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

Motivation

- 100-car Naturalistic Study: Over 78% of crashes involved driver inattention
- It is estimated that drivers engage in potentially distracting secondary tasks about 30% of their time [Ranney, 2008]
- In-vehicle technologies, cell phones and navigation systems are estimated to increase exponentially [Broy, 2006]

Types of Distraction

- **VISUAL** – eyes looking somewhere beside the road
- **COGNITIVE** – driver thinking about something besides driving
- **AUDITORY** – driver speaking over phone or with co-passenger
- **PSYCHOLOGICAL/ PHYSICAL** – driver emotions, physical limitation

Are They Distracted???

Driver's facial and head movement can tell us something!!!

Highlights of this study

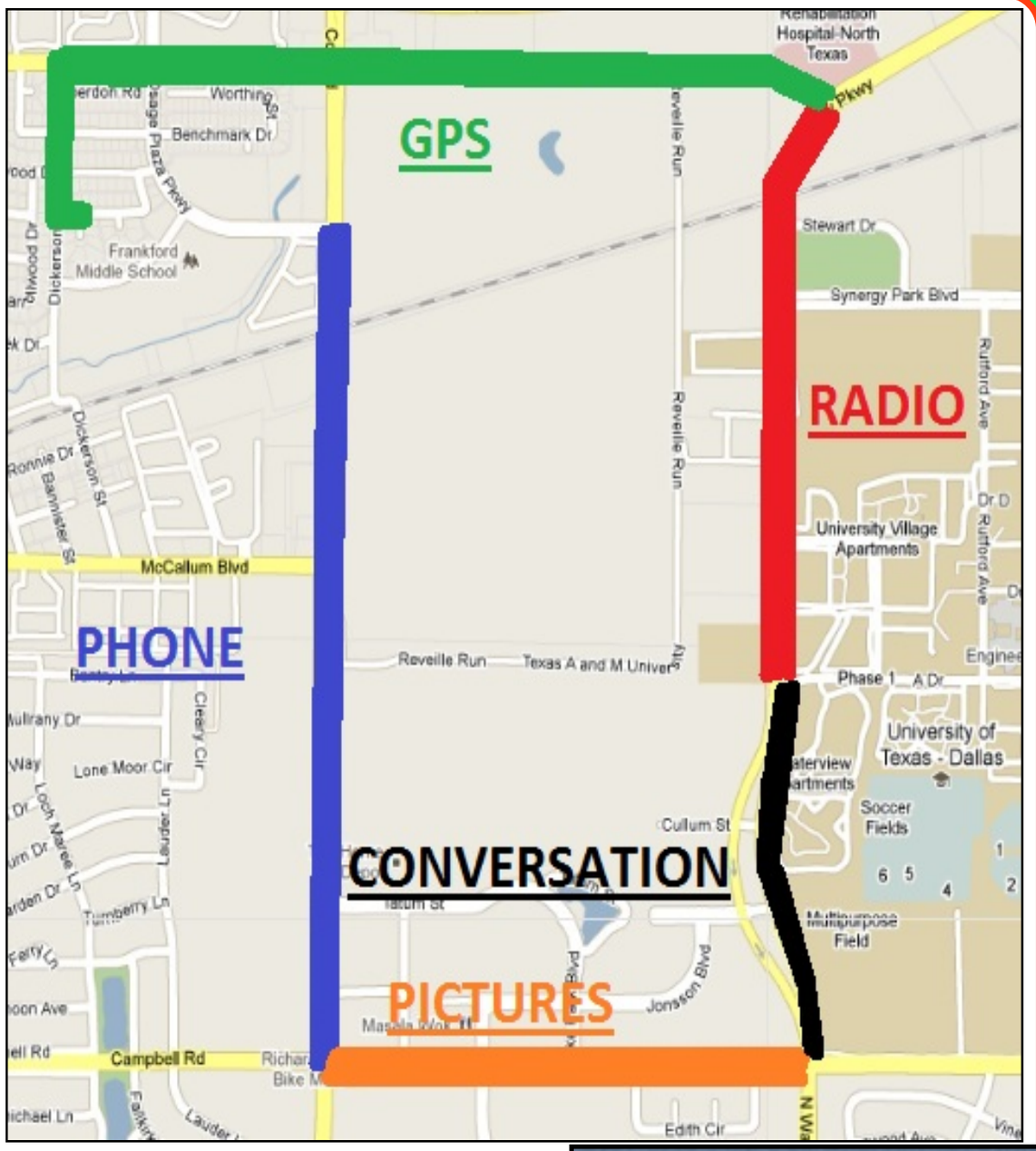
- **Detection of driver visual and cognitive distraction based on facial information**
- **Rely on human perceptive evaluation to annotate visual and cognitive distraction levels**
- **Exploration of the relationship between head/ facial movement and driver distraction**

UTDrive

- **Front facing camera**
  - PBC-700
  - 320 x 240 at 30fps
- 4 - channel Microphone array
  - 25kHz
- CAN Bus for Steering wheel, Vehicle speed, Brake, Gas
- **Road facing camera**
  - 320 x 240 at 15fps

Protocol

- 20 drivers: 10 male, 10 female
  - Valid US Driving License
  - At least 18 years of age
- Good Day light, dry weather
- 2 runs of driving per subject
- First run – with 7 tasks
- Second run – neutral driving (without tasks)

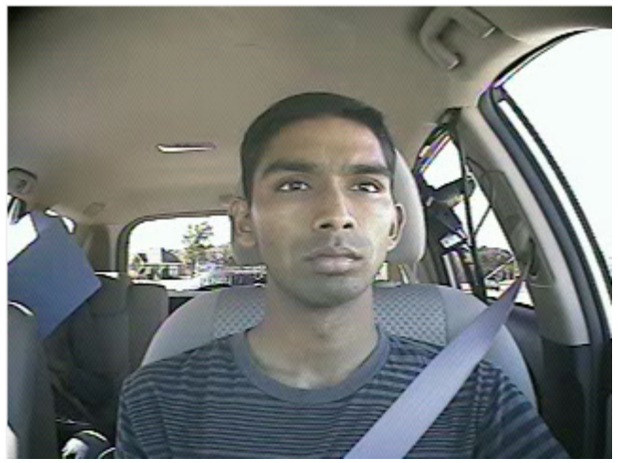


Secondary tasks

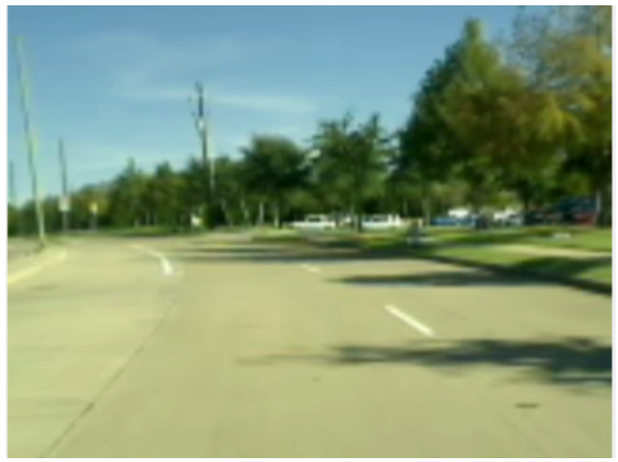
- Radio
- GPS - Operating
- GPS - Following
- Phone - Operating
- Phone - Talking
- Pictures
- Conversation

Preprocessing

- 10-second driver videos and its corresponding road video are randomly chosen from the database (480 videos)
  - 3 samples x 8 tasks x 20 drivers = 480
- The speed of the UTDrive vehicle is greater than 0km/h in the chosen videos



Driver Video



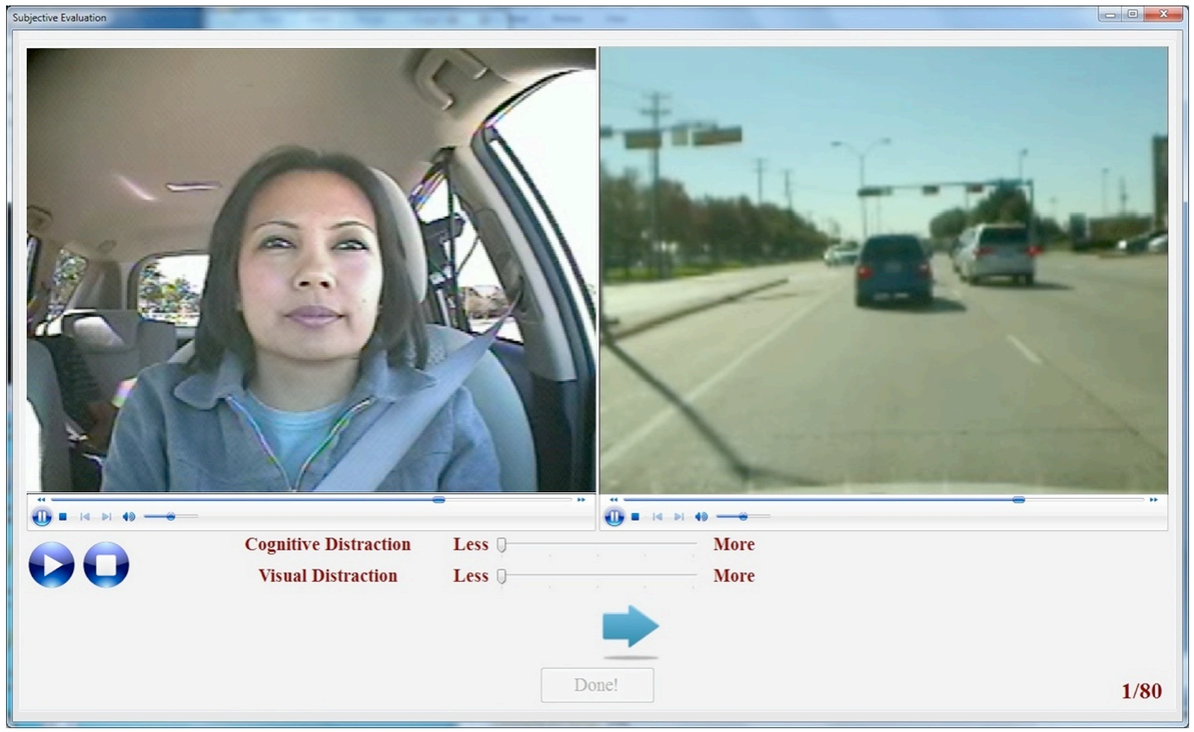
Road Video

Perceptual Evaluation

- We separately evaluate the perceived visual and cognitive distractions

- Evaluators watch both road and driver videos

- Each video is evaluated by 3 different observers and the average is used

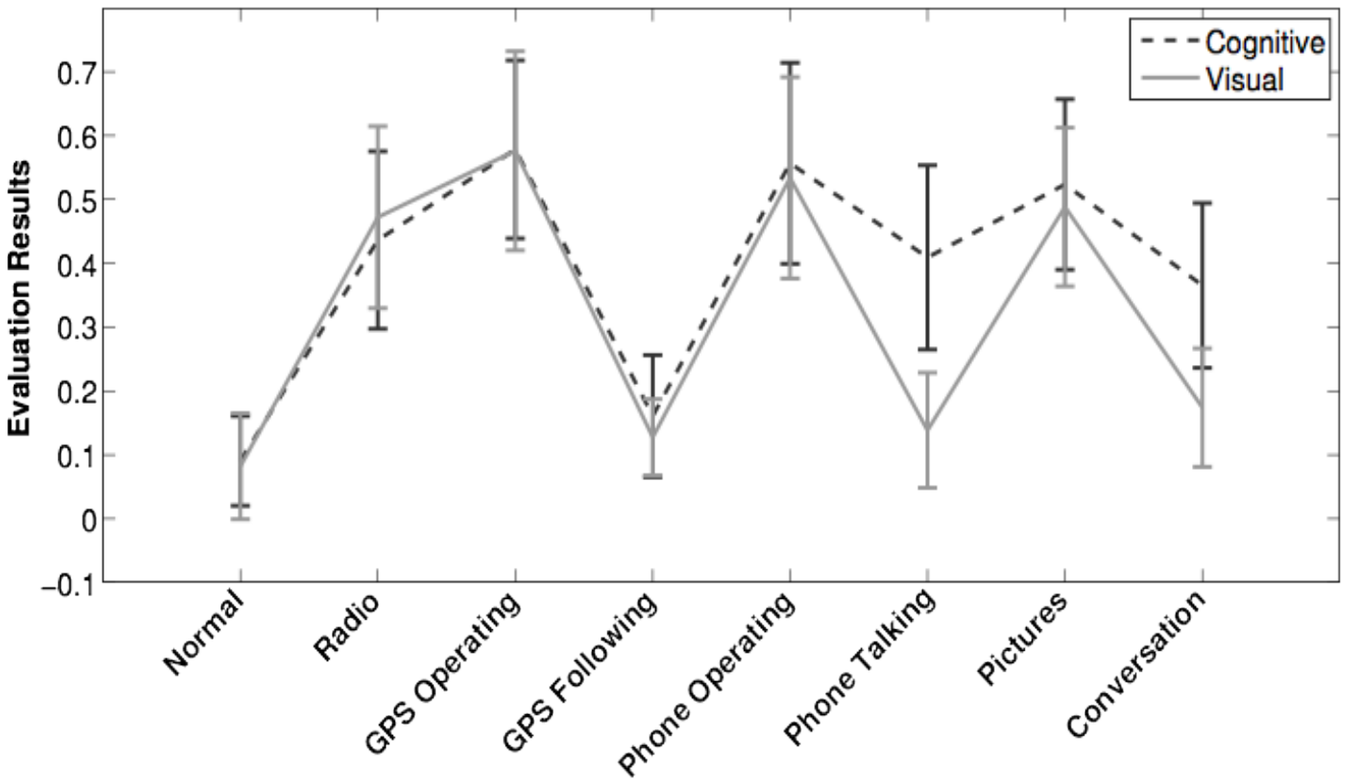


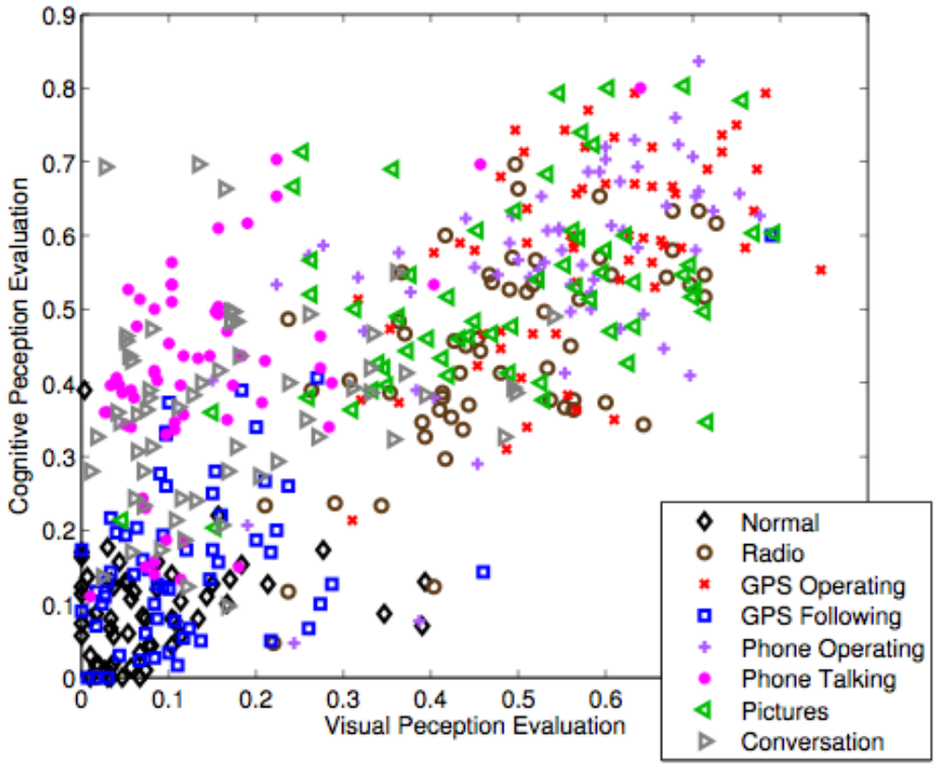
GUI for evaluation

✓ Advantages

- Labels assigned to localized segments
- Videos can be assessed by many raters

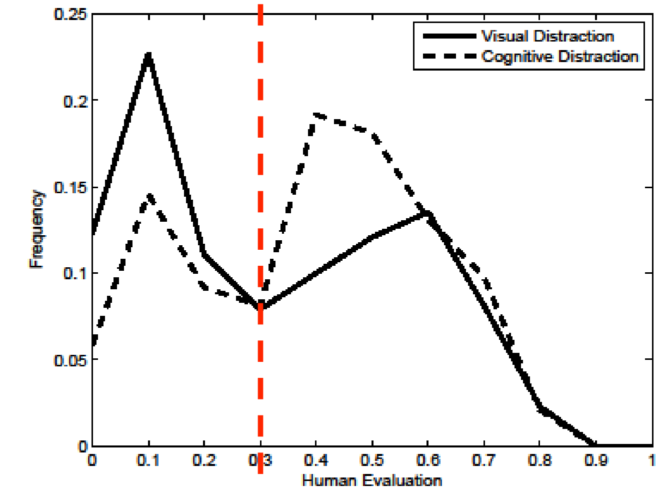
Perceived Visual and Cognitive Distractions





Mean values for perceived cognitive and visual distractions

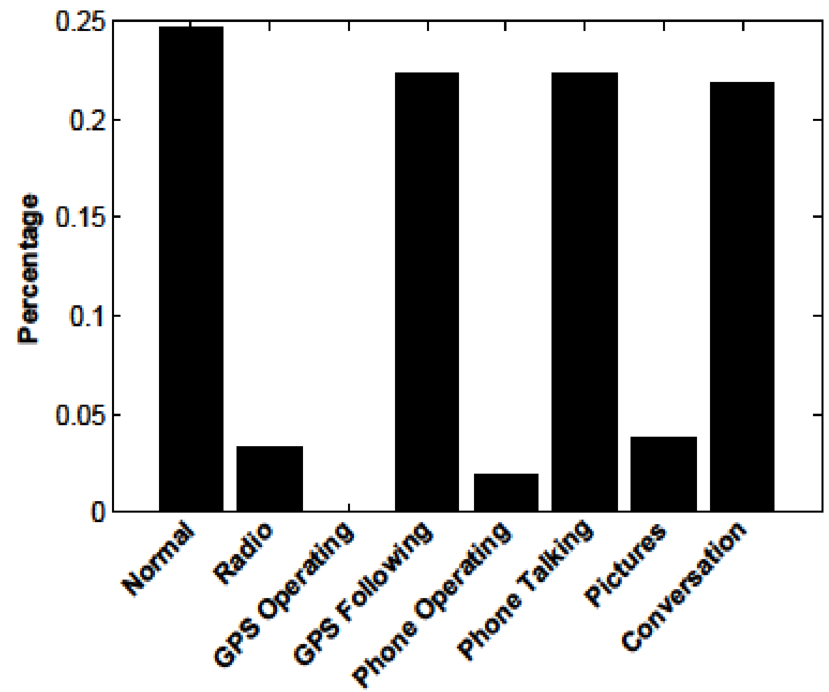
Binary Classes

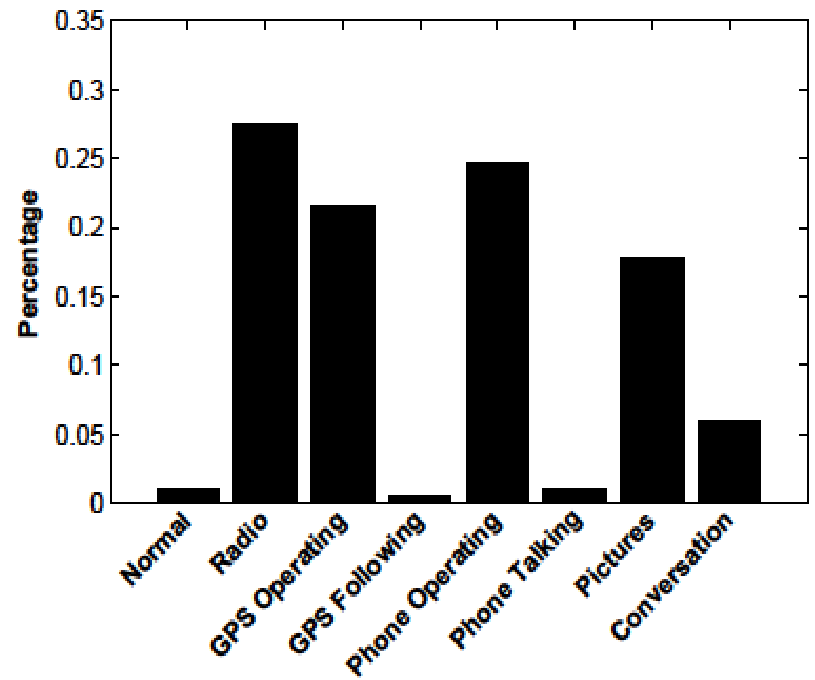


Undistracted Distracted

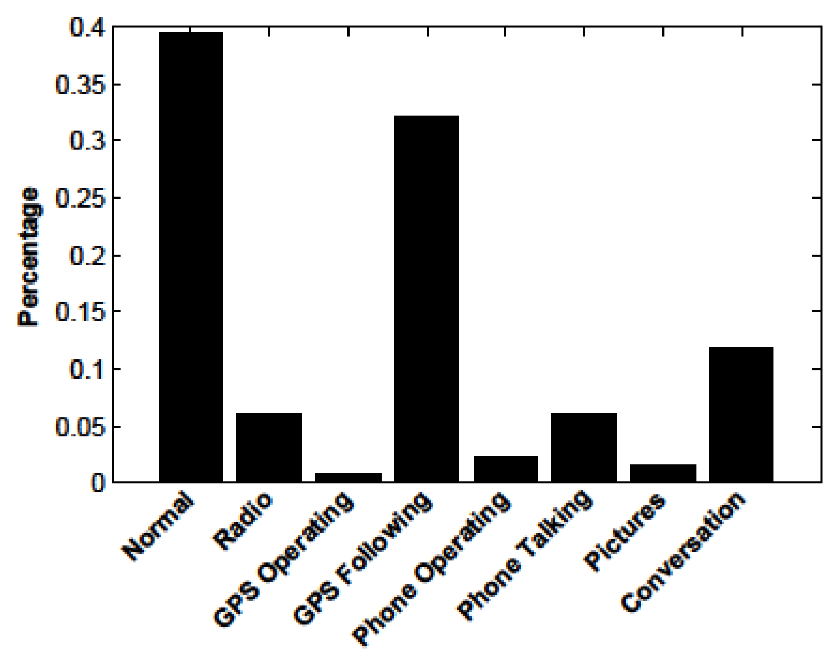
Undistracted Distracted

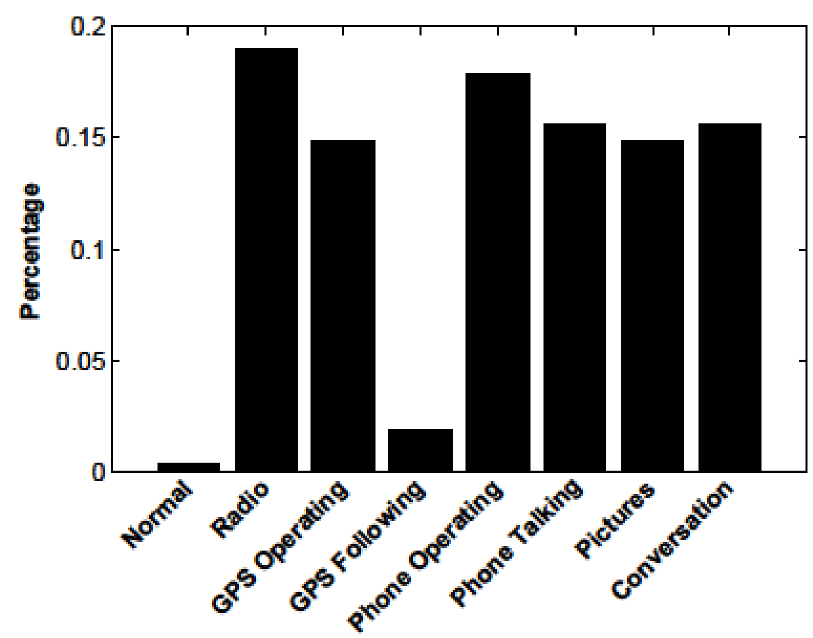
Visual



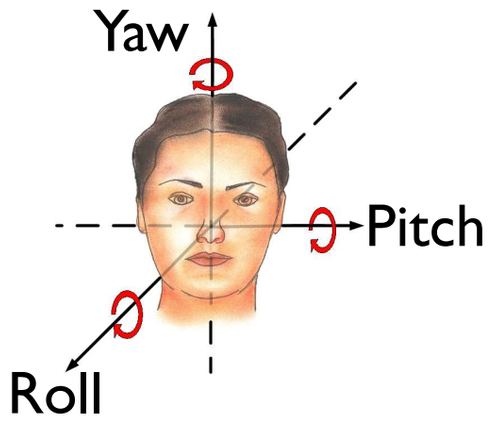


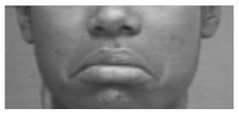
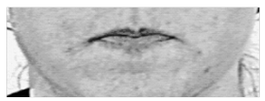
Cognitive





Head/Facial Features



AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer
AU5  Upper Lid Raiser	AU6  Cheek Raiser	AU7  Lid Tightener
AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU43  Eyes Closed
AU12  Lip Corner Puller	AU15  Lip Corner Depressor	AU17  Chin Raiser
AU23  Lip Tightener	AU24  Lip Pressor	AU26  Jaw Drop

Source: http://www.cs.cmu.edu/~face/facs.htm

- Frontal Facing video Information:
  - Head pose (yaw, pitch and roll)
  - Action Units
    - High level eye features
- Extracted with the Computer Expression Recognition Toolbox (CERT)

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

Feature Extraction

- Low level features
  - CERT AUs
  - CERT head pose
- High level features
  - Statistics
  - LEOR and EOR
- 186 in total

Low Level Feature		
Action Unit		
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)	
Head Related Features		
Head Yaw (Yaw)	Head Pitch (Pitch)	Head Roll (Roll)
High Level Features		
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.)		

LEOR and EOR

- Studies have shown that when the eyes-off-the-road (EOR) duration is greater than 2 seconds, the chances of accidents increase.
  - Total duration of glance (EOR Dur.)
  - Longest glance (LEOR Duration)
- A driver dependent box is set
  - EOR is detected when head orientation is out of the box

Binary Classification Results

(20- fold driver independent crossvalidation)

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	80.6	80.8
KNN	12	71.8	71.5	71.6	4	76.6	75.5	76	5	78.7	77.9	78.3
SVM1	4	72	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	79.3
QDC	5	71.4	70.4	70.9	3	76.8	74.5	75.6	4	80.9	79.2	80

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	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	10	70.9	72.3	71.6

LDC - linear discriminant classifier, **KNN** - k-nearest neighbor classifier, **SVM1** - support vector machine with linear kernel, **SVM2** - support vector machine with quadratic kernel, **QDC** - quadratic discriminant classifier

Precision, Recall and F-score

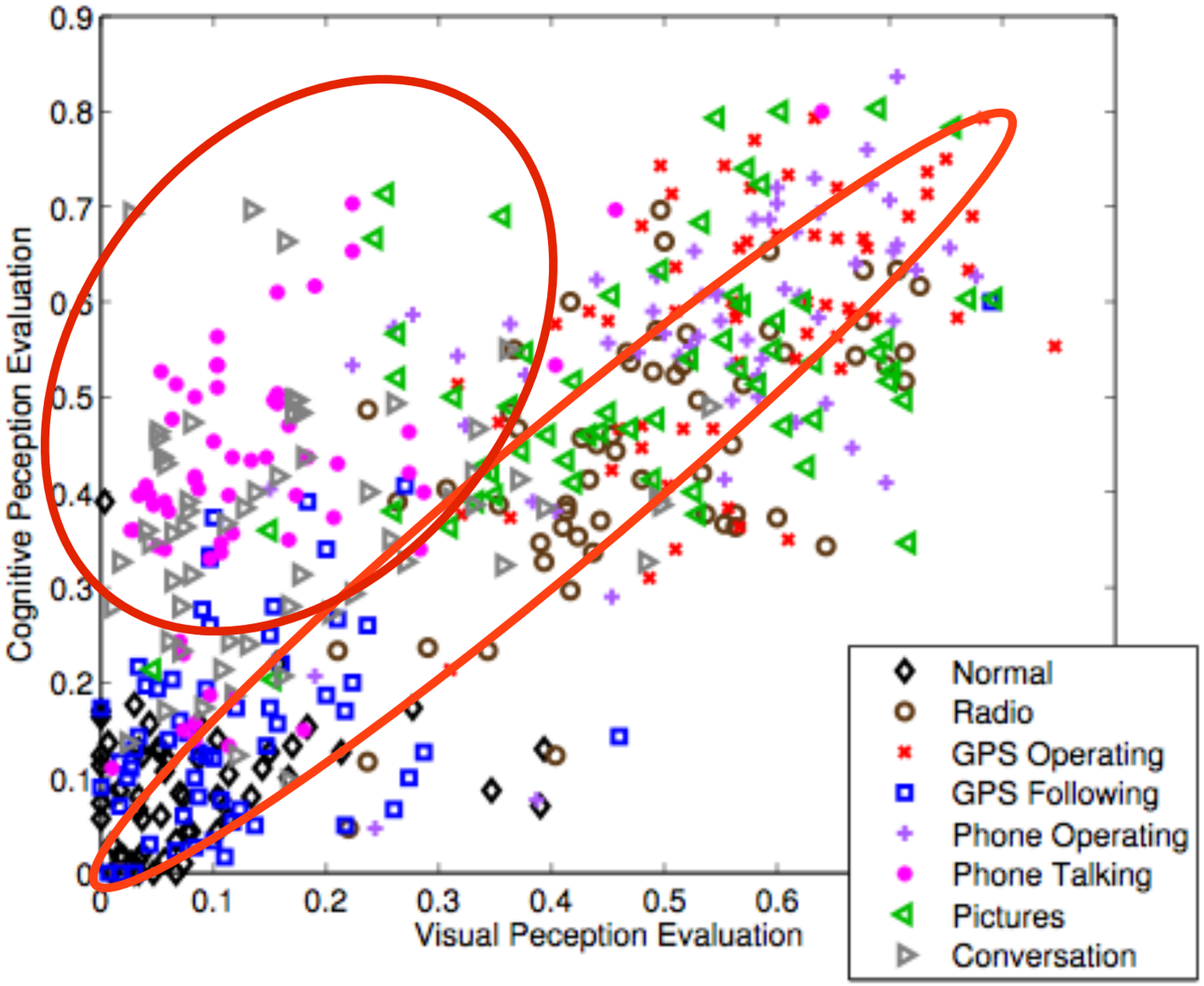
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F = 2 * \frac{Precision * Recall}{Precision + Recall}$$

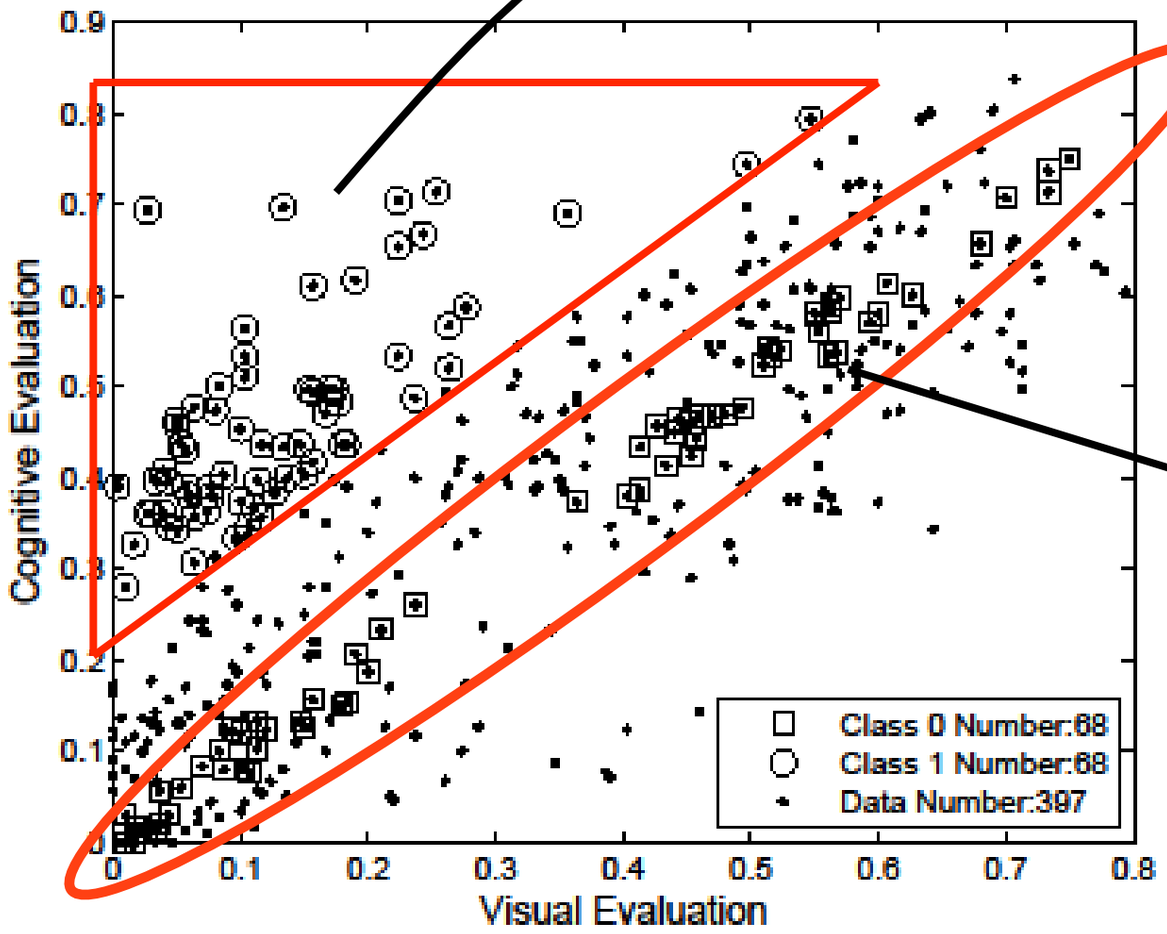
	Actual Class	
Predicted Class	TP (true positive)	FP (false positive)
	FN (false negative)	TN (true negative)

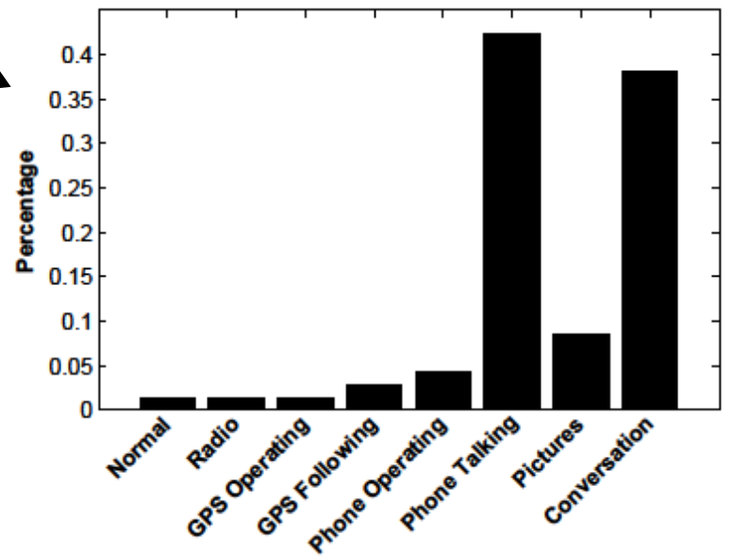
Perceived Visual and Cognitive Distractions Scatter Plot

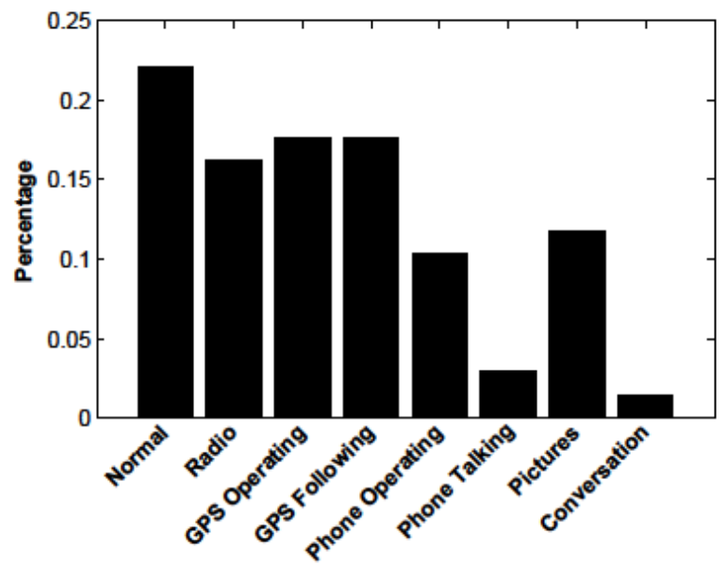


- Visual and cognitive scores are correlated
  - Visual distractions induces cognitive distractions
- There are videos with high cognitive scores and low visual scores

A Different Binary Class Problem







- Data are split into two new classes
  - Class 1 – visual distraction ≈ cognitive distraction
  - Class 2 – cognitive distraction > visual distraction

Logistic Regression Analysis

- In logistic regression, the contribution of a set of features can be statistically estimated by comparing two nested models.

H_0: $\pi(f) = \frac{e^{\beta_0}}{e^{\beta_0} + 1}$ - model with just the intercept

H_1: $\pi(f) = \frac{e^{\beta_0} + e^{\beta_1 f_1}}{e^{\beta_0} + e^{\beta_1 f_1} + 1}$ - model with a single feature

The likelihood ratio between the models is related to chi-square
Goal: Compare each feature at a time

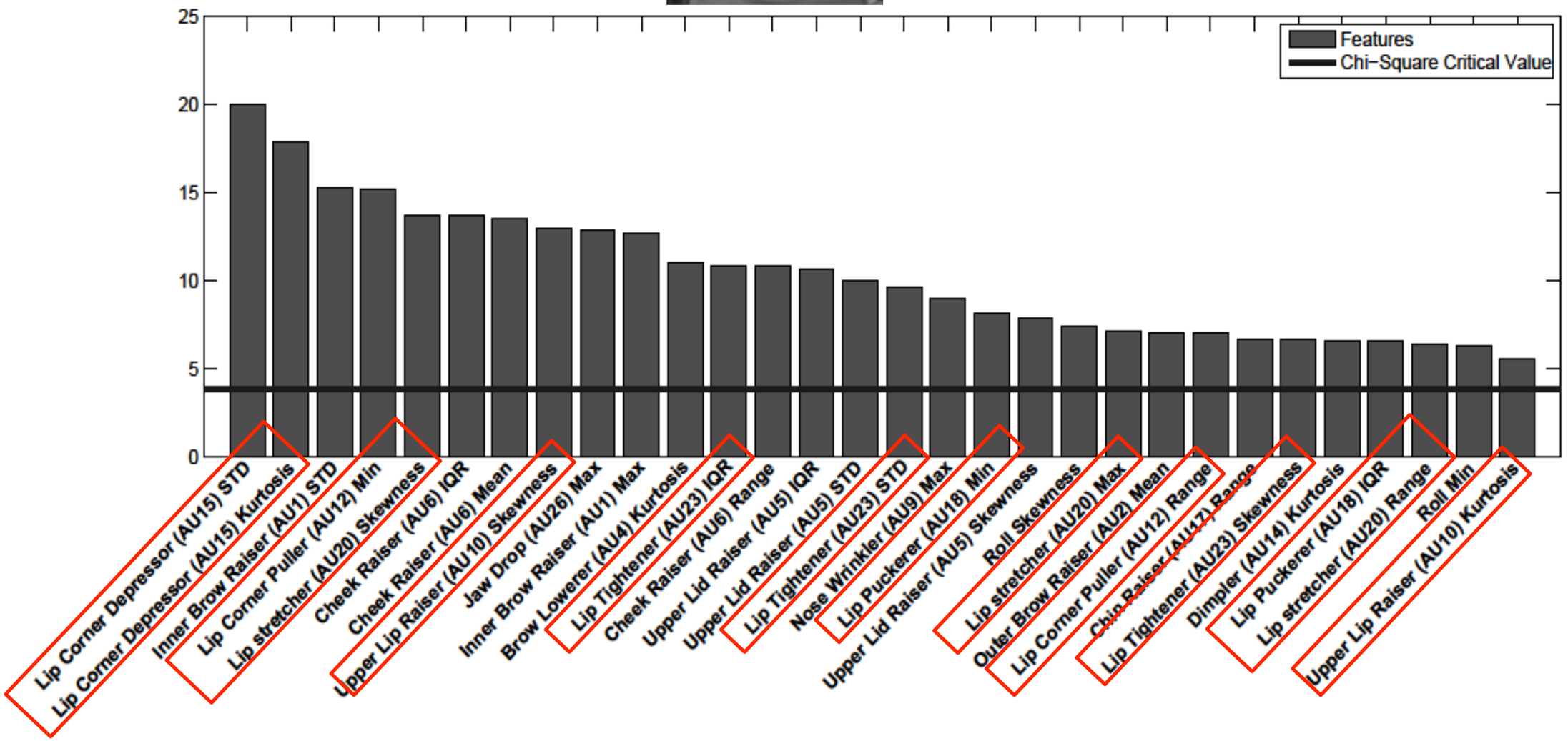
Logistic Regression Analysis

Lip Related AUs

AU15

AU12

AU23

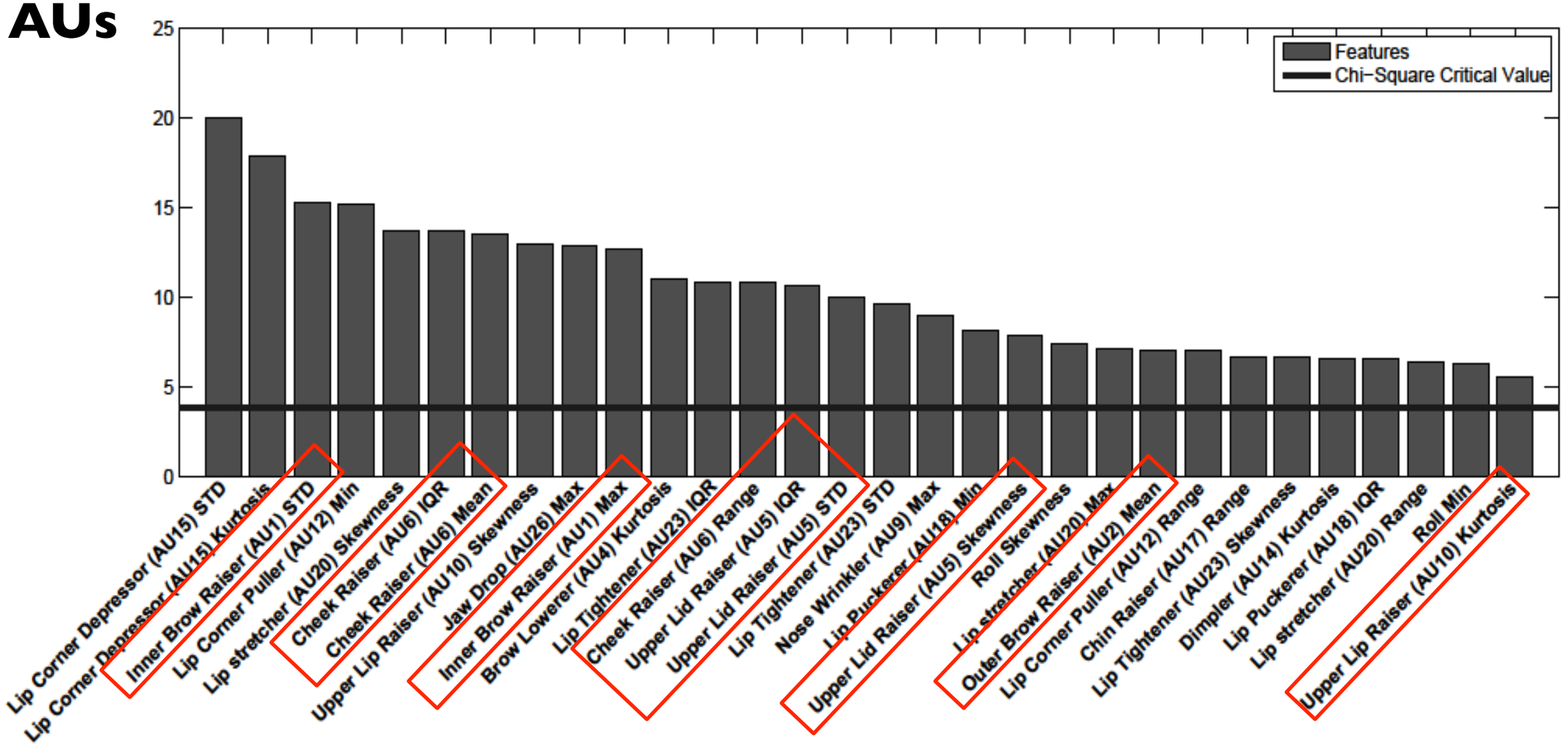


The horizontal line indicates the threshold for which the individual features are statistical significant at p-value=0.05.

Logistic Regression Analysis

Upper face Related AUs

AU6 **AU5** **AU2** **AU1**



The horizontal line indicates the threshold for which the individual features are statistical significant at p-value=0.05.

Conclusions

- Facial information is useful for driver distraction detection.
- Gaze features and AUs provide valuable information for visual distraction detection.
- AUs play an important role in cognitive distraction detection.
- AUs are also useful for detecting when cognitive distraction is not induced by visual distraction.

Future work

- Include multimodal signals for visual and cognitive distraction detection
  - CAN-Bus
  - Audio
  - Road Camera
- Include other cognitive tasks
- Cover a wide range of scenarios under different road and environment conditions
- Build road dependent driver modals

Thank you!

Questions?

MSP@UTDALLAS.EDU