Yaw IQR; Pitch Min; Steering IQR; Jitter STD Min; Brake Kurtosis; Speed STD; Brake Min; Brake Skewness; Pitch IQR
Conversation	11	Audio Skewness; Audio Kurtosis; EOR Dur.; Audio IQR; Speed Max; Audio Max; Yaw Mean; Brake STD; Steering Min; Audio Range; Speed Range
Multiclass Task	**#Feat.**	**Selected Features**
	9	Eye STD; Steering Min; Eye Skewness; EOR Dur.; Yaw Min; Pitch Max; Speed Range; Speed Max; Jitter STD Max

TABLE V
TASK INDEPENDENT AND TASK DEPENDENT GMMS. THE TABLE REPORTS THE *area under the curve* (AUC) AND *equal error rate* (EER).

	Task Dependent GMMs (AUC / EER)		
Mixture Number	Feature = 10	Feature = 15	Feature = 20
4	0.976 / 0.087	0.963 / 0.078	0.914 / 0.165
8	0.970 / 0.087	0.979 / 0.078	0.937 / 0.155
16	0.974 / 0.087	0.987 / 0.078	0.975 / 0.178
	Task Independent GMMs (AUC / EER)		
Mixture Number	Feature = 10	Feature = 15	Feature = 20
4	0.869 / 0.233	0.943 / 0.132	0.768 / 0.311
8	0.833 / 0.252	0.771 / 0.292	0.638 / 0.463
16	0.967 / 0.078	0.957 / 0.100	0.852 / 0.222

By changing the threshold on $R(X|\Theta_n, \Theta_t)$, we estimate the *receiver operating characteristic* (ROC). The ROC curve is a plot describing the relationship between true positive rate and false positive rate. Each point in the graph corresponds to a different threshold on $R(X|\Theta_n, \Theta_t)$. From the ROC, we calculate the *area under the curve* (AUC) and the *equal error rate* (EER). A good classifier will have a higher AUC and a lower EER. These metrics are upper bounded by 1 and lower bounded by 0. We evaluate different configurations for the GMMs by changing the number of mixtures and the number of features. Table V shows the performance of the task dependent GMMs for different configurations. The best performance is achieved with 10 features and 16 mixtures. Fig. 6 shows the corresponding ROC curve for the task dependent GMMs with this configuration (dashed line).

The task dependent GMMs are not practical in real applications, since it assumes that the potential secondary task is known. This problem can be addressed by using the classifiers presented in section V-C to infer the most likely secondary task. Then, we can use the corresponding GMMs in Equation 2. An alternative approach is to train a single GMM with the data extracted from all the secondary tasks (i.e., Θ_n/Θ_t). This task-independent GMM will capture the driving behaviors observed across tasks. We follow the second approach because (i) it can generalize better to secondary tasks not considered in this study, (ii) it is a simple approach, and (iii) it gives good results, as described below. These GMMs are trained with the best 10, 15 and 20 features selected by the SFSS in the multiclass evaluation (Sec. V-C).

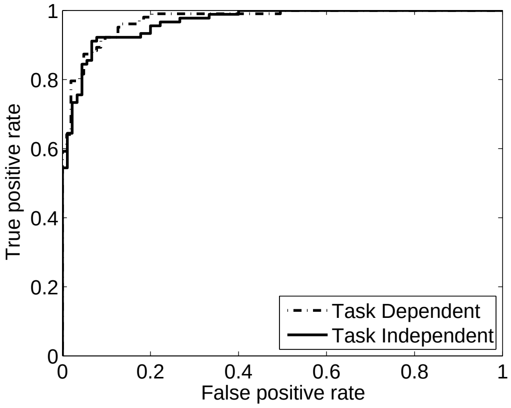


Fig. 6. ROC curve between task dependent and task independent GMMs for one fold (16 mixtures and 10 features).

Table V shows the results of task independent GMMs for different configurations. As expected, the AUC values are lower and the EER values are higher than the corresponding results achieved with task dependent GMMs. However, good performances are observed for certain configurations (e.g., 16 mixtures and 10 or 15 features). Fig. 6 shows the ROC curve for the task independent GMMs with 16 mixtures and 10 features (solid line).

B. Deviations from Expected Driving Behaviors

While the ratio $R(X|\Theta_n, \Theta_t)$ is useful to distinguish between task and normal conditions, the metric provides an objective value of the deviations in a recording from the

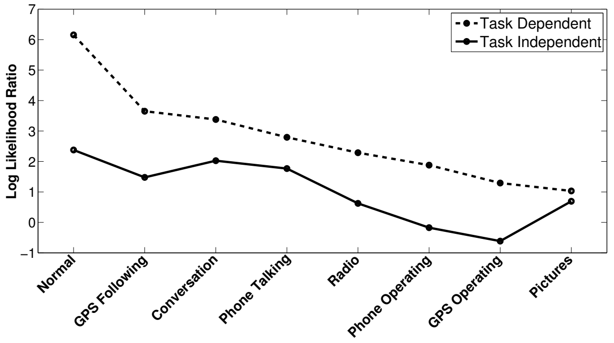


Fig. 7. Average value of $R(X|\Theta_n, \Theta_t)$ across tasks. Lower values implies higher deviations from normal driving behaviors.

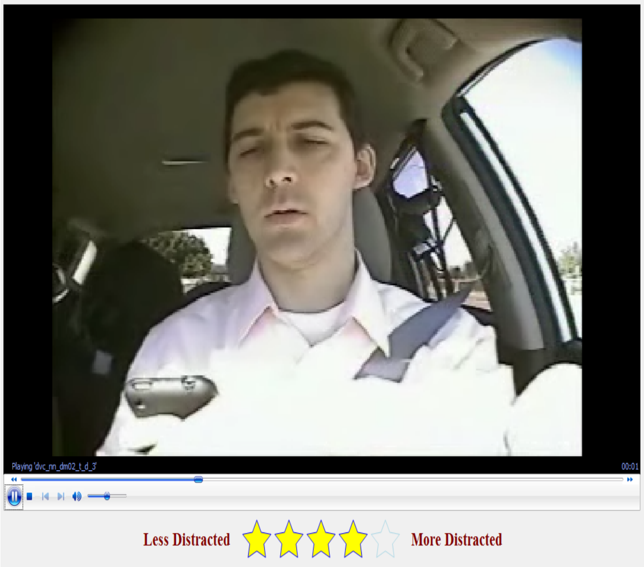
expected normal driving behaviors. We illustrate this point in Fig. 7, which shows the mean of the log likelihood ratio across the normal and seven task conditions using both task-dependent and task-independent models (16 mixtures and 10 features). The tasks are ranked in descend order for the task dependent models. Notice that lower values for $R(X|\Theta_n, \Theta_t)$ implies higher deviations from normal patterns. The tasks *GPS - Following* and *Conversation* induce driving behaviors that are closer to the expected normal patterns. The tasks that deviate the most from normal behaviors are *Radio*, *Phone - Operating*, *GPS - Operating*, and *Picture*. These results are consistent with the binary classification performances, where tasks inducing less deviated behaviors have lower accuracies (see Table III). One possible explanation is that the selected features are not suitable to describe the distractions induced by these tasks. However, the perceptual evaluation analysis presented in Section VII-A reveals that the perceived distraction scores by external evaluators are consistent with the findings reported in this section (see Fig. 9). These results validate the use of $R(X|\Theta_n, \Theta_t)$ for classification, and for quantifying the deviation of driving behaviors from normal patterns.

VII. Driver Distraction Metric

The ultimate goal of this work is to provide a metric for driver distraction using multimodal features. Some studies consider recordings in which drivers perform secondary tasks as positive examples of *distraction*. Controlled recordings are considered as *normal* [10], [32]. However, the results presented in Section VI reveal that secondary tasks induce different distraction levels. To build a distraction warning system, therefore, we need a metric that captures the intrinsic distraction induced during the recordings. For this purpose, we conducted a perceptual evaluation to derive a ground truth for driver distraction (Sec. VII-A). We use this metric to build regression models to quantify the driver' distraction level (Sec. VII-B).

A. Perceptual Evaluation for Driver Distraction

The corpus was split into 5 sec videos with synchronized audio. Each class was equally represented in the evaluation – seven tasks and normal conditions. For each of the 20 subjects in our database, 24 videos were randomly chosen (three for each task). In total, 480 unique videos were selected for evaluation.


Playing 'dvc_nn_dm02_t_d_3'
00:01

Less Distracted ★★★★☆ More Distracted

Fig. 8. Subjective evaluation GUI. The subjects are asked to rate the perceived distraction level of the drivers (1 for less distracted, 5 for more distracted).

Nine students from UT Dallas were asked to evaluate the perceived distraction level of the drivers. Fig. 8 shows the *graphical user interface* (GUI) that was built for this subjective evaluation. After watching each video, the evaluators rated on a scale from 1 to 5 the level of distraction of the driver (1 - *Less distracted*; 5 - *More distracted*). The definitions of distraction presented in Section I was read to the evaluators before the evaluation to unify their understanding of distraction. To minimize the duration of the evaluation, each evaluator was requested to complete only 160 videos, corresponding to one video per task, and per driver (20 drivers × 8 tasks). The average duration of the evaluation was approximately 15 minutes. The presentation of the videos was randomized to avoid biases. In total, each video was evaluated by 3 independent evaluators.

Fig. 9 shows the error plots with the perceived distraction level of the drivers. The figure provides the average and standard deviation values for the seven tasks and normal conditions. The results suggest that *GPS - Operating*, *Phone - Operating* and *Pictures* are perceived by the evaluators as the most distracting tasks. *GPS - Following* is not perceived to be as distracting as other tasks such as *Phone - Talking* and *Conversation*.

B. Distraction Evaluation and Metric

A linear regression model is built to measure the driver distraction level. The multimodal features described in Table I are used as independent variables. The average distraction levels obtained from the subjective evaluations are used as a dependent variable (see Equation 3). The evaluation in this section only considers the subset of the data that was perceptually evaluated.

$$y = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \cdots + \beta_F f_F \qquad (3)$$

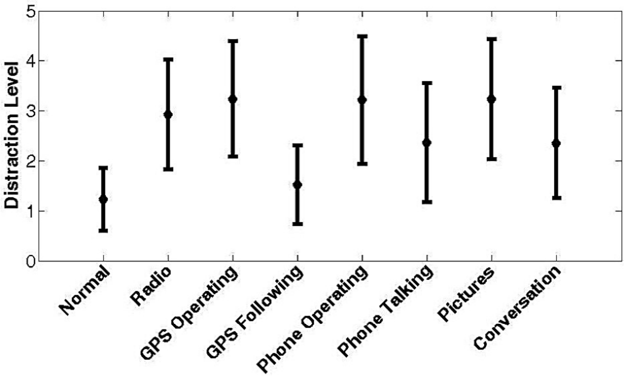


Fig. 9. Average distraction levels based on the subjective evaluations of all the subjects across classes.

TABLE VI
LINEAR REGRESSION VERSUS SVR

	Training		Testing	
	Correlation	**MSE**	**Correlation**	**MSE**
Linear Regression	0.66	0.66	0.61	0.74
SVR	0.69	0.60	0.55	1.11

A 20-fold cross validation approach is implemented to maximize the usage of the corpus. We implement the SFFS scheme using stepwise method. The stepwise approach uses the F-statistic to compare two nested regression models (the features of one model are a subset of the features of the other). At each step, the p-value is calculated for the F-statistic of the models with and without a potential independent variable. We add the best independent variable when the p-value of the F-statistics is below the entrance tolerance (0.05). Likewise, we remove selected independent variables when the p-value of the F-statistics is above the exit tolerance (0.10). This procedure continues until no change is made to the selected independent variable set. Notice that the training data changes across folds. We selected nine features in average (minimum 8, maximum 12). Although the selected feature sets vary, some features are frequently selected such as EOR Freq., eye skewness, head roll skewness, and eye STD.

Table VI gives the performance of the regression model in terms of correlation and *mean square error* (MSE). Predicted values are clipped if the scores are outside the range [1-5]. The average coefficient of determination R^2 across the folds is 0.42, which gives a correlation of $\rho^{train} = 0.66$. Table VI gives the average results for the testing partition across the 20 folds. The model predicts the perceived driver distraction level with $\rho^{test} = 0.61$. Given that the correlation during training and testing are similar, we conclude that the regression models have good generalization. These results show the strong correlation between the proposed metric and the perceptual evaluations. These models can serve as a valuable tool for new active safety systems that aim to monitor the distraction level of the drivers.

We also trained *support vector regression* (SVR) with linear kernel as an alternative to the linear regression model described in Equation 3. The result indicates that this approach does not generalize well (see Table VI). SVR tends to over-fit the data, giving lower performance during testing.

VIII. CONCLUSION

The paper presented our efforts to monitor the distraction level of the drivers. The study considered noninvasive sensors to capture the behaviors of drivers engaged in common secondary tasks such as operating a phone and a GPS. The evaluation relies on real driving recordings using the UTDrive platform. The study presented statistical analyses to identify relevant multimodal features extracted from a frontal camera, a microphone array and the CAN-Bus signal. Binary and multiclass recognition experiments indicated that the proposed features can be used to distinguish between normal and task conditions. Furthermore, the paper proposed a framework based on GMMs to quantify the deviations of the driver behavior from the expected normal patterns. The approach achieved promising results suggesting that it is possible to measure the distraction level of the drivers. Motivated by these results, a linear regression model is built as a metric of driver distraction. The prediction of the proposed model strongly correlates with subjective evaluations describing distractions.

A limitation of the study is that the corpus was recorded using a predefined route, during which the drivers were asked to perform secondary tasks in sequential order. The protocol, which was inspired by previous studies [15], [40], [41], [43], was used to collect a controlled corpus, reducing some of the multiple variability sources observed in real driving scenarios. Also, the study relies on recordings collected in urban roadways with specific speed limits and traffic signals. However, studies have shown that driver behavior changes under different environment conditions [25]. The effects of secondary tasks are also likely to depend on the traffic conditions. Therefore, our next data collection will consider different conditions including other residential roads and highways. It will also include different weather and illuminations conditions. Another limitation of the corpus is that secondary tasks were always collected during the first lap and the reference recordings during the second lap. This protocol ignores learning effects on the drivers. In the future, we will use a modified protocol consisting of 3 laps. During the first lap, the drivers will get familiar with the car. The recordings will not be considered in the analysis. During the second and third laps, the subjects will drive either normal or performing secondary tasks. The order will be randomized to minimize the learning effects.

This study provides a strong foundation for further research in the area of active safety systems. We are exploring other visual features that may be relevant to characterize driver distraction. For example, we hypothesize that facial expressions may provide insights about the cognitive load of the drivers. To capture the variability introduced by recordings during less restrictive conditions, we are planning to include route or driving maneuver dependent models (e.g., specialized model for “changing lane”). A real-time algorithm with such capabilities will have an impact on the design of feedback systems that are able to alert the drivers when their attention falls below an acceptable level. This driver-centric active safety system will help to prevent accidents, improving the overall driving experience on the roads

Acknowledgment

The authors would like to thank the anonymous reviewers for their thoughtful and insightful comments. The authors would also like to thank Dr. John Hansen for his support with the UTDrive Platform. We want to thank the Machine Perception Lab (MPLab) at The University of California, San Diego for providing the CERT software. The authors are also thankful to Ms. Rosarita Khadij M Lubag for her support and efforts with the data collection.

References

[1] T. Ranney, W. Garrott, and M. Goodman, "NHTSA driver distraction research: Past, present, and future," National Highway Traffic Safety Administration, Technical Report Paper No. 2001-06-0177, June 2001.

[2] V. Neale, T. Dingus, S. Klauer, J. Sudweeks, and M. Goodman, "An overview of the 100-car naturalistic study and findings," National Highway Traffic Safety Administration, Technical Report Paper No. 05-0400, June 2005.

[3] T. Ranney, "Driver distraction: A review of the current state-of-knowledge," National Highway Traffic Safety Administration, Technical Report DOT HS 810 787, April 2008.

[4] M. Broy, "Challenges in automotive software engineering," in *Proceedings of the 28th international conference on Software engineering (ICSE 2006)*, Shanghai, China, May 2006.

[5] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," in *IEEE Intelligent Vehicles Symposium*, Xi'an, Shaanxi, China, June 2009, pp. 875–880.

[6] M. Kutila, M. Jokela, G. Markkula, and M. Rue, "Driver distraction detection with a camera vision system," in *IEEE International Conference on Image Processing (ICIP 2007)*, vol. 6, San Antonio, Texas, USA, September 2007, pp. 201–204.

[7] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2726–2738, December 2005.

[8] C. Berka, D. Levendowski, M. Lumicao, A. Yau, G. Davis, V. Zivkovic, R. Olmstead, P. Tremoulet, and P. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, Space, and Environmental Medicine*, vol. 78, no. 5, pp. 231–244, May 2007.

[9] I. Damousis and D. Tzovaras, "Fuzzy fusion of eyelid activity indicators for hypovigilance-related accident prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 491–500, September 2008.

[10] Y. Liang, M. Reyes, and J. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 340–350, June 2007.

[11] F. Putze, J.-P. Jarvis, and T. Schultz, "Multimodal recognition of cognitive workload for multitasking in the car," in *International Conference on Pattern Recognition (ICPR 2010)*, Istanbul, Turkey, August 2010.

[12] A. Perez, M. Garcia, M. Nieto, J. Pedraza, S. Rodriguez, and J. Zamorano, "Argos: An advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 463–473, June 2010.

[13] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 300–311, June 2010.

[14] A. Giusti, C. Zocchi, and A. Rovetta, "A noninvasive system for evaluating driver vigilance level examining both physiological and mechanical data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 127–134, March 2009.

[15] P. Angkititrakul, M. Petracca, A. Sathyanarayana, and J. Hansen, "UTDrive: Driver behavior and speech interactive systems for in-vehicle environments," in *IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007, pp. 566–569.

[16] T. Ersal, H. Fuller, O. Tsimhoni, J. Stein, and H. Fathy, "Model-based analysis and classification of driver distraction under secondary tasks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 692–701, September 2010.

[17] F. Tango and M. Botta, "Evaluation of distraction in a driver-vehicle-environment framework: An application of different data-mining techniques," in *Advances in Data Mining. Applications and Theoretical Aspects*, ser. Lecture Notes in Computer Science, P. Perner, Ed. Berlin, Germany: Springer Berlin / Heidelberg, 2009, vol. 5633, pp. 176–190.

[18] P. Angkititrakul, D. Kwak, S. Choi, J. Kim, A. Phucphan, A. Sathyanarayana, and J. Hansen, "Getting start with UTDrive: Driver-behavior modeling and assessment of distraction for in-vehicle speech systems," in *Interspeech 2007*, Antwerp, Belgium, August 2007, pp. 1334–1337.

[19] I. Trezise, E. Stoney, B. Bishop, J. Eren, A. Harkness, C. Langdon, and T. Mulder, "Inquiry into driver distraction: Report of the road safety committee on the inquiry into driver distraction," Road Safety Committee, Parliament of Victoria, Melbourne, Victoria, Australia, Technical Report No. 209 Session 2003-2006, August 2006.

[20] T. Rahman, S. Mariooryad, S. Keshavamurthy, G. Liu, J. Hansen, and C. Busso, "Detecting sleepiness by fusing classifiers trained with novel acoustic features," in *12th Annual Conference of the International Speech Communication Association (Interspeech'2011)*, Florence, Italy, August 2011, pp. 3285–3288.

[21] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Human-Computer Interaction*, ser. Lecture Notes in Computer Science, M. Lew, N. Sebe, T. Huang, and E. Bakker, Eds. Berlin, Germany: Springer Berlin / Heidelberg, December 2007, vol. 4796, pp. 6–18.

[22] E. Vural, M. Bartlett, G. Littlewort, M. Cetin, A. Ercil, and J. Movellan, "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions," in *International Conference on Pattern Recognition (ICPR 2010)*, Istanbul, Turkey, August 2010, pp. 3874–3877.

[23] J. Jain and C. Busso, "Analysis of driver behaviors during common tasks using frontal video camera and CAN-Bus information," in *IEEE International Conference on Multimedia and Expo (ICME 2011)*, Barcelona, Spain, July 2011.

[24] M. Su, C. Hsiung, and D. Huang, "A simple approach to implementing a system for monitoring driver inattention," in *IEEE International Conference on Systems, Man and Cybernetics (SMC 2006)*, vol. 1, Taipei, Taiwan, October 2006, pp. 429–433.

[25] M. Kutila, M. Jokela, T. Mäkinen, J. Viitanen, G. Markkula, and T. Victor, "Driver cognitive distraction detection: Feature estimation and implementation," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 221, no. 9, pp. 1027–1040, September 2007.

[26] C. Busso and J. Jain, "Advances in multimodal tracking of driver distraction," in *Digital Signal Processing for In-Vehicle Systems and Safety*, J. Hansen, P. Boyraz, K. Takeda, and H. Abut, Eds. New York, NY, USA: Springer, December 2011, pp. 253–270.

[27] L. Bergasa, J. Nuevo, M. Sotelo, R. Barea, and M. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63–77, March 2006.

[28] C. Ahlström and K. Kircher, "Review of real-time visual driver distraction detection algorithms," in *International Conference on Methods and Techniques in Behavioral Research (MB 2010)*, Eindhoven, The Netherlands, August 2010, pp. 2:1–2:4.

[29] K. M. Bach, M. Jaeger, M. Skov, and N. Thomassen, "Interacting with in-vehicle systems: understanding, measuring, and evaluating attention," in *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, Cambridge, United Kingdom, September 2009.

[30] Q. Wu, "An overview of driving distraction measure methods," in *IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design (CAID CD 2009)*, Wenzhou, China, November 2009.

[31] M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 108–120, March 2007.

[32] A. Azman, Q. Meng, and E. Edirisinghe, "Non intrusive physiological measurement for driver cognitive distraction detection: Eye and mouth movements," in *International Conference on Advanced Computer Theory and Engineering (ICACTE 2010)*, vol. 3, Chengdu, China, August 2010.

[33] Z. Zhu and Q. Ji, "Real time and non-intrusive driver fatigue monitoring," in *IEEE International Conference on Intelligent Transportation Systems*, Washington, DC, October 2004, pp. 657–662.

[34] J. McCall and M. Trivedi, "Driver behavior and situation aware brake assistance for intelligent vehicles," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 374–387, February 2007.

[35] C. Tran, A. Doshi, and M. Trivedi, "Modeling and prediction of driver behavior by foot gesture analysis," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 435–445, March 2012.

[36] A. Sathyanarayana, P. Boyraz, Z. Purohit, R. Lubag, and J. Hansen, "Driver adaptive and context aware active safety systems using CAN-bus signals," in *IEEE Intelligent Vehicles Symposium (IV 2010)*, San Diego, CA, USA, June 2010.

[37] J. Yang, T. Chang, and E. Hou, "Driver distraction detection for vehicular monitoring," in *Annual Conference of the IEEE Industrial Electronics Society (IECON 2010)*, Glendale, AZ, USA, November 2010.

[38] J. Harbluk, Y. Noy, P. Trbovich, and M. Eizenman, "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," *Accident Analysis and Prevention*, vol. 39, no. 2, pp. 372–379, March 2007.

[39] A. Sathyanarayana, S. Nageswaren, H. Ghasemzadeh, R. Jafari, and J.H.L.Hansen, "Body sensor networks for driver distraction identification," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES 2008)*, Columbus, OH, USA, September 2008.

[40] K. Takeda, J. Hansen, P. Boyraz, L. Malta, C. Miyajima, and H. Abut, "International large-scale vehicle corpora for research on driver behavior on the road," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1609–1623, December 2011.

[41] H. Abut, H. Erdoğan, A. Erçil, B. Çürüklü, H. Koman, F. Taş, A. Argunşah, S. Coşar, B. Akan, H. Karabalkan *et al.*, "Real-world data collection with "UYANIK"," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, H. Erdoğan, J. Hansen, and H. Abut, Eds. New York, NY, USA: Springer, 2009, pp. 23–43.

[42] P. Green, "The 15-second rule for driver information systems," in *Intelligent Transportation Society (ITS) America Ninth Annual Meeting*, Washington, DC, USA, April 1999.

[43] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 97 – 120, March 2005.

[44] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, pp. 22–35, September 2006.

[45] J. Whitehill and J. Movellan, "A discriminative approach to frame-by-frame head pose tracking," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, Amsterdam, The Netherlands, September 2008.

[46] P. Caffier, U. Erdmann, and P. Ullsperger, "Experimental evaluation of eye-blink parameters as a drowsiness measure," *European Journal of Applied Physiology*, vol. 89, no. 3-4, pp. 319–325, May 2003.

[47] S. Klauer, T. Dingus, V. Neale, J. Sudweeks, and D. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," National Highway Traffic Safety Administration, Blacksburg, VA, USA, Technical Report DOT HS 810 594, April 2006.

[48] T. Victor, J. Engström, and J. Harbluk, "Distraction assessment methods based on visual behavior and event detection," in *Driver distraction: theory, effects, and mitigation*, M. Regan, J. Lee, and K. Young, Eds. Boca Raton, FL, USA: CRC Press, October 2008, pp. 135–165.

[49] W. Mendenhall and T. Sincich, *Statistics for Engineering and the Sciences*. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.

[50] J. Jain and C. Busso, "Assessment of driver's distraction using perceptual evaluations, self assessments and multimodal feature analysis," in *5th Biennial Workshop on DSP for In-Vehicle Systems*, Kiel, Germany, September 2011.

Nanxiang Li (S'2012) received his B.S. degree (2005) in Electrical Engineering from Xiamen University, Fujian, China. He received his M.S. degree (2009) in Electrical Engineering from University of Alabama, Tuscaloosa, Alabama, USA. He is currently pursuing the Ph.D. degree in Electrical Engineering at the University of Texas at Dallas (UTD), Richardson, Texas, USA. He joined the Multimodal Signal Processing (MSP) laboratory at UTD in 2011. His research interests include human attention modeling in the context of driving behavior analysis, and multimodal interfaces with emphasis on gaze estimation. He has also worked on human tracking and recognition using gait information.

Jinesh J Jain obtained his Bachelor of Engineering degree (2008) in Electronics and Communication with high honors from PES Institute of Technology, Bangalore, India affiliated to Visveswaraya Technological University, Belgaum, and his M.S degree (2011) in Electrical Engineering from the University of Texas at Dallas (UTD), Richardson, Texas, USA. From 2010 to 2011, he was a Research Assistant at the Multimodal Signal Processing (MSP) laboratory at UTD. He received the Hewlett Packard Best Paper Award at the IEEE ICME 2011 (with Dr. Carlos Busso). His research interests are in multimodal signal processing, pattern recognition and machine learning. He is currently working at Belkin International Inc, Playa Vista, California in the field of Non-Intrusive Load Monitoring systems and Home Automation Systems. He has also worked on multi-person detection, estimation and tracking, speech processing algorithms for Cochlear implants and Pattern classification of fMRI images for Face Recognition.

Carlos Busso (S'02-M'09) is an Assistant Professor at the Electrical Engineering Department of The University of Texas at Dallas (UTD). He received his B.S (2000) and M.S (2003) degrees with high honors in electrical engineering from University of Chile, Santiago, Chile, and his Ph.D (2008) in electrical engineering from University of Southern California (USC), Los Angeles, USA. Before joining UTD, he was a Postdoctoral Research Associate at the Signal Analysis and Interpretation Laboratory (SAIL), USC. He was selected by the School of Engineering of Chile as the best Electrical Engineer graduated in Chile in 2003. At USC, he received a Provost Doctoral Fellowship from 2003 to 2005 and a Fellowship in Digital Scholarship from 2007 to 2008. He received the Hewlett Packard Best Paper Award at the IEEE ICME 2011 (with J. Jain). He is the co-author of the winner paper of the Classifier Sub-Challenge event at the Interspeech 2009 emotion challenge. At UTD, he leads the Multimodal Signal Processing (MSP) laboratory [http://msp.utdallas.edu]. His research interests are in digital signal processing, speech and video processing, and multimodal interfaces. His current research includes the broad areas of in-vehicle modeling of driver behavior, affective computing, multimodal human-machine interfaces, modeling and synthesis of verbal and nonverbal behaviors, sensing human interaction, and machine learning methods for multimodal processing.