ANALYSIS OF FACIAL FEATURES OF DRIVERS UNDER COGNITIVE AND VISUAL DISTRACTIONS

Nanxiang Li and Carlos Busso

Multimodal Signal Processing (MSP) Laboratory

Department of Electrical Engineering, The University of Texas at Dallas, Richardson TX-75080, USA

nxl056000@utdallas.edu, busso@utdallas.edu

ABSTRACT

Drivers are exposed to a growing risk of being distracted with the recent development of in-vehicle systems for navigation, communication and infotainment. As a result, there is a need for tracking systems that can monitor the drivers' attention. This study investigates driver distractions using a multimodal corpus collected from real world driving scenarios. The paper focuses on facial cues automatically extracted from a frontal camera facing the driver. We conducted subjective evaluations by external observers to assess the perceived visual and cognitive distraction of drivers performing secondary tasks. The data is divided into two classes – *distracted* and *normal*. This partition is separately created for visual and cognitive scores. Binary classifiers are built with features describing *action units* (AU) and gaze (e.g., head poses). The classifiers achieve 80.8% F-score for visual distractions, and 73.8% F-score for cognitive distractions. The study identifies features that are relevant for detecting both types of distractions. Furthermore, the paper presents a logistic regression analysis to identify facial features that are useful for detecting samples in which cognitive distraction scores are not related to visual distraction scores. The analysis reveals the benefits of using AU in cognitive related distraction detection.

***Index Terms*—** Driver distraction, facial features, action units, cognitive distraction, visual distraction

1. INTRODUCTION

Driver distraction detection is an important yet challenging task, given the various potential distraction sources. Studies have suggested the multidimensional nature of drivers' distraction, including visual, cognitive, auditory, psychological and physical distractions [1]. A key challenge is that each distraction's type affects the driver in different ways, so it is important to study them separately. Among the types of distractions, visual and cognitive aspects are prominent causes of vehicle accidents [2]. Visual distractions affect the drivers' visual attention, reducing their road awareness. Cognitive distractions are related mind-off-the-road situations, where drivers are immersed in internal thoughts such as daydreaming and thinking. These distractions directly impair the drivers' ability for visual scanning and decision making [3,4].

This paper explores the use of facial features to detect drivers' visual and cognitive distractions. The motivation of this study is rooted in the cognitive appraisal theory, which states that human emotions are generated as the result of perceiving, assessing and responding to events [5]. In the context of driving behaviors, the desires and actions of performing secondary tasks can affect the drivers' emotions (e.g., nervous, surprise, happy). The driving scenario can also induce cognitive states such as stress and confusion, which will also affect drivers' facial expressions. Based on the well-established relationship between facial expression and human emotion and cognitive states [6], we hypothesize that facial expression can be used to capture cues signaling distracted behaviors. This hypothesis is particularly important for cognitive distractions, for which visible features to assess cognitive load are less obvious. The proposed features can play a key role in the next generation of in-vehicle active safety systems.

Although emotion has being considered in driver distraction studies [7], few research groups have considered facial expression. A common approach is to explore direct measurements derived from the driving activity including lateral (lane position) and longitudinal (speed) control measures [8, 9]. Some studies focused on the performance of secondary tasks such as event detection and driver response [10, 11]. Other measurements used in this area are *electroencephalography* (EEG), size of eye pupils and eye movement [1, 9, 10]. Facial features can provide complementary information to these measurements.

This study uses perceptual evaluations from external evaluators to assess the perceived visual and cognitive distraction levels of drivers engaged in secondary tasks. The corpus consists of recording from subjects driving a car in normal public roads. The scores provided by the evaluators are used to split the recordings into *normal* and *distracted* classes. This partition is separately done for visual and cognitive distractions. Using binary classification, we explored facial features, including head orientation and *action units* (AUs), which are automatically extracted from video recordings. The binary classifiers achieve F-score of 80.8% and 73.8% for visual and cognitive distractions, respectively. Gaze and AU features are useful for detecting visual distractions, while AU features are particularly important for cognitive distractions. A scatter plot of the perceptual evaluation in the visual-cognitive space shows that in many cases the cognitive and visual distraction scores are correlated. However, some recordings have higher cognitive distraction scores than visual distraction scores. This result is consistent with the fact that visual demanding tasks induce cognitive load, but not vice versa. Based on this observation, we used logistic regression analysis to identify facial features that are useful for detecting the cases

Table 1. Seven secondary tasks considered in this study. The tasks phone and GPS were split, since we observed different driver behaviors while operating and using the devices [12, 13].

Task Name	Description
Radio	Driver tunes the radio to predetermined stations
GPS-Operating	Driver inputs a predetermined address into GPS
GPS-Following	Driver follows the GPS instruction to the destination
Phone-Operating	Driver dials an airline automatic flight information system using a cellphone
Phone-Talking	Driver interacts with the flight information system to retreive flight information
Picture	Driver describes pictures shown by the passenger to reflect the distraction caused by road sighs and billboard
Conversation	Driver discusses the driving experience with the passengers and answers general questions to the passenger

where cognitive distraction is not related to visual distraction. Most of the features that are statistically significant ($p = 0.05$) under chi-square statistical test are AU features. This result suggests that AUs are important for cognitive distraction detection.

2. DATA COLLECTION

2.1. Database

The multimodal database used for this study includes real world driving scenarios in which 20 participants were asked to drive a customized 2006 Toyota RAV4. The participants' age ranges from 20 to 40 with an average of 25.4. Each participant drove the car along the same predefined 5.6-mile route twice. During the first run, the drivers were asked to perform various common secondary tasks during specific segments in the route (Table 1). During the second run, they were asked to drive normally following the same route. The driving data is recorded using both a frontal camera facing the driver (30 fps) and a road camera facing the road (15 fps). Both cameras provide the same resolution (320x240). The corpus also includes the *controller area network* (CAN) Bus signal and audio recorded with a microphone array. This study uses only the video cameras. Table 1 describes the selected secondary tasks, which are commonly performed by individuals while driving. Each task was performed under the same route segment. This controlled, sequential approach reduces the road dependency in the analysis and induces similar learning effect across drivers. The data was collected during the day under good weather condition. Each participant took about 40 minutes to finish the experiment, producing over 12 hours of driving recordings. The corpus is further described in Jain et al. [12].

2.2. Perceptual Evaluation

This study relies on human perceptual evaluation to assess the perceived visual and cognitive distraction levels induced by different secondary tasks. Three non-overlapped 10-second video segments were randomly selected for each participant under 8 different driving conditions (7 tasks and normal). Altogether, 480 videos (3 videos $\times$ 8 conditions $\times$ 20 drivers = 480) were extracted from the database. Eighteen subjects were invited to evaluate both the perceived visual and cognitive distraction levels of the drivers. We unified the understanding of visual and cognitive distractions by providing a precise definition. We followed the description given by Ranney et al. [3]. Visual distraction is defined as eye-off-the-road – drivers looking away from the roadway. Cognitive distraction is defined as mind-off-the-road – drivers being lost/busy in thoughts. We also mentioned cues that they could use for inferring both types of distractions (e.g., gaze and head direction for visual distractions; facial expression, secondary task and driving performance for cognitive distractions). The evaluation included videos showing the road and the driver. By showing both videos, the evaluators can easily determine whether the observed behaviors were related to the primary driving task, increasing the reliability of the perceived distraction scores. The evaluators gave a continuous score to indicate the perceived distraction level by adjusting a sliding bar, for which its extreme values are 0-*less distracted* and 1-*more distracted*. To minimize the duration of the evaluation, each evaluator is only asked to assess 80 videos for visual distraction, followed by a different set of 80 videos for cognitive distraction. The visual and cognitive distraction levels of the driver in each video are assessed by 3 different evaluators. This study uses the average across evaluators to label the drivers' distraction levels.

In our previous study, we showed the consistency among different evaluators [14]. We also compared the scores provided by external observers with other two commonly used distraction measurements: gaze metrics and self evaluations. It was found that perceptual evaluation by external observers is superior: (1) it provides reliable scores for short videos based on the driving behaviors and route conditions, (2) it incorporates the unbiased perception of various subtle cues that a driver may choose to ignore, (3) it gives consistent scores from many evaluators.

3. FEATURE EXTRACTION

3.1. Facial Action Unit

Previous works on detecting driver behavior suggested that facial features provide valuable information about drivers performing secondary tasks [12, 15]. In addition to the obvious facial movement associated with secondary tasks such as talking, we hypothesize that facial expression can play an important role in cognitive distraction detection. In particular, features related to brow motion and eye lids movements may signal cognitive load. The cognitive appraisal theory indicates that human emotion results from the evaluation of events, including the available coping strategies [5]. Since face is a prominent expressive communicative channel, this theory implies that there exist a link between cognitive load and facial expressions. In fact, human behavioral studies have used facial expression to analyze the affective state, cognitive processes, and social interaction of an individual [16].

One of the most popular frameworks for describing facial expression is the *facial action coding system* (FACS) [17]. The *action units* (AUs) defined in the FACS represent the muscle activity that generates different facial expressions. This system has been used for distinguishing between genuine and simulated pain [18], and detecting deceptive behaviors [19]. Generally, FACS is coded by professionally trained individuals that are

Table 2. Visual features for the analysis. Low level features are time series signals over which we estimate statistics over windowed segments (i.e., high level feature).

Low Level Feature		
Action Unit (AUs)		
Inner Brow Raiser (AU1)	Dimpler (AU14)	Lip Tightener (AU23)
Outer Brow Raiser (AU2)	Lip Corner Depressor (AU15)	Lip Pressor (AU24)
Brow Lowerer (AU4)	Chin Raiser (AU17)	Lips part (AU25)
Upper Lid Raiser (AU5)	Lip Stretcher (AU20)	Jaw Drop (AU26)
Nose Wrinkler (AU9)	Cheek Raiser (AU6)	Lip Suck (AU28)
Upper Lip Raiser (AU10)	Lid Tightener (AU7)	Blink (AU45)
Lip Corner Puller (AU12)	Lip Puckerer (AU18)	
Gaze Related Features		
Head Yaw (Yaw)	Head Pitch (Pitch)	Head Roll (Roll)
High Level Features (10 sec.)		
Statistics		
Mean	Minimum (Min)	Skewness
Standard Deviation (STD)	Range	Kurtosis
Maximum (Max)	Inter-Quatile Range (IQR)	
Global features		
Longest Eyes-Off-Road Duration (LEOR Dur.)		
Eyes-Off-Road Duration (EOR Dur.)		

FACS certified. The computer vision community has made significant progress to automatize the extraction of AUs. In particular, Bartlett et al. [20] developed algorithms that were included in the *Computer Expression Recognition Toolbox* (CERT), a robust toolbox for estimating facial features. The software estimates 20 AUs, from frontal faces (see Table 2). In fact, this toolbox has been used for drowsy driver detection [21]. This study estimates these 20 AUs for each of the 10-sec videos. Then, the eight statistics listed in Table 2 are calculated for each of the 20 AUs (e.g., skewness of AU25). Altogether, a feature vector with 160 *high level features* (HLFs) is generated per video.

3.2. Gaze Related Features

Gaze metrics provide useful information about people's attention, and has been used for detecting driver distraction [9, 10]. In this study, we extracted features related to head pose to approximate the drivers' gaze behavior. The assumption is that most of time people move their heads for visual scanning, especially when the glance range is big. Low level features correspond to the 3-D head orientations represented by yaw, pitch and roll angles, which are extracted by CERT. Details about the head pose estimation can be found in Whitehill et al. [22].

For each of the three head pose angles, eight statistics (Table 2) are calculated, generating 24 HLFs. In addition, we also derived two off-the-road gaze behavior features: total *eye-off-the-road* (EOR) duration, and *longest eye-off-the-road* (LEOR) duration. Both metrics have been used in other driver distraction studies [9, 10]. EOR is related to the time required to complete a task, and LEOR captures the duration of each glance (longer glances are more dangerous than few short glances). Both metrics are estimated by defining a reference field centered at the road. In this study, we used a rectangle relevant field by setting thresholds on yaw and pitch angles. The drivers are regarded as looking off-the-road when their estimated head pose

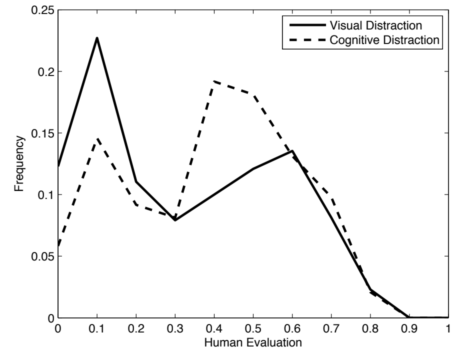


Fig. 1. Histograms of perceived visual and cognitive distractions.

is beyond these thresholds. Considering that participants vary in their heights and seat positions, thresholds are individually defined with respect to their normal driving condition. Using these thresholds, the feature EOR is calculated by counting the frames in the recordings in which the driver is detected with his/her eyes off-the-road. The feature LEOR is calculated by counting the longest consecutive eye-off-the-road frames. Notice that normal driving behavior such as mirror checking can be detected as an eye-off-the-road action, thus simply using these metrics for detecting visual distraction can be misleading.

In total, 186 HLFs are estimated for all the videos used in the perceptual evaluation. The features include 26 gaze related features and 160 AUs related features. Due to hand occlusions, extreme glance behaviors or adverse illumination conditions, CERT may fail to recognize the face in the frame producing empty values for the LLFs. In these cases, we interpolated the missing value when the number of frames with missing values is less than one third of the total number of frames. Otherwise, the samples are excluded from the analysis (83 out of 480). The study considers 397 samples. Notice that the interpolation introduces a delay that can affect real-time applications.

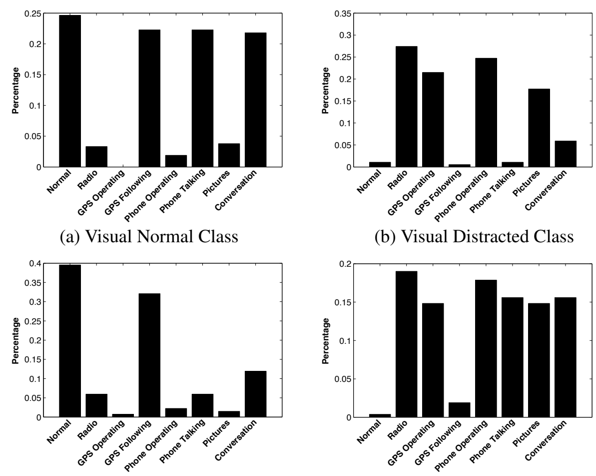
4. DETECTION OF DISTRACTED BEHAVIORS

4.1. Data Partition

To explore the effectiveness of facial features in the drivers' distraction detection, we implemented binary classification to recognize the distracted and normal driving behaviors. The data is split into two classes labeled as either *distracted* or *normal* by setting thresholds on the human perceptual evaluations. This distracted-normal dichotomy is separately done for the perceived visual and cognitive distraction scores. The thresholds are determined based on the distributions of the scores as shown in Fig. 1. The figure shows that both types of distractions cluster into two groups. The threshold s=0.3 is used for both visual and cognitive distractions to define the classes *normal* (low distraction scores) and *distracted* (high distraction scores) (visual: 211-normal, 186-distracted; cognitive: 134-normal, 263-distracted). Fig. 2 shows the distribution of the recordings for each class across the eight driving conditions (7 tasks and normal). For visual distraction, the *distracted* class consists of data from *Radio*, *GPS-Operating*, *Phone-operating*, and *Picture*. For cognitive

Table 3. Binary classification for normal and distracted behaviors (Feat. # = feature dimension; P = precision; R = recall; F = F-score).

	Visual Distraction											
	Gaze Feature				**AUs Feature**				**All Feature**			
	Feat #	**P (%)**	**R (%)**	**F (%)**	**Feat #**	**P (%)**	**R (%)**	**F (%)**	**Feat #**	**P (%)**	**R (%)**	**F (%)**
LDC	6	71.9	71.3	71.6	3	77.3	76.3	76.8	4	81.0	80.6	80.8
KNN	12	71.8	71.5	71.6	4	76.6	75.5	76.0	5	78.7	77.9	78.3
SVM1	4	72.0	71.3	71.6	4	77.2	76.3	76.8	4	80.6	80.4	80.5
SVM2	6	71.9	70.9	71.4	4	76.3	75.3	75.8	4	79.5	79.0	79.3
QDC	5	71.4	70.4	70.9	3	76.8	74.5	75.6	4	80.9	79.2	80.0
	Cognitive Distraction											
	Gaze Feature				**AUs Feature**				**All Feature**			
	Feat #	**P (%)**	**R (%)**	**F (%)**	**Feat #**	**P (%)**	**R (%)**	**F (%)**	**Feat #**	**P (%)**	**R (%)**	**F (%)**
LDC	4	71.7	68.9	70.3	8	74.3	72.4	73.3	24	73.8	73.4	73.6
KNN	10	70.6	71.1	70.8	10	71.8	67.6	69.6	29	67.6	68.1	67.8
SVM1	15	72.4	70.8	71.6	11	70.0	68.5	69.2	21	73.8	73.9	73.8
SVM2	8	68.7	69.4	69.1	8	73.9	69.3	71.5	10	73.2	72.4	72.8
QDC	5	67.3	69.1	68.2	8	70.4	71.6	71.0	10	70.9	72.3	71.6



(a) Visual Normal Class (b) Visual Distracted Class

(c) Cognitive Normal Class (d) Cognitive Distracted Class

Fig. 2. Distributions of *normal* and *distracted* classes for visual and cognitive distractions. The y-axis provides the percentage of tasks in each class.

distraction, the distracted class includes recordings from most of the secondary tasks except *GPS-following*. It is important for the *distracted* class to include various secondary tasks such that the results can capture representative features across tasks.

4.2. Binary Classification

Various machine learning methods are considered in the binary classification experiment: *linear discriminant classifier* (LDC), *k-nearest neighbor classifier* (KNN), *support vector machine* with linear kernel (SVM1); *support vector machine* with quadratic kernel (SVM2) and *quadratic discriminant classifier* (QDC). The performance of the classifer is estimated with 20-fold driver independent crossvalidation scheme. In each fold, data from 19 drivers is used for training and the data from the remaining driver is used for testing. To understand what features are important for driver distraction, for each training set, we applied *forward feature selection* (FFS) to the 186 features based on inter-intra class distance ratio. The top 30 features with the highest inter-intra class ratio are selected and used for the evaluation. This criterion does not depend on any particular classifier. While a wrapper-based method can give higher accuracies, this approach was selected since the focus of the analysis is on the features rather than on the classifiers. Different number of features from this set are tested for each classifier, and the one with the best performance is reported in Table 3.

Due to the unbalanced data partition, the table reports the average precision (P), average recall (R) and F-score (F). The average precision corresponds to the mean precision (i.e., fraction of retrieved samples that are relevant) of *normal* and *distracted* classes. The average recall corresponds to the mean recall (i.e., fraction of relevant samples that are correctly classified) between the classes. F-score combines both metrics by using equation 1.

$$F = 2 \times \frac{precision \times recall}{precision + recall} \tag{1}$$

The results show that we achieve the best performance when both gaze and AUs features are considered with F-score of 80.8% (LDC) for visual distraction and 73.8% (SVM1) for cognitive distraction. Using only AUs gives better performance than using only gaze features for both visual and cognitive classifications. Interestingly, the performance for visual binary classification improves when fusing all the features. Glance behavior features and AUs provide complementary information that can be used to detect visual distractions. For cognitive distraction, combining both set of features does not improve the performance. Glance behavior features do not add additional cues to the ones given by AUs. The performance of the classifiers highlights the importance of using facial features to recognize cognitive distractions.

The rank of the best features provided by FFS varies across the 20 folds. We study the top five features across folds for cognitive and visual distractions. For cognitive distraction, the top five features in order are head Yaw Mean; Lip Corner Depressor (AU15) STD; Lip Puckerer (AU18) Max; Lip Tightener (AU23) IQR; head Roll Mean. For visual distraction, the top five features in order are Lip Tightener (AU23) IQR; Jaw Drop (AU26) Max; Head Yaw Mean; Lip Suck (AU28) Mean; Blink (AU45) Min. The features Lip Tightener (AU23) IQR and head Yaw Mean are frequently selected for both visual and cognitive distractions. For cognitive distraction, Roll Mean was previously found useful for

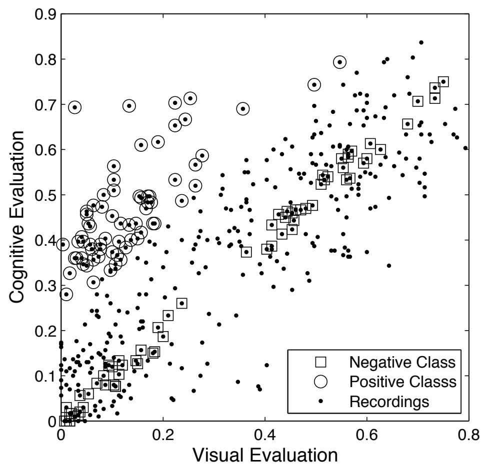


Fig. 3. Positive and negative classes. The study aims to identify relevant features characterizing samples in which cognitive distraction scores are not related to visual distraction scores.

phone talking detection since people tend to tilt their heads when using the phone [13]]. Given that phone talking is a cognitive exclusive task, this result is expected. For visual distraction, Blink (AU45) Min is related to eye movement. Therefore, it can provide useful information about visual attention.

5. VISUAL & COGNITIVE DISTRACTIONS

Fig. 3 displays the scatter plot of the evaluation scores for visual and cognitive distractions. The figure shows most recordings have similar scores for both type of distractions (points in the diagonal). However, there are many samples where the cognitive distraction score is higher than the visual distraction score (e.g., upper left corner). Furthermore, The close but still different performance between the visual and cognitive classification suggests that these two type of distractions are not completely correlated. This result is expected since visual distraction can induce cognitive distraction but not vice versa. This section presents a logistic regression analysis to identify facial and gaze features that are useful for detecting samples in which cognitive distraction scores are not related to visual distraction scores. For this purpose, the recordings are divided into two new classes based on the difference between the scores assigned to cognitive and visual distractions. The positive class corresponds to the cases when cognitive distraction is not related to visual distraction (i.e., higher cognitive scores). The negative class corresponds to the cases when the two distractions are related (i.e., similar scores). Thresholds are applied to define the two classes. For the positive class, we set an upper threshold equal to the mean plus one standard deviation of the differences between cognitive and visual distraction scores. Samples for which the score difference is higher than this threshold are assigned to the positive class (68 recordings). For the negative classes, we defined a region centered in the diagonal of the scatter plot such that the number of samples matches the ones in the positive class (balanced classes). Fig. 3 highlights the samples that are selected for each of the classes. Most of the samples in the positive class are related to cognitive tasks (*Phone-Talking* and *Conversation*).

The feature analysis follows the approach introduced by Busso et al. [23], which is based on logistic regression (Eq. 2).

$$\pi(f) = \frac{e^{\beta_0+\beta_1 f_1+\beta_2 f_2+\cdots+\beta_F f_F+\varepsilon}}{e^{\beta_0+\beta_1 f_1+\beta_2 f_2+\cdots+\beta_F f_F+\varepsilon}+1} \tag{2}$$

In logistic regression, the contribution of a set of features can be statistically estimated by comparing two nested models. If the likelihood ratio between the models is multiplied by minus two times its natural logarithm, the resulting statistic is approximately distributed with a chi-square function with 1 degree of freedom. We compare the model with just the intercept (Equation 3) with the model with a single feature (Equation 4). Then, we estimate the chi-square statistic for each feature and use its value to rank-order them (Fig. 4). Features that are more discriminative between the positive and negative classes will lead to better models, which will be reflected in the chi-square statistic.

$$H_0 \quad : \quad \pi(f) = \frac{e^{\beta_0}}{e^{\beta_0}+1} \tag{3}$$

$$H_1 \quad : \quad \pi(f) = \frac{e^{\beta_0+\beta_1 f_1}}{e^{\beta_0+\beta_1 f_1}+1} \tag{4}$$

Fig. 4 shows the top 30 features ranked by the logistic regression analysis. These features are all significant (p-value = 0.05) under chi-square distribution test (see horizontal line). Most of the features are AUs except for two high level features related to head Roll. It is not surprising that many AUs related to lips movement are selected given that positive class are closely related to talking activities. In addition, AUs describing the eyebrows and cheeks are selected among the best features. They capture emotional and cognitive cues displayed by the drivers. These features can provide useful information to detect tasks that increase the cognitive load, without affecting the visual demand.

6. CONCLUSIONS AND FUTURE WORK

The results from this study indicate that facial information is useful for driver distraction detection. Gaze features and AUs provide valuable information for visual distraction detection. The results indicate that AUs play an important role in cognitive distraction detection. In addition, AUs are also useful for detecting when cognitive distraction is not induced by visual distraction.

Since the cognitive tasks considered in this study closely related to talking activities, our future work will include the analysis of other cognitive tasks (e.g. thinking, solving math problems). In addition, other modalities such as CAN-Bus, and audio signals, which can provide complementary information, will be studied. We will also collect data under different driving conditions (e.g., day/night, road/highway). Studies have shown that driver behaviors change under different driving conditions. For example, driving on city roads is different from driving on highways. By collecting more data, we will expand our study to cover a wide range of scenarios, building appropriate models for driver distraction under different road and environment conditions.

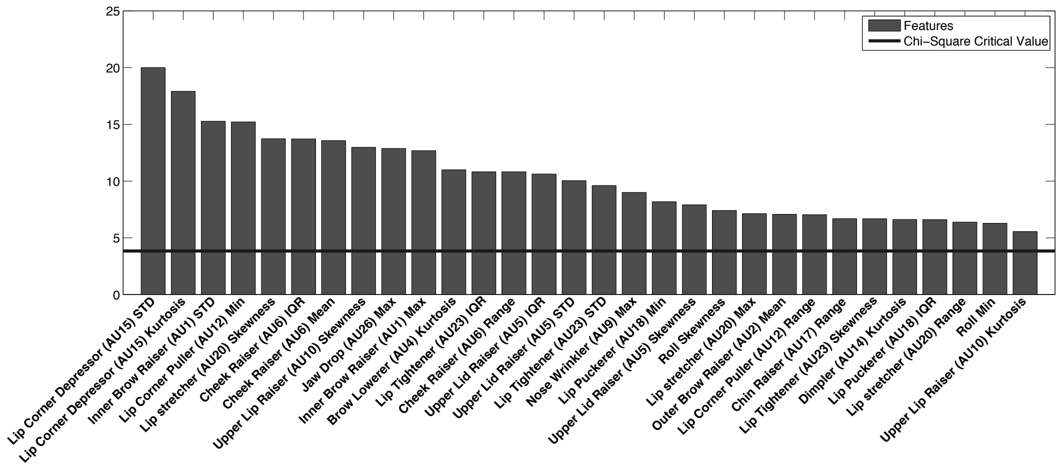


Fig. 4. Most discriminative features for positive and negative classes according to the logistic regression analysis. The horizontal line indicates the threshold for which the individual features are statistical significant at p-value=0.05.

7. REFERENCES

[1] 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.

[2] Y. Liang, M.L. Reyes, and J.D. 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.

[3] 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.

[4] 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*, New Orleans, LA, USA, September 2004, vol. 48.

[5] K.R. Scherer, "Appraisal theory," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Power, Eds., pp. 637–663. John Wiley & Sons, Ltd, Chichester, UK, January 2005.

[6] P. Ekman and E.L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)*, Oxford University Press, New York, NY, USA, 1997.

[7] C. Gordon, "Driver distraction: An initial examination of the 'attention diverted by' contributory factor codes from crash reports and focus group research on perceived risks," *Zuletzt geprüft am*, vol. 10, pp. 2011, 2005.

[8] J.L. Harbluk, Y.I. Noy, P.L. 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.

[9] J.C. McCall and M.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.

[10] 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.

[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] 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.

[13] N. Li, J.J. Jain, and C. Busso, "Modeling of driver behavior in real world scenarios using multiple noninvasive sensors," *IEEE Transactions on Multimedia*, vol. In press, 2013.

[14] N. Li and C. Busso, "Using perceptual evaluation to quantify cognitive and visual driver distractions," in *Smart Mobile In-Vehicle Systems – Next Generation Advancements*, H. Abut, J. H. L. Hansen, G. Schmidt, and K. Takeda, Eds. Springer, New York, NY, USA.

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

[16] C.L. Lisetti and D.J. Schiano, "Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect," *Pragmatics and Cognition*, vol. 8, no. 1, pp. 185–235, 2000.

[17] P. Ekman and W.V. Friesen, *Facial Action Coding System: A Technique for Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[18] K.D. Craig, S.A. Hyde, and C.J. Patrick, "Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain," *Pain*, vol. 46, no. 2, pp. 161–171, August 1991.

[19] M.G. Frank and P. Ekman, "Appearing truthful generalizes across different deception situations," *Journal of personality and social psychology*, vol. 86, no. 3, pp. 486–495, March 2004.

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

[21] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M.Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," *Human–Computer Interaction*, pp. 6–18, 2007.

[22] J. Whitehill and J.R. 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.

[23] C. Busso, S. Lee, and S.S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 582–596, May 2009.