Assessment of driver's distraction using perceptual evaluations, self assessments and multimodal feature analysis

Jinesh J Jain and Carlos Busso

Multimodal Signal Processing Lab, The University of Texas at Dallas, USA
E-mail: jinesh.jain@utdallas.edu, busso@utdallas.edu

Abstract Developing feedback systems that can detect the attention level of the driver can play a key role in preventing accidents by alerting the driver about possible hazardous situations. Monitoring driver's distraction is an important research problem, especially with new forms of technology that are made available to drivers, which can interfere with the primary driving task. An important question is how to define reference labels that can be used as ground truth to train machine learning algorithms to detect distracted drivers. The answer to this question is not simple since drivers are affected by visual, cognitive, auditory, psychological and physical distractions. This paper explores and compares three different approaches to characterize driver's distraction: perceptual evaluation from external evaluators, self evaluations collected from post driving questionnaires, and analysis of the differences observed across multimodal features from normal patterns.

Keywords Driver distraction, active safety, driver perception, subjective evaluation, driving performance.

1. INTRODUCTION

One of the main causes of car accidents is drivers that become distracted by secondary tasks. There is a growing concern that the number of accidents may increase due to the development of new in-vehicle technologies, which can negatively affect the attention level of the driver [5]. According to the study reported by *The National Highway Traffic Safety Administration* (NHTSA), over 25% of police-reported crashes involved inattentive drivers [14]. This finding is not surprising since it is estimated that about 30% of the time that drivers are in a moving vehicle, they are engaged in secondary tasks [13]. Detecting distracted drivers is an important research problem to prevent accidents and increase the security on the roads.

Common secondary tasks can deviate the attention of the drivers from the primary driving task. These in-cab demands can produce visual, cognitive, auditory, psychological and physical distractions. Therefore, it is very important to understand the effect induced by different secondary tasks on the drivers [5]. A key step in the analysis is to define reference metrics or criteria to assess the attention level of the driver. These reference labels can be used as ground truth to train machine learning algorithms to detect distracted drivers.

There are many measures that can be used to determine driver distraction. These measures include lateral control measures (e.g., lane-related measures), longitudinal control measures (e.g., accelerator-related measures, brake and deceleration-related measures), obstacle and event detection (e.g., probability of detection measures), driver response measures (e.g., stimulus-response measures), vision related measures (e.g., visual allocation to roadway) and manual-related measures (e.g., hands-on wheel frequency) [20]. Unfortunately, not all these metrics can be directly used to define labels to train machine learning algorithms to predict distracted drivers.

This paper aims to compare approaches to characterize driver distractions. The study relies on a database collected from subjects driving the UTDrive platform – a car equipped with multiple nonintrusive sensors [1, 9]. Following the definition provided by the Australian Road Safety Board [18], this study considers distractions due to the involvement in common secondary tasks that voluntarily or involuntarily deviate the attention from the primary task of driving. We consider operating a phone, operating a *global positioning system* (GPS), operating a built-in car radio, conversation with another passenger, and describing pictures. According to the taxonomy given by Wierwille *et al.* [20], these in-cab tasks include visual-manual (e.g., operating a cellphone), visual only (e.g., following a GPS), and manual primarily (e.g., changing the radio) tasks. Notice that this study does not include distractions or impairments due to alcohol, drugs or fatigue.

The study addresses the problem of describing driver distraction from three complementary approaches. The first approach corresponds to self evaluations collected from post driving questionnaires. After the recording, the drivers completed a questionnaire, in which they rated on a scale of 1 to 5 how distracted they felt while performing the in-cab tasks. The second approach consists in using subjective evaluations to quantify the distraction level perceived by external evaluators. We randomly selected 480 5-sec videos from the database. The raters were asked to annotate these videos by assigning a scale between 1 (less distracted) and 5 (more distracted). The third approach consists in analyzing the differences observed across multimodal features between normal and task conditions. The study includes features derived from the *controller area network* (CAN) Bus data of the vehicle (e.g., steering wheel angle, vehicle speed, brake value, and gas pedal pressures), from a frontal video camera (e.g., head pose and eye closure), and from an array of microphones (e.g., energy). This analysis provides unbiased insights about the differences observed in driving behaviors in the presence of in-cab tasks.

The results indicate that the three approaches give consistent observations. Distractions induced by visual intensive tasks such as tuning the built-in radio, operating the GPS and dialing a number in a cellphone are well captured by the three metrics. However, the metrics do not seem to appropriately characterize cognitive distractions induced by tasks such as talking on a cellphone and following the GPS instructions.

The paper is organized as follows. Section 2 summarizes previous work describing metrics to characterize distracted drivers. Section 3 describes the design and protocol used to record the audiovisual database used in this work. Section 4 describes the results from self evaluations. Section 5 describes the subjective evaluation by external raters and its analysis. Section 6 studies the deviations observed in multimodal features when the driver is engaged in secondary tasks. Section 7 concludes the paper with discussion, future directions and final remarks.

2. RELATED WORK

Several studies have proposed and evaluated measurements to characterize driver distractions. This section summaries some of the proposed metrics.

A common distraction metric is to measure secondary task performance [2]. The drivers are asked to complete artificial detection tasks not related to the primary driving task, such as identifying objects or events, and solving mathematical problems. The performance is measured as the effectiveness (accuracy) and efficiency (required time) to complete the task. There are various approaches that fall under this category. Examples include *peripheral detecting task* (PDT), *visual detection task* (VDT), *tactile detection task* (TDT) and *signal detection task* (SDT) [5, 19]. Most of the studies are conducted using car simulators, in which the stimulus can be controlled.

Studies have proposed surrogate schemes to evaluate the distraction level when the driver operates an in-vehicle technology. These methods are particularly suitable for early stages in the product design cycle of a device that is intended to be used inside the car. The *lane change test* (LCT) is one example [12]. Using a car simulator, the driver is asked to change lanes according to signals on the road while operating a particular device. The distraction level is measured by analyzing the driving performance. Another example is the *visual occlusion* approach, which has been used by automotive human factor experts as a measure of the visual demand of a particular task [6]. In this approach, the field of view is temporally occluded mimicking the eye-off-the-road patterns for visual or visual-manual tasks. During the occlusion interval (usually set equal to 1.5 sec), the subject can manipulate the controls of the device, but cannot see the interface or the control values. The time to complete the task provides an estimation of the required visual demand. However, these metrics are not suitable for our goal of defining ground truth labels to describe the distraction level of recordings collected in real traffic conditions.

Another type of attention measurement corresponds to primary task performance metrics. They determine the attention level of the driver by directly measuring the car response [2]. These measures include *lateral control* such as lane excursions, and steering wheel pattern, *longitudinal control*, such as speed maintenance and break pedal pattern, and *car following performance*, such as the distance to the leading car. Notice that these measurements may only capture distractions produced by visual intense tasks, since studies have shown that metrics such as lane keeping performance are not affected by cognitive load [5]. Lee *et al.* suggested that it is important to study the entire brake response process [11]. In this direction, they considered the *accelerator release time* (i.e., the time between the leading car brakes and the accelerator is released), the *accelerator to brake* (i.e., the movement time from accelerator release to initial brake depress), and the *brake to maximum brake* (i.e., the time from the initial brake depress to maximum deceleration). From these measurements, they found that the accelerator release time was the most sensitive metric of braking performance.

Movement of the eyes usually indicates where the attention is allocated [19]. Therefore, studies have proposed eye glance behavior to characterize inattentive drivers [2]. This is an important aspect since tasks with visual demand require foveal vision, which forces the driver to take the eyes off the road [19]. The proposed metrics range from detailed eye-control metrics, such as within-fixation metrics, saccade profiles, pupil control, and eye closure pattern, to coarse visual behavior metric, such as head movement [19]. The total eye-off-the-road to complete a task is accepted as a measure of visual demand associated to secondary tasks. It is correlated with the number of lane excursions committed during the task [21]. The farther away from the road that a driver fix his/her eyes, the higher the detrimental effect on his/her driver performance [19]. Also, longer glances have higher repercussions than few short glances [21]. In fact, when the eye-off-the-road duration is greater than 2 seconds, the chances of accidents increases [2, 10]. Considering these observations, metrics such as total glance duration, glance frequency and mean single glance duration have been standardized by the *International Organization for Standardization* (ISO). Another interesting metric is the *percent road center* (PRC), which is defined as the percentage of time within 1 minute that the gaze falls in the 8° radius circle centered at the center of the road.

Subjective assessments have been proposed to measure driver distraction. The most common techniques are scales for subjective mental workload. Examples include the *NA-*

SA task load index (NASA-TLX), *driving activity load index* (DALI), *subjective workload assessment technique* (SWAT), *modified Cooper Harper scale* (MCH), and the *rating scale mental effort* (RSME) [22]. For assessment of fatigue, studies have used the *Karolinska sleepiness scale* (KSS) [4].

One important aspect that needs to be defined in many of the aforementioned driver distraction measurements is the corresponding values or thresholds that are considered acceptable for safe driving [22]. In some cases, organizations have defined those values. For example, the *Alliance of Automobile Manufacturers* (AAM) stated that the total duration required to complete a visual-manual task should be less than 20s. In other cases, a secondary task such as manual radio tuning is used as a reference. To be considered as an acceptable, safe task, the deviation in driving performance should be lower than the one induced by the reference task.

This study compares three approaches that can be used to characterize distracted drivers. First, we study self evaluations collected from post driving questionnaires. Second, we consider perceptual evaluations from external evaluators. These two assessments correspond to subjective assessments of driver behavior. Third, we analyze the differences observed across multimodal features from normal patterns. Since we are considering CAN-Bus information and features from the driver extracted from a frontal camera, these metrics belong to primary task performance and eye glance behavior.

3. METHODOLOGY

The goal of this study is to compare different assessment methods that can be used as ground truth to detect inattentive drivers. We are particularly interested in distractions observed when drivers are involved in secondary tasks such as interacting with another passenger, operating a phone, GPS or radio. Although car simulators have been used in previous studies, we are interested in studying realistic scenarios in real driving conditions. Therefore, the recordings will include various uncontrollable but important variables such as congestions, irregular lighting conditions and traffic signals. This section describes the design and recording of the database, including the car platform and the protocol used for the study.

3.1. UTDrive platform

To collect a corpus in real driving conditions, this study relies on the UTDrive car (Fig. 1-a). This is a research platform developed at *The Center for Robust Speech Systems* (CRSS) at *The University of Texas at Dallas* (UTD)[1]. Its goal is to serve as a research platform to develop driver behavior models that can be deployed into human-centric active safety systems. The UTDrive car has been custom fit with data acquisition systems comprising various modalities. It has a frontal facing video camera (PBC-700H), which is mounted on the dashboard facing the driver (see Fig. 1-b). The placement

(a) UTDrive Car (b) Setup

Fig. 1. Car platform used for the recording. (a) picture of the UTDrive car (b) placement of the frontal camera and the microphone array.

Fig. 2. Dewesoft software used for recording and exporting the data. The figure shows the frontal and road videos. It also shows the instantaneous values of various CAN-Bus signals.

and small size of the camera are suitable for recording frontal views of the driver without obstructing his/her field of vision. The resolution of the camera is set to 320×240 pixels and records at 30 fps. Another camera is placed facing the road, which records at 15 fps at 320×240 resolution. The video from this camera can be used for lane tracking. Likewise, the UTDrive car has a microphone array placed on top of the windshield next to the sunlight visors (see Fig. 1-b). The array has five omnidirectional microphones to capture the audio inside the car. We can also extract and record various CAN-Bus signals, including vehicle speed, steering wheel angle, brake value, and acceleration. A sensor is separately placed on the gas pedal to record the gas pedal pressure.

The modalities are simultaneously recorded into a Dewetron computer, which is placed behind the seat of the driver. A Dewesoft software is used to retrieve synchronized information across modalities. Figure 2 shows the interface of the Dewesoft software, which displays the frontal and road videos and various CAN-Bus signals. For further details about the UTDrive car, readers are referred to [1].

Fig. 3. Route used for the collection of the data. The subjects drove this 5.6 miles long route twice. First, they were asked to perform a series of tasks starting with operating the radio and ending with a conversation with a passenger. Then, they drove the route without any in-vehicle distractions.

3.2. Database and Protocol

A multimodal database was recorded for this study, using the UTDrive car. Twenty students or employees of the university were asked to drive while performing a number of common secondary tasks. They were required to be at least 18 years old and have a valid driving license. The average and standard deviation of the age of the participants are 25.4 and 7.03, respectively. The recordings were conducted during dry days with good light condition to reduce the impact of the environment variables. Although wet weather can lead to different challenges for the driver, studies have shown that crashes related to distraction are more likely to occur during dry days with less traffic congestion [8]. By collecting the data during dry days, we have relevant information for the study. The subjects were advised to take their time while performing the tasks to prevent potential accidents.

We defined a 5.6 mile route in the vicinities of the university (see Fig. 3). The route includes traffic signals, heavy and low traffic zones, residential areas and a school zone. We decided to exclude streets with high speed limit (e.g., highways or freeways) from the analysis to minimize the risks in the recording. The participants took between 13 and 17 minutes to complete the route.

The drivers drove this route twice. During the first run, the participants were asked to perform a number of secondary tasks while driving. Among the tasks mentioned by Stutts *et al.* [17] and Glaze and Ellis [7], we selected the following tasks: tuning the built-in car radio, operating and following a GPS, dialing and using a cellphone, describing pictures and interacting with a passenger. Some dangerous tasks such as text messaging, grooming and eating were not included for security reasons. The details of the selected seven tasks are given below.

Radio: The driver is asked to tune the built-in car radio to some predetermined stations. The radio is in its standard place, on the right side of the driver.
GPS - Operating: A predefined address is given to the driver who is asked to enter the information in the GPS. The device is mounted in the middle of the windshield. The driver is allowed to adjust it before starting the recording.
GPS - Following: After entering the address in the GPS, the driver is asked to follow the instructions to the destination.
Phone - Operating: The driver dials the number of an airlines automatic flight information system (toll-free). A regular cellphone is used for this task. Handsfree cellphones are not used to include the inherent mechanical distraction.
Phone - Talking: After dialing the number, the driver has to retrieve the flight information between two given US cities.
Pictures: The driver has to look and describe randomly selected pictures, which are displayed by another passenger sitting in the front passenger seat. The purpose of this task is to collect representative samples of distractions induced when the driver is looking at billboards, sign boards, shops or any object inside or outside the car.
Conversation: A passenger in the car asks general questions to establish a spontaneous conversation.

According to the driver resources-based taxonomy defined by Wierwille *et al.* [20], the selected secondary tasks includes visual-manual tasks (e.g., *GPS - Operating* and *Phone - Operating*), visual only tasks (e.g., *GPS - Following* and *Pictures*) and manual primarily task (e.g., *Radio*). The set also includes tasks characterized by cognitive demand (e.g., *Phone - Talking*) and auditory/verbal demands (e.g., *Conversation*).

During the second run, the drivers were asked to drive the same route without performing any of the aforementioned tasks. This data is collected as a normal reference to compare the deviation observed in the driver behaviors when he/she is engaged in secondary tasks. Since the same route is used to compare normal and task conditions, the analysis is less dependent on the selected road.

4. SELF EVALUATION OF DISTRACTION

The first approach considers self assessment from the drivers. The assumption is that drivers are somehow aware of the visual, cognitive, auditory, psychological and physical distractions induced by common secondary tasks. At the very least, they should be able to rank-order the tasks that are more distracting to them. While this assumption may not always hold, this approach provides a subjective value from the drivers to characterize the secondary tasks that are most distracting to them.

The drivers answered a questionnaire after completing the recording. They were asked whether they felt comfortable driving the car and whether their driving performance was representative of their normal driving behavior. Then, they were

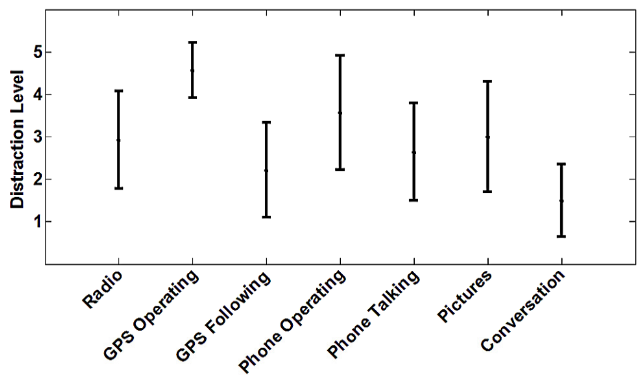

Fig. 4. Average distraction levels based on self evaluations across the drivers. The figure shows the mean and standard deviation of the values assigned to each task.

asked to rate how distracted they felt while performing each of the secondary task considered in this study. They used a Likert-scale with extreme values corresponding to 1 – *less distracted*, and 5 – *more distracted*.

Figure 4 presents the error plots with the average and standard deviation values for the secondary tasks. The drivers felt that *GPS - Operating* was the most distracting task, followed by *Phone - Operating*, *Pictures* and *Radio*. *Conversation* was felt as the least distracting task. This result agrees with the findings of Strayer *et al.*, which concluded that conversation does not increase crash risk [16].

Notice that this approach only provides a coarse value for each task. This average value should be assigned as a distraction label to each of the recording collected when the drivers were performing the corresponding task. However, certain subtask within a task may be more distracting (e.g., looking in the car for the cellphone). This approach ignores this intrinsic within-task variability. An alternative approach is to individually assess small segments of the recording by conducting subjective evaluations. This approach is explored in Section 5.

5. SUBJECTIVE EVALUATION OF DISTRACTION

The second approach consists in conducting subjective experiments to assess the perceived distraction level of the drivers. Instead of measuring coarse labels for the secondary tasks, the evaluation is conducted over localized segments of the recording. The underlying assumption is that the previous driving experience of the external evaluators will allow them to accurately identify and rank distracting scenarios or actions.

One advantage of this approach is that a number quantifying the perceived distraction level is assigned to localized segments in the recording. Therefore, it is possible to identify various multimodal features that correlate with this distraction metric. Using these features, regression models can be designed to directly identify inattentive drivers. Another advantage is that the videos can be assessed by many raters so the aggregated values are more accurate.

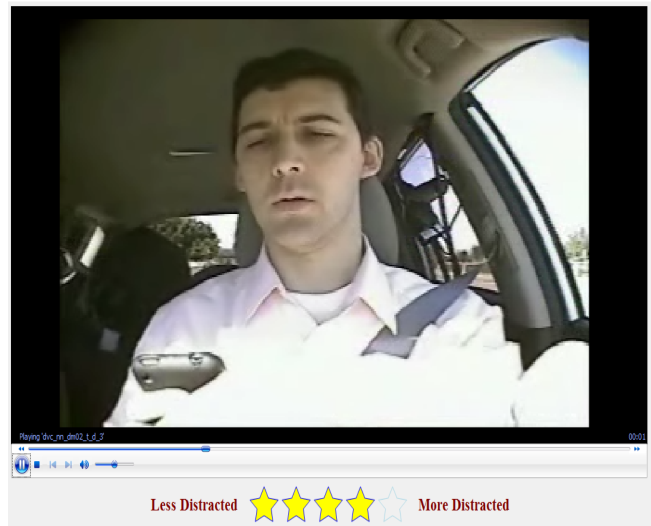

Fig. 5. GUI for the subjective evaluation of distraction. The subjects are required to rate the perceived distraction level of the driver (1 – *less distracted*, to 5 – *more distracted*).

The database contains approximately seven hours of data. However, only a portion of the corpus was considered for the evaluation. The corpus was split into 5 sec videos with their corresponding synchronized audio. For each driver, three videos were randomly selected for each of the seven secondary tasks (Sec. 3.2). Three videos from normal condition were also randomly selected. Therefore, 24 videos per driver are considered, which give altogether 480 unique videos (3 videos $\times$ 8 conditions $\times$ 20 drivers = 480). Twelve students at UTD were recruited to participate in the evaluation. A *graphical user interface* (GUI) was built for the subjective test (see Fig. 5). To unify the understanding of the raters, the definition of distraction adopted in this paper (see Sec. 1) was read before the evaluation. After watching a video, the evaluators rated the perceived distraction level on a scale of 1 – *less distracted*, to 5 – *more distracted*. To minimize the duration of the evaluation, each evaluator was requested to assess only 160 videos. The average duration of the evaluation was approximately 15 minutes. To avoid biases, the presentation of the videos was randomized for each evaluator. In total, three independent evaluators assessed each video. The correlation between the assessments provided by the evaluators is in average $\rho = 0.63$, which suggests that the assigned values are consistent across raters.

Figure 6 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 visually intensive tasks such as *Radio*, *GPS - Operating*, *Phone - Operating* and *Pictures* are perceived by the evaluators as the most distracting tasks.

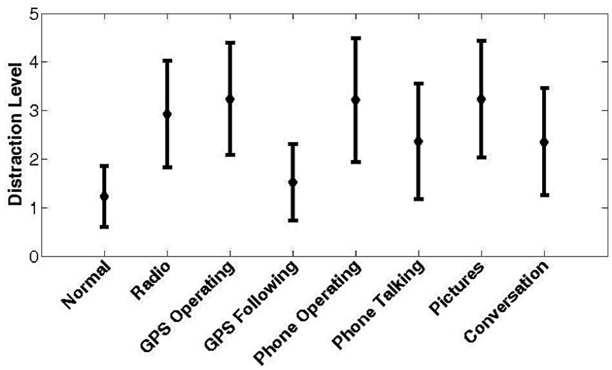

Fig. 6. Average perceived distraction levels based on subjective evaluations. The figure shows the mean and standard deviation of the values assigned to each task by external raters.

The tasks that involve cognitive load such as *GPS - Following* are perceived as the least distracting tasks. This result is also observed with *Phone - Talking*, which is perceived as less distracting than *Phone - Operating*. Notice that studies have shown that the mechanical distraction in operating a cellphone is a minor factor compared to verbal and cognitive distraction [2]. These results suggest that the proposed subjective evaluation is effective in capturing visual distraction but not cognitive distraction.

Figures 4 and 6 show similar patterns, which validates the use of external evaluators. The main differences are observed in *GPS - Operating* and *Conversation*. On the one hand, *GPS - Operating* is felt by the drivers as a highly distracting task ($\mu = 4.57$ for self evaluations, $\mu = 3.24$ for external evaluations). This result suggests that certain demands on the drivers are not easily captured by subjective evaluations. On the other hand, the average value assigned to *Conversation* is higher in the subjective evaluations by the external raters ($\mu = 1.49$ for self evaluations, $\mu = 2.25$ for external evaluations). Drivers felt that having a spontaneous conversation did not affect their driving performance.

6. DEVIATIONS IN MULTIMODAL FEATURES

The third approach consists in analyzing the deviations observed in multimodal features when the drivers are engaged on secondary tasks. We expect to identify modalities that can be used to characterize driver distractions by measuring the differences in the features extracted during task and normal conditions (i.e., without performing any task). This approach provides unbiased metrics to describe driver distractions.

6.1. Modalities

In this study, we extract features from the *controller area network* (CAN)-Bus data, which provide direct information about the car performance during the recording. We also consider features automatically extracted from the frontal video camera, which describe the eye glance behavior of the driver. Finally, we extract acoustic information from the microphone array. The raw information from all these modalities are extracted using the Dewesoft software.

Car Performance Features: As mentioned in Section 3.2, we have access to important CAN-Bus information such as the steering wheel angle, vehicle speed, brake pedal pressure, and gas pedal pressure. We extract car performance features derived from these signals. From the steering wheel angle, we calculated its jitter by computing the variance over 5 sec windows centered at each frame. This feature reveals small corrections or drifts made by the drivers. From the brake and gas pedal pressures, we estimate their derivatives, which describe their relative temporal changes. The car speed is directly used as a feature, since studies have shown that the vehicle speed is reduced when drivers are engaged in secondary tasks [5].

Eye Glance Behavior Features: The frontal video camera provides valuable information about the eye glance behavior of the driver. Features describing the head pose and eye closure rate are automatically extracted using the AFECT software [3]. The toolkit is robust against large data sets and varied illumination conditions. AFECT processes frame-by-frame the video, avoiding error propagation due to tracking and localization. We estimate the yaw and pitch movements of the head which are used as features. The head roll movement is not considered since we expect that it is not directly affected by secondary tasks. The analysis also considers the percentage of eye closure, which is defined as the percentage of frames in which the eyeballs are directed below a given threshold. This threshold is set at the point where the eyes are looking straight at the frontal camera.

Acoustic features: The audio in the car is another modality that we expect to be relevant in describing distracted drivers. Several of the secondary tasks considered in this work have characteristic acoustic signals (e.g., *GPS - Following*, *Phone - Talking*, *Pictures* and *Conversation*). In this study, the energy of the audio signal is extracted and used as a feature.

The AFECT software gives reliable information when the head rotation is within $\pm 10°$ range. If the face is occluded or rotated beyond this range the software fails to provide information. Therefore, we consider only segments in which the car moved straight (absolute value of the steering wheel angle is smaller than $20°$). Therefore, we avoid segments that require the driver to glance. Any remaining gap in the data due to face rotation or hand obstruction is interpolated. Likewise, we do not consider segments in which the vehicle speed is lower than 5 kph, since we are only interested in the driver behaviors observed when the car is moving.

6.2. Analysis of Multimodal Features

The analysis aims to identify prominent modalities to characterize driver distraction. The proposed approach is to measure

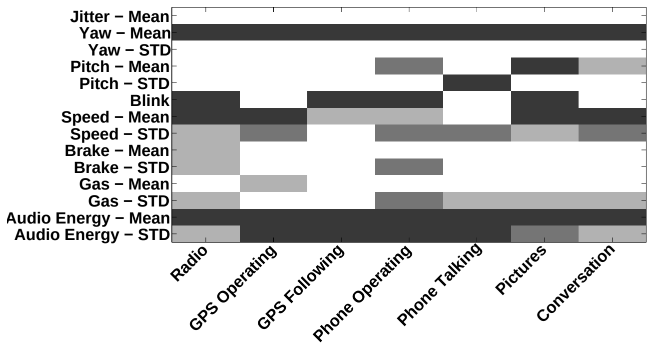

Fig. 7. Results of the Matched Pairs t-test: Features vs Tasks. For a particular task, gray regions indicate the features that are found to have significant differences (dark gray, p-value=0.05; gray, p-value=0.10; light gray p-value=0.20).

the differences observed in the features during task and normal conditions. However, the route segments have different properties (e.g., speed limits). To reduce route variability, we only compare the recordings for task and normal conditions over the same route segment.

Similar to Section 5, we split the corpus into 5 sec windows for the analysis. In each window, we estimate the mean and standard deviation of the modalities presented in Section 6.1, which are used as features for the analysis. A matched pairs hypothesis test is conducted to determine whether the differences in the features between task and normal conditions are significant across drivers. Since the database contains 20 drivers, a t-test for small sample size is used. The results are given in Figure 7, which highlights the features that were found significant at p-value=0.05 (dark gray), p-value=0.10 (gray) and p-value=0.20 (light gray).

Figure 7 shows that the mean of head yaw movement presents significant differences across each of the seven secondary tasks. Drivers tend to look to their right when they are engaged in secondary tasks. This result agrees with previous studies that indicate that drivers have to take the eye off the road to complete secondary task [19]. The figure also indicates that the mean and standard deviation of the audio energy are significantly different across tasks. It will be interesting to study whether these patterns are also observed in more naturalistic recordings, or when the drivers are engaged in other secondary tasks. The figure suggests that the mean speed of the vehicle is an important feature. A detailed analysis indicated that drivers tend to reduce the car speed when they are performing a secondary task [9]. This result agrees with the study of Engström *et al.*, which suggests that visual load correlated with a reduce of speed [5]. They explained this pattern as a compensatory effect to keep an acceptable driving performance. Other salient features are the percentage of eye closure (i.e., blink), and the mean of head pitch movements.

Secondary tasks that impose visual demands such as *Radio*, *GPS - Operating*, *Phone - Operating* and *Pictures* present many salient multimodal features (see Fig. 7). However, we did not find many features that are affected by tasks that increase the cognitive load of the driver (e.g., *GPS - Following* and *Phone - Talking*). This result indicates that the selected features may not capture cognitive distractions.

The salient multimodal features identified in this section can be used to characterize distracted drivers. As suggested by Young *et al.*, there are many approaches that can be used to define performance criteria [22]. We can define a common task as a reference (e.g., *Radio*). A recording from the database will be labeled as ‘distracted’ if the deviations observed in its features exceed the deviations observed in this reference task. Alternatively, a threshold can be defined for each of the proposed features. If the differences in the driver behaviors exceed these thresholds, the recording will be labeled as ‘distracted.’

7. DISCUSSION AND CONCLUSIONS

This study presented three methodologies to describe the distraction level of the drivers. The ultimate goal is to define labels that can be used to train human-centric active safety systems. The first approach was based on self assessments. It aims to identify and rank-order distracted tasks as perceived by the drivers themselves. The second approach was to subjectively evaluate localized recordings of the drivers by conducting perceptual evaluations by external raters. The experiments revealed substantial inter-evaluator agreement, which validates the approach. The third approach was to identify salient features across different modalities (CAN-Bus signals, eye glance behavior, and acoustic signal). This method gives unbiased metrics to describe the deviation in behaviors observed when the driver is involved in the secondary tasks.

Consistent results are observed across approaches. Distraction induced by visual intensive tasks are well captured by the proposed metrics. However, these measurements do not seem to appropriately describe cognitive distractions.

An interesting result is that none of the proposed approaches was able to characterize *Phone - Talking* as a distracting task. This result differs from previous studies that show the detrimental effect in driving performance produced by talking on a cellphone [15, 16]. Under this condition, drivers seems to have inattention blindness or selective withdrawal of attention [15]. They fail to see an object even though they are looking directly at it. Unfortunately, this type of distraction is hard to determine with the proposed metrics. Victor *et al.* suggests that cognitive load reduces the horizontal and vertical variability of gaze direction [19]. It produces longer on-road fixation and reduces glance frequency to mirror and speedometer. Although these patterns were not observed, finding new detailed features to capture these behaviors are planned as part of our future work. Likewise, we will complement the proposed subjective evaluations with mental workload scales to investigate whether these instruments are able to capture cognitive distractions.

After defining relevant metrics to describe distracted drivers, our next goal is to use these labels to build machine learning algorithms to detect inattentive drivers. The intended driver behavior monitoring system will provide feedbacks to inattentive drivers, preventing accidents, and increasing the security on the roads.

References

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

[2] K. M. Bach, M.G. Jaeger, M.B. Skov, and N.G. 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.

[3] M. Bartlett, G. Littlewort, T. Wu, and J. Movellan. Computer expression recognition toolbox. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, Amsterdam, The Netherlands, September 2008.

[4] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, 12(2):596–614, June 2011.

[5] 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*, 8(2):97 – 120, March 2005.

[6] J.P. Foley. Now you see it, now you dont: Visual occlusion as a surrogate distraction measurement technique. In M.A. Regan, J.D. Lee, and K.L. Young, editors, *Driver distraction: theory, effects, and mitigation*, pages 123–134. CRC Press, Boca Raton, FL, USA, October 2008.

[7] A. L. Glaze and J.M. Ellis. Pilot study of distracted drivers. Technical report, Transportation and Safety Training Center, Virginia Commonwealth University, Richmond, VA, USA, January 2003.

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

[9] 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.

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

[11] J.D. Lee, D.V. McGehee, T.L. Brown, and M.L. Reyes. Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44:314–334, Summer 2002.

[12] S. Mattes and A. Hallén. Surrogate distraction measurement techniques: The lane change test. In M.A. Regan, J.D. Lee, and K.L. Young, editors, *Driver distraction: theory, effects, and mitigation*, pages 107–122. CRC Press, Boca Raton, FL, USA, October 2008.

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

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

[15] D. L. Strayer, J. M. Cooper, and F. A. Drews. What do drivers fail to see when conversing on a cell phone? In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, volume 48, New Orleans, LA, USA, September 2004.

[16] D.L. Strayer, J.M. Watson, and F.A. Drews. Cognitive distraction while multitasking in the automobile. In B.H. Ross, editor, *The Psychology of Learning and Motivation*, volume 54, pages 29–58. Academic Press, Burlington, MA, USA, February 2011.

[17] J. C. Stutts, D. W. Reinfurt, L. Staplin, and E. A. Rodgman. The role of driver distraction in traffic crashes. Technical report, AAA Foundation for Traffic Safety, Washington, DC, USA, May 2001.

[18] I. Trezise, E.G. 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. Technical Report No. 209 Session 2003-2006, Road Safety Committee, Parliament of Victoria, Melbourne, Victoria, Australia, August 2006.

[19] T.W. Victor, J. Engström, and J.L. Harbluk. Distraction assessment methods based on visual behavior and event detection. In M.A. Regan, J.D. Lee, and K.L. Young, editors, *Driver distraction: theory, effects, and mitigation*, pages 135–165. CRC Press, Boca Raton, FL, USA, October 2008.

[20] W. Wierwille, L. Tijerina, S. Kiger, T. Rockwell, E. Lauber, and A. Bittner Jr. Final report supplement – task 4: Review of workload and related research. Technical Report DOT HS 808 467, U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, USA, October 1996.

[21] 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.

[22] K.L. Young, M.A. Regan, and J.D. Lee. Measuring the effects of driver distraction: Direct driving performance methods and measures. In M.A. Regan, J.D. Lee, and K.L. Young, editors, *Driver distraction: theory, effects, and mitigation*, pages 85–105. CRC Press, Boca Raton, FL, USA, October 2008.