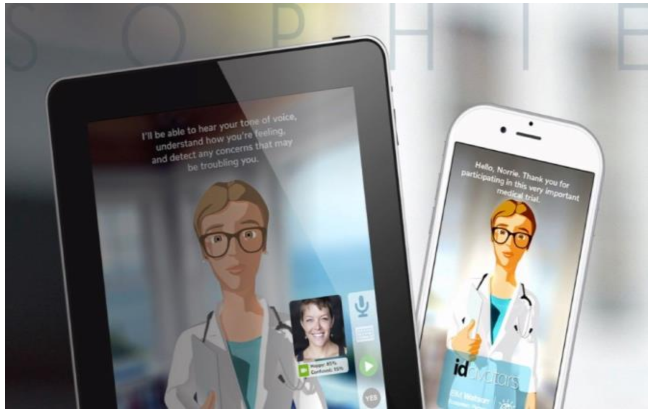
:) Affectiva

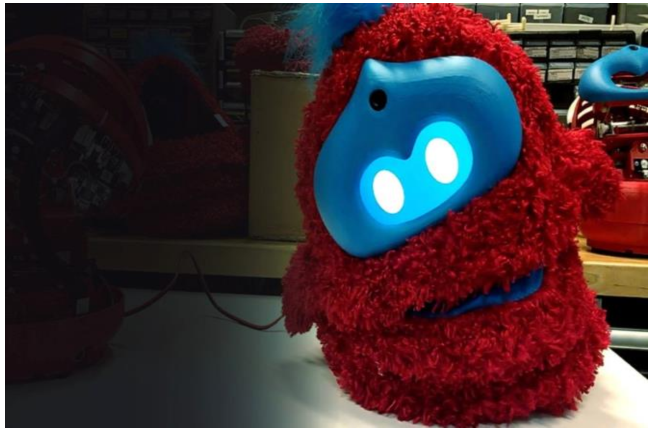
# Multi-Modal Emotion Estimation

Presenter: *Dr. Mohammad Mavadati and Dr. Taniya Mishra*

# Our Emotions, influence how we live and experience life!

## But, we’re also surrounded by **High IQ and no EQ devices**



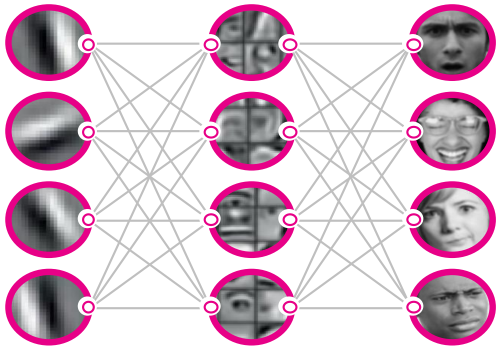
# Affectiva mission: **humanize technology with Human Perception AI**

Pioneers **of Human Perception AI.**

AI software that understands all **things human – nuanced human emotions,** complex **cognitive states**, behaviors, activities, interactions and objects people use.

**Only multi-modal in cabin sensing AI.**

Using deep learning, computer vision, voice analytics and massive amounts of data, Affectiva analyzes face and voice to understand state of humans in vehicle.



*Face:*
7 emotions, indicators of attention, drowsiness, distraction, positive / negative, 20+ facial expressions and demographics

*Voice:*
Arousal, laughter, anger, gender

# Emotion AI detects emotion and cognitive states the way people do

## People communicate through multiple modalities

**55%**
Facial expressions and gestures

**38%**
How the words are said

**7%**
The actual words

*Source: Journal of Consulting Psychology.*

## Affectiva's multi-modal Emotion AI

### Face

7 emotions, indicators of attention, drowsiness, distraction, positive / negative, 20+ facial expressions and demographics

### Multi-modal

- Developing early and late fusion of modalities for deeper understanding of complex states
- Expanding beyond face and voice

### Voice

Arousal, laughter, anger, gender

# Emotion AI is **a multi-modal and multi-dimensional problem**

**Multi-modal** - Human emotions manifest in a variety of ways including your tone of voice and your face

**Many expressions** - Facial muscles generate hundreds of facial actions, speech has many different dimensions - from pitch and resonance, to melody and voice quality

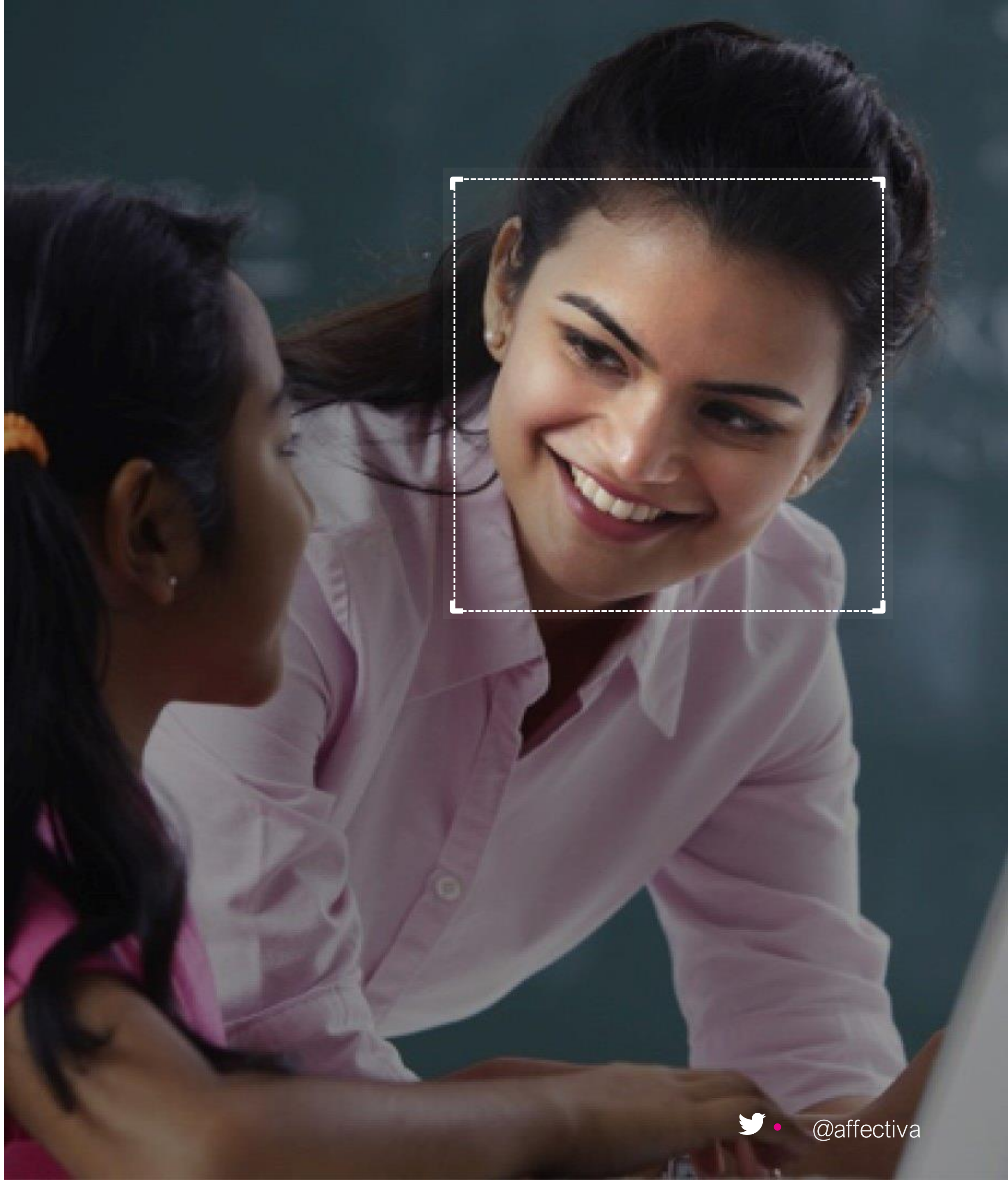
**Highly nuanced** – Emotional and cognitive states can be very nuanced and subtle, like an eye twitch or your pause patterns when speaking

**Non-deterministic** - Changes in facial or vocal expressions, can have different meanings depending on the person's context at that time

**Temporal lapse**- As an individual's state unfold over time, algorithms need to measure moment by moment changes to accurately capture of mind

**Context** – Understanding complex state of mind requires contextual knowledge of the surrounding environment and how an individual is interacting with it

**Massive data** - Emotion AI algorithms need to be trained with massive amounts of real world data that is collected and annotated



# Display and perception of emotion is not perfectly aligned

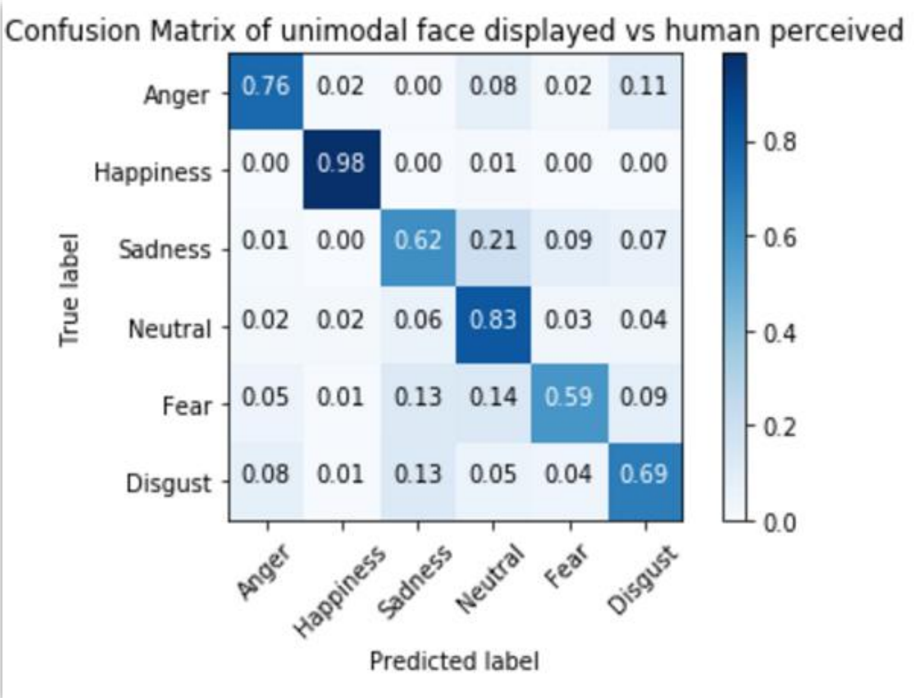
**CREMA-D*:** large scale study of emotion and perception

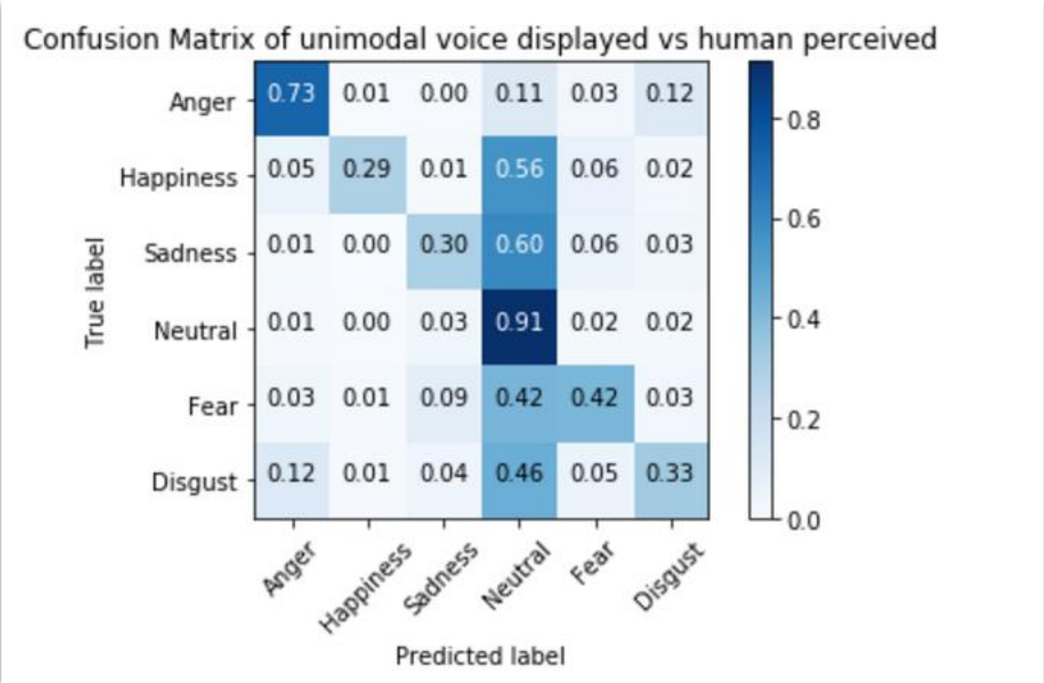
- 91 participants
- 6 emotions of varying intensities
- 7442 emotion samples.
- 2443 observers

Human recognition of intended emotion based on

- voice-only: 40.9%
- face-only: 58.2%
- face and voice: 63.6%

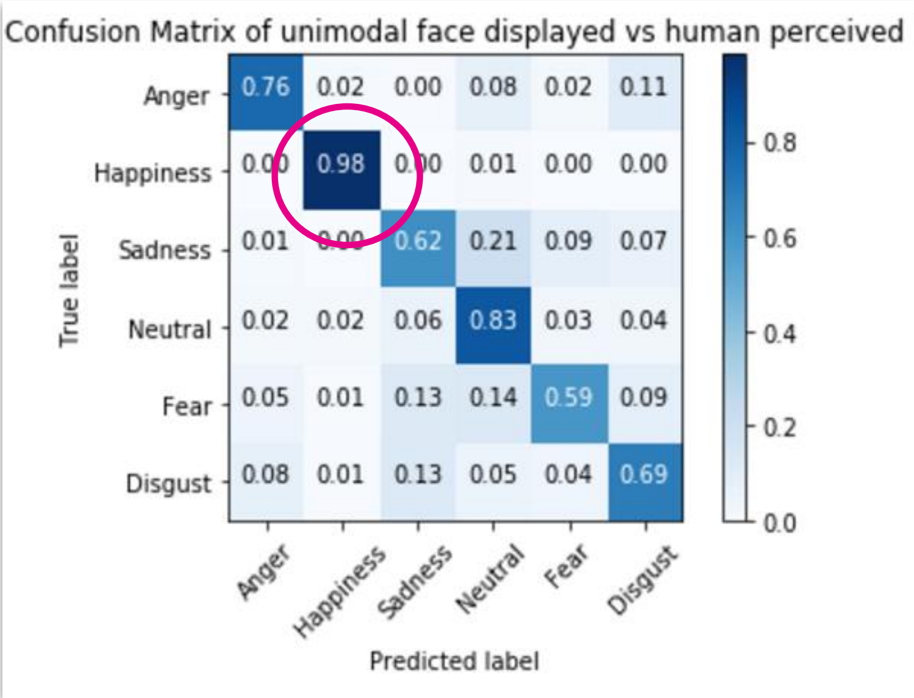
# Difference in emotion perception from Face vs. Speech modalities

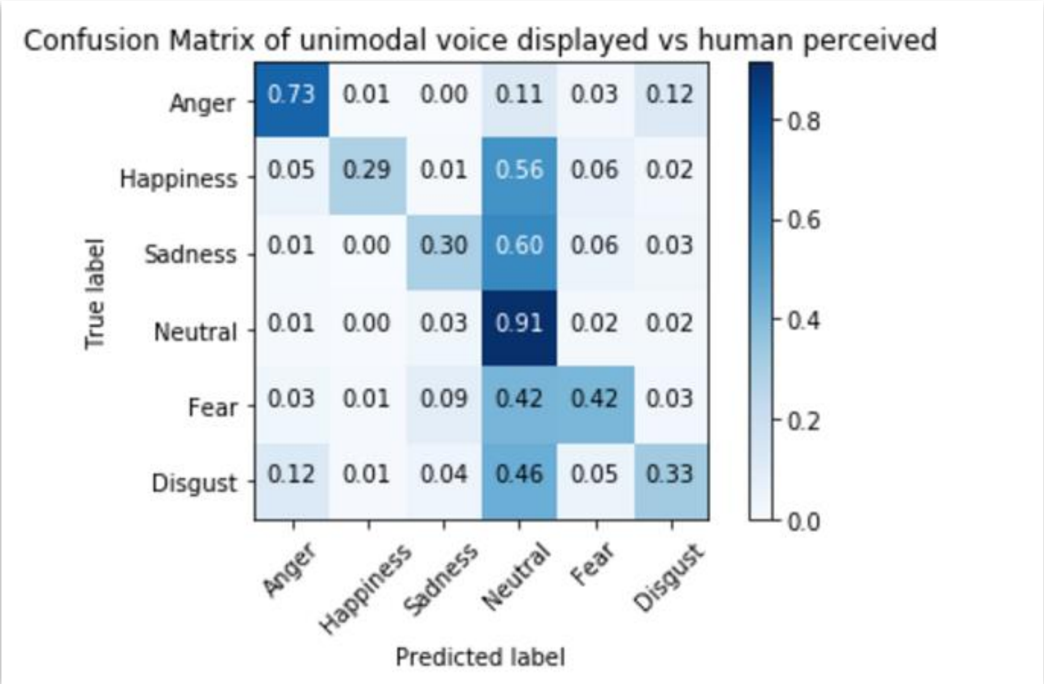


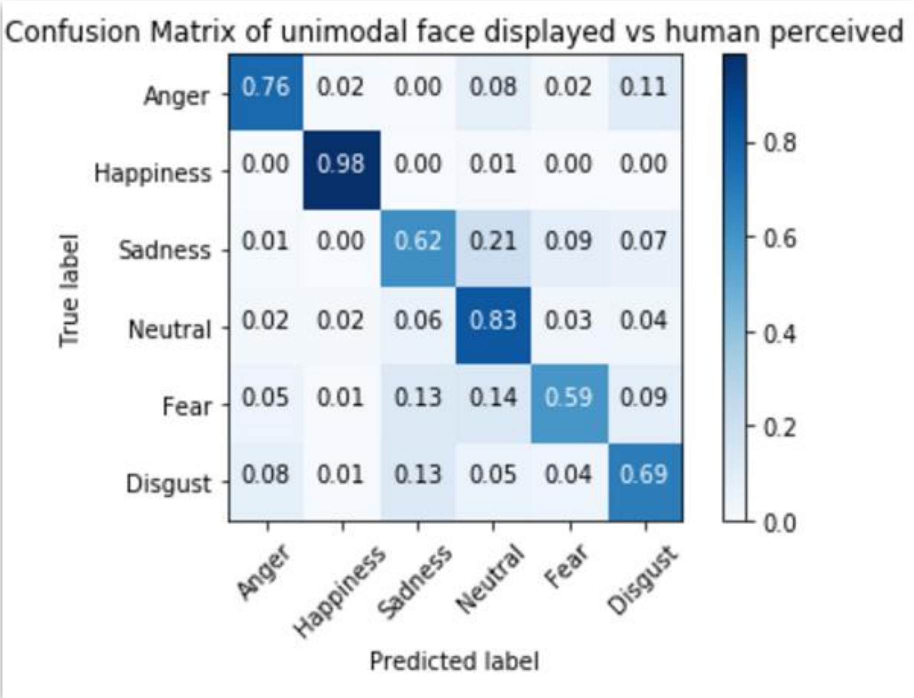

Confusion matrices showing emotions displayed by humans, recognized by other human observers

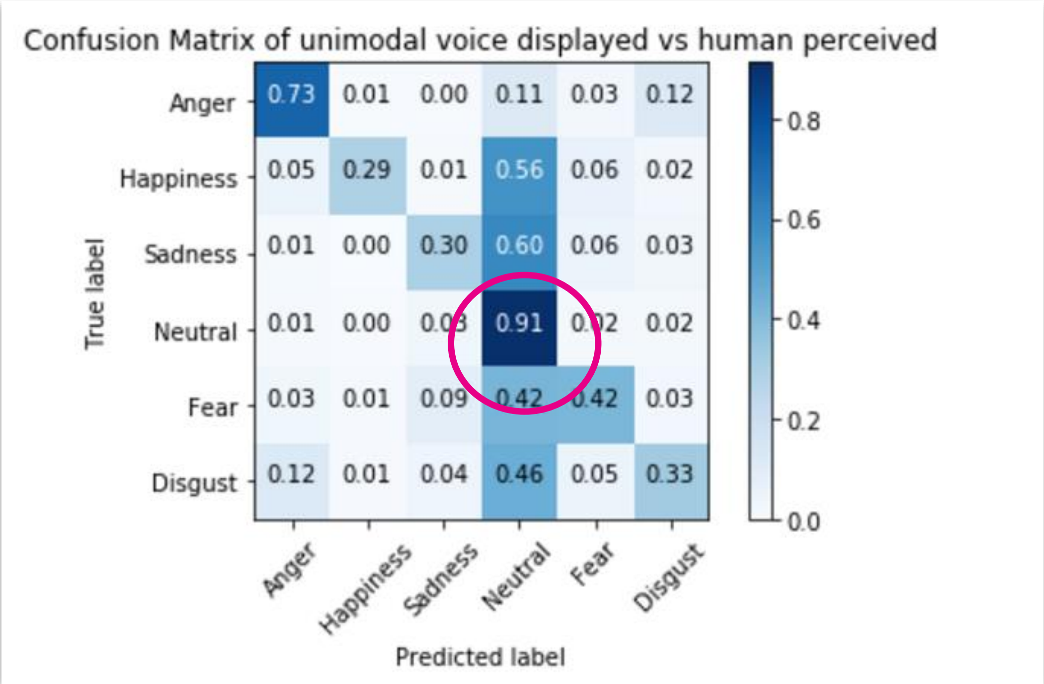
# Difference in emotion perception from Face vs. Speech modalities



# Difference in emotion perception from Face vs. Speech modalities



# Emotion AI at Affectiva

# How it **works**

# Data driven approach to Emotion AI

## Data

**Multi-Modal Data Acquisition**

Large amounts of real world video & audio data; different ethnicities and contexts

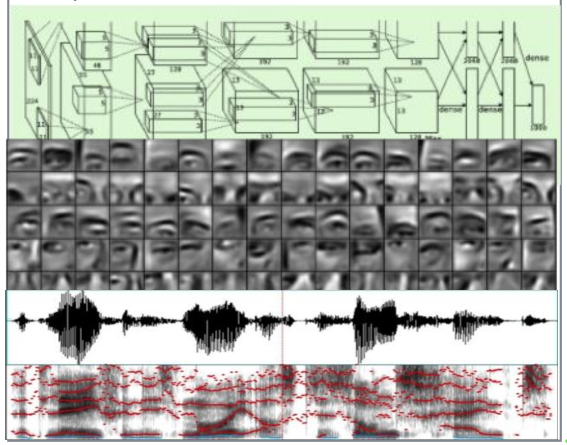
**Data Annotation Infrastructure**

Manual and automated labeling of video and speech

## Algorithms

**Training & Validation**

Parallelize deep learning experiments on a massive scale



## Evaluation

**Output**

Multi-modal classifiers for machine perception, e.g., expressions, emotions, cognitive states and demographics

**Product Delivery: APIs and SDKs**

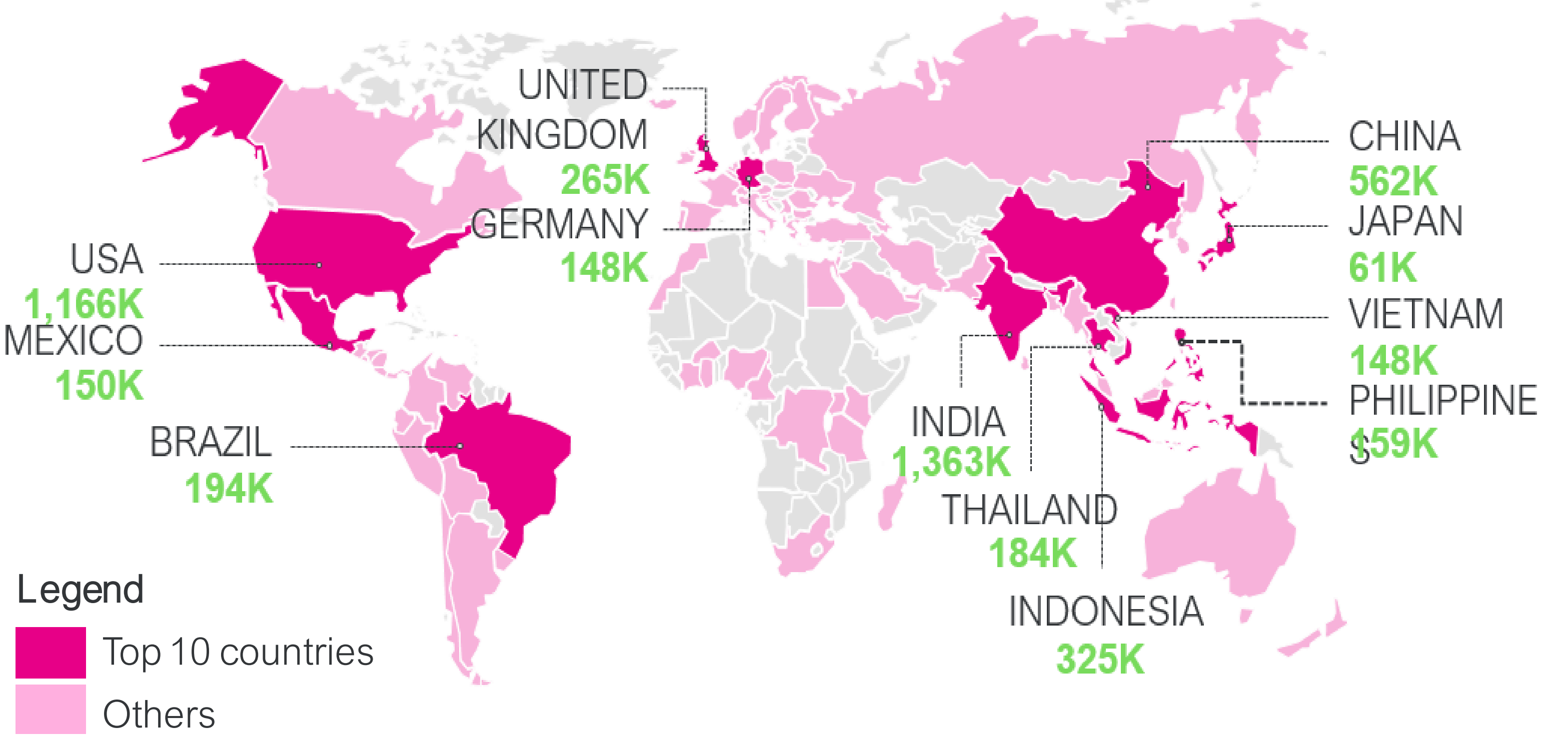
The classifiers and run-time system are optimized for the cloud or on device or embedded

iOS

HTML

# Data matters ...

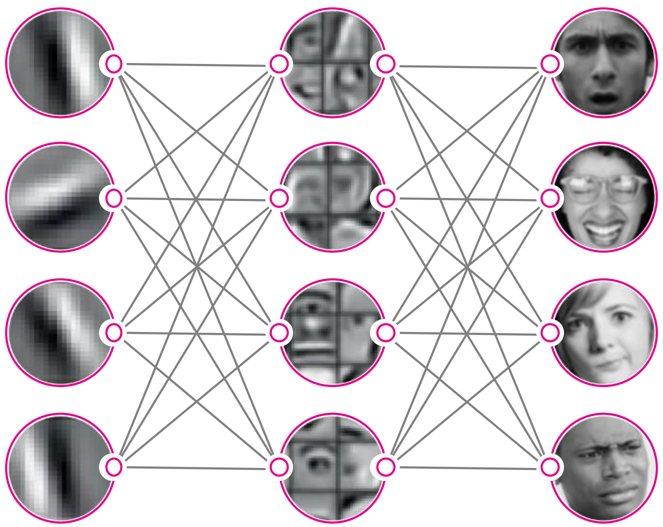
# Massive proprietary data and annotations power our AI

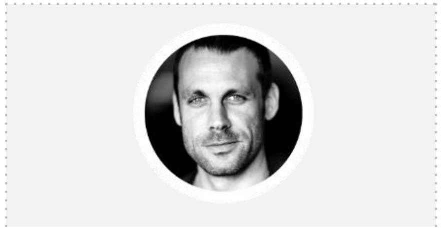



| | |
|---|---|
| **4Bn** FRAMES | **7.5MM** FACES |
| **836MM** AUTO FRAMES | **87** COUNTRIES |

- ✓ **Foundation: Large, diverse & real world data built in the past 7 years**
- ✓ **Growing automotive in-cabin data with scalable data acquisition strategy**

# Deep **learning**





| | |
|---|---|
| Anger | **0.09133** |
| Contempt | **0.62842** |
| Disgust | **0.20128** |
| Fear | **0.00001** |
| Happiness | **0.00041** |

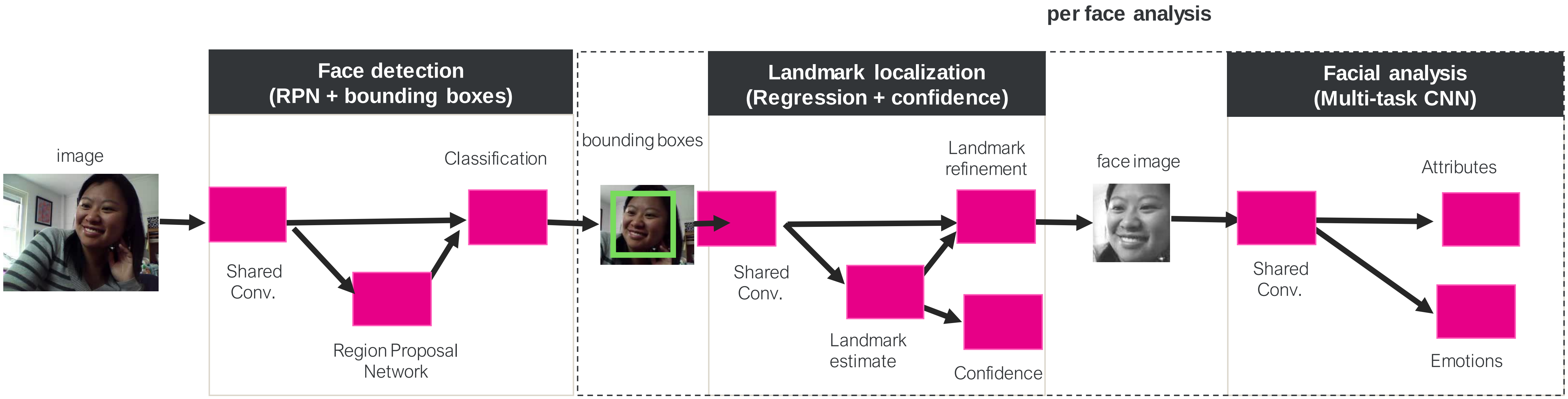
**Affectiva's focus is on deep learning**

- It allows modeling of more complex problems with higher accuracy than other machine learning techniques
- Allows for end-to-end learning of one or more complex tasks jointly
- Solves a variety of problems: classification, segmentation, temporal modeling

# Vision pipeline

The current vision SDK consists of steps

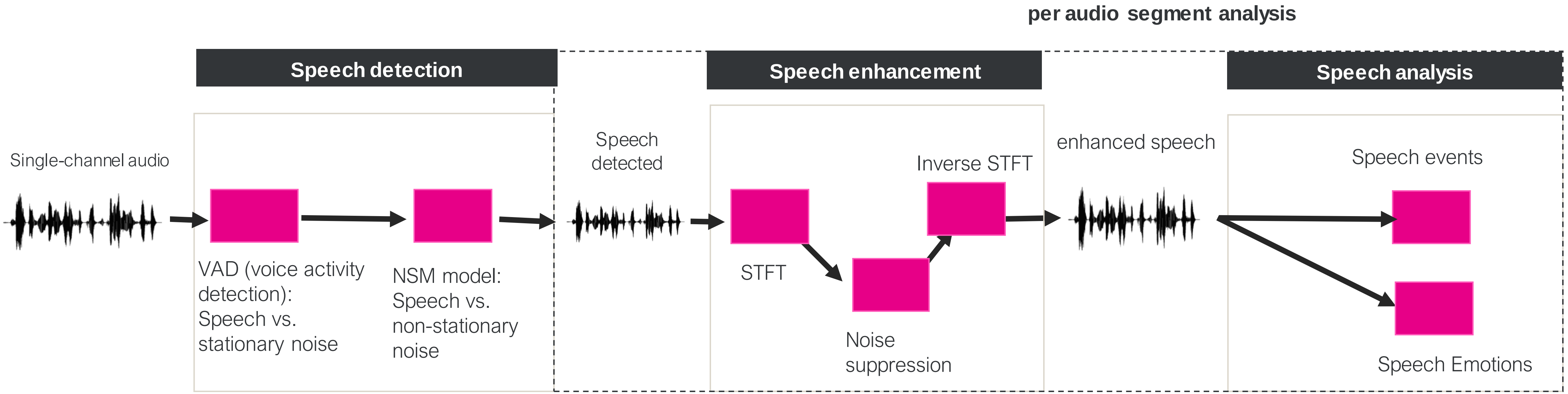
- **Face detection:** given an image, detect faces
- **Landmark localization:** given a image + bounding box, detect and track landmarks
- **Facial analysis:** detect facial expression/emotion/attributes



# Speech pipeline

The current speech pipeline consists of these steps:

- **Speech detection:** given audio, detect speech
- **Speech enhancement:** given noisy speech speech segment, mask noise
- **Speech analysis:** detect speech events/emotion/attributes



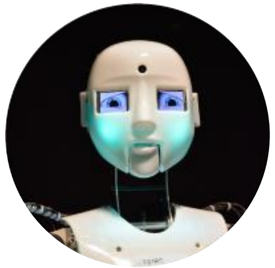
# Multi-Modal Applications


Media and entertainment


Advertising


Human resources


Automotive


Robotics


Healthcare and quantified self


Video communication


Online education


Devices


Gaming

# Multimodal for Automotive

# Affectiva Automotive AI

# Human Perception AI fuels deep understanding of people in a vehicle

Delivering valuable services to vehicle occupants depends on a deep understanding of their current state

## Third Party Solutions

**In-Cab Context**
Infotainment content
Inanimate objects
Cabin environment

**External Context**
Weather
Traffic
Signs
Pedestrians

**Personal Context**
Identity
Likes/dislikes & preferences
Occupant state history
Calendar

+

## Affectiva **Automotive AI**


=

## Advanced Vehicle Services

**Safety**
Next generation driver monitoring
Smart handoff & safety drivers
Proactive intervention

**Occupant Experience**
Individually customized baseline
Adaptive environment
Personalization across vehicles

**Monetization**
Differentiation among brands
Premium content delivery
Purchase recommendations

# Affectiva Automotive AI

## Modular and extensible deep learning platform for in-cabin human perception AI

- Core technology is shared and reused across different modules
- Modular packaging enables light-weight deployment of capabilities for a specific use case
- Extend existing capabilities by adding more modules

### Driver Monitoring

- Drowsiness levels
- Distraction levels
- Cognitive load

### Occupant State

- **Facial and vocal emotion**
- **Mood (valence)**
- **Multimodal emotion: frustration**
- **Engagement**

### Occupant Activities

- Talking
- Texting
- Cellphone in hand

### Cabin State

- Occupant location and presence
- Objects left behind
- Child left behind

### Core Technology

**Face & head tracking**

- 3D Head pose

**Facial expression recognition**

- 20 Facial expressions:
  e.g. smile, eye brow raise
- Drowsiness markers:
  eye closure, yawn, blink

**Object detection**

- Object classes:
  mobile device, bags
- Object location

**Voice detection**

- Voice activity detection

**Flexible Platform**

- Support Near IR sensors
- Support ARM ECU
- Support multiple camera positions

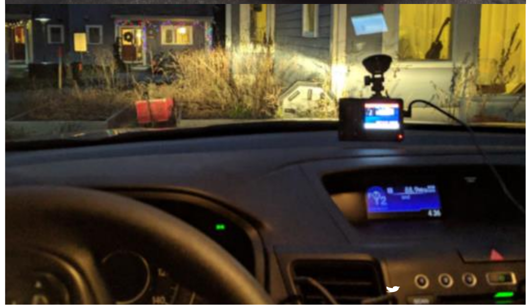
# Automotive data collection for multimodal analysis

# Automotive Data Acquisition

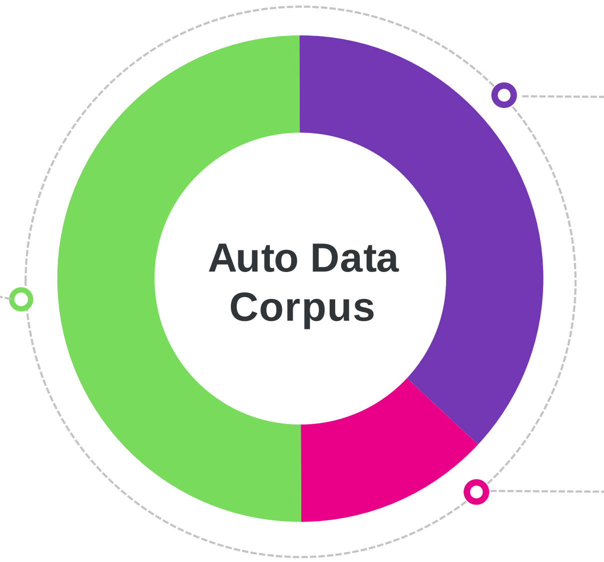
To develop a deep understanding of the state of occupants in a car, one needs large amounts of data. With this data we can develop algorithms that can sense emotions and gather people analytics in real world conditions.

## Spontaneous occupant data

Using Affectiva Driver Kits and Affectiva Moving Labs to collect naturalistic driver and occupant data to develop metrics that are robust to real-world conditions

In-Car Data Acquisition (Quarterly)
42,000 miles and 2,000+ hours driven
200+ drivers on 3 continents




## Data partnerships

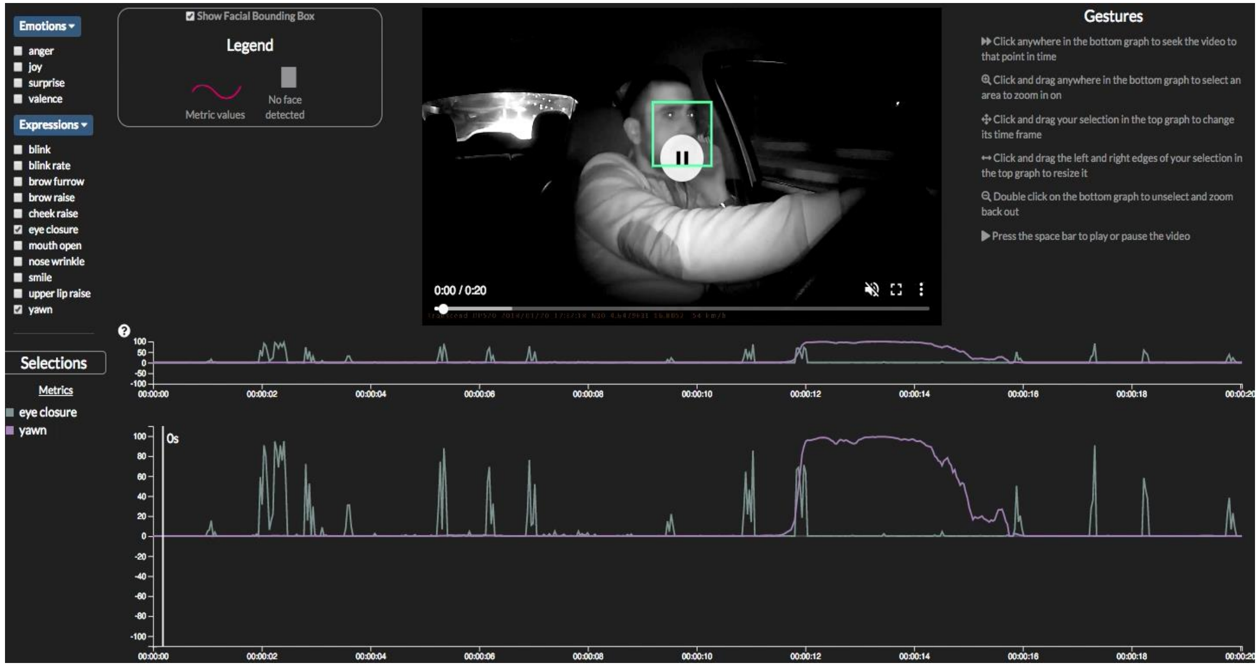
Acquire 3rd party natural in-cab data through academic and commercial partners (MIT AVT, fleet operators, ride-share companies)

## Simulated data


Collect challenging data in safe lab simulation environment to augment the spontaneous driver dataset and bootstrap algorithms (e.g. drowsiness, intoxication) multi-spectral & transfer learning.

# Automotive AI data

Automotive AI 1.0 tracks metrics for driver monitoring as well as emotion estimation




Driver Drowsiness

Detecting eye closure and yawning events


Emotion detection

Detect driver emotions including surprise and joy

# Multimodal frustration: A case study

# Why detect frustration?

Frustration is “the occurrence of an obstacle that prevents the satisfaction of a need” [Lawson, 1965].

A frustrated driver can be a dangerous driver.

- Frustration has been shown to be accompanied by various driving behaviors, such as, horn honking, purposeful tailgating and flashing high beams [Hennessey and Wiesenthal, 1999].
- Overtaking was found to be correlated with a state of frustration [Kinnear et al., 2015]
- Malta et al. found that the intensity of pedal actuation signals --- hard braking --- correlated with frustration [Malta et al., 2011].

Automatic in-cabin sensing of affective states such as frustration can utilize that information to provide effective interventions that attempt to minimize unsafe behavior. For example, If driver is irritated because of a traffic jam, agent suggests an alternative route.

# In-lab data collection to elicit Frustration

- Participants were asked to do 6 timed tasks requiring interactions with a voice agent (Alexa) to mimic interactions with car HMI in 2 sessions.

  Multi-tasking: interacting with the voice agent while driving
  Uni-tasking: only interacting with the voice agent; no driving

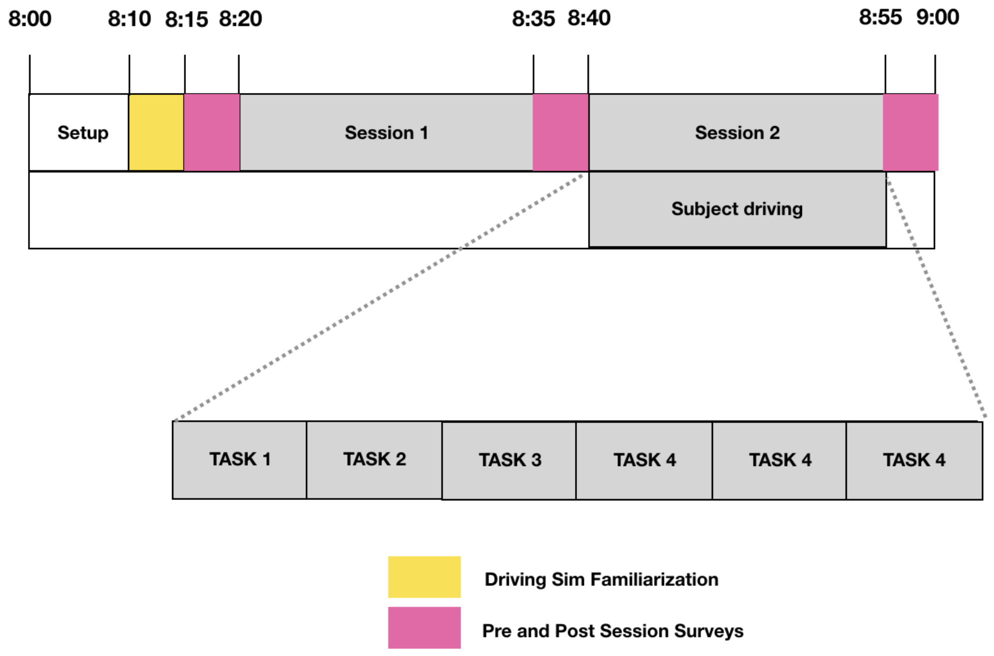
- Tasks designed to mimic real interactive conversations that people might have with an in-car assistant.

  difficulty

  Make a shopping list
  Set a timer/alarm
  Request system to say something funny
  Request a particular song by name
  Request a particular radio station call number and frequency
  Dictate an email to a particular person

- Wizard-of-Oz setting: dialogue from Alexa pre-recorded and played by study administrator.

- 105 participants: 55 female, 47 male and 3 did not specify gender



# Instrumentation

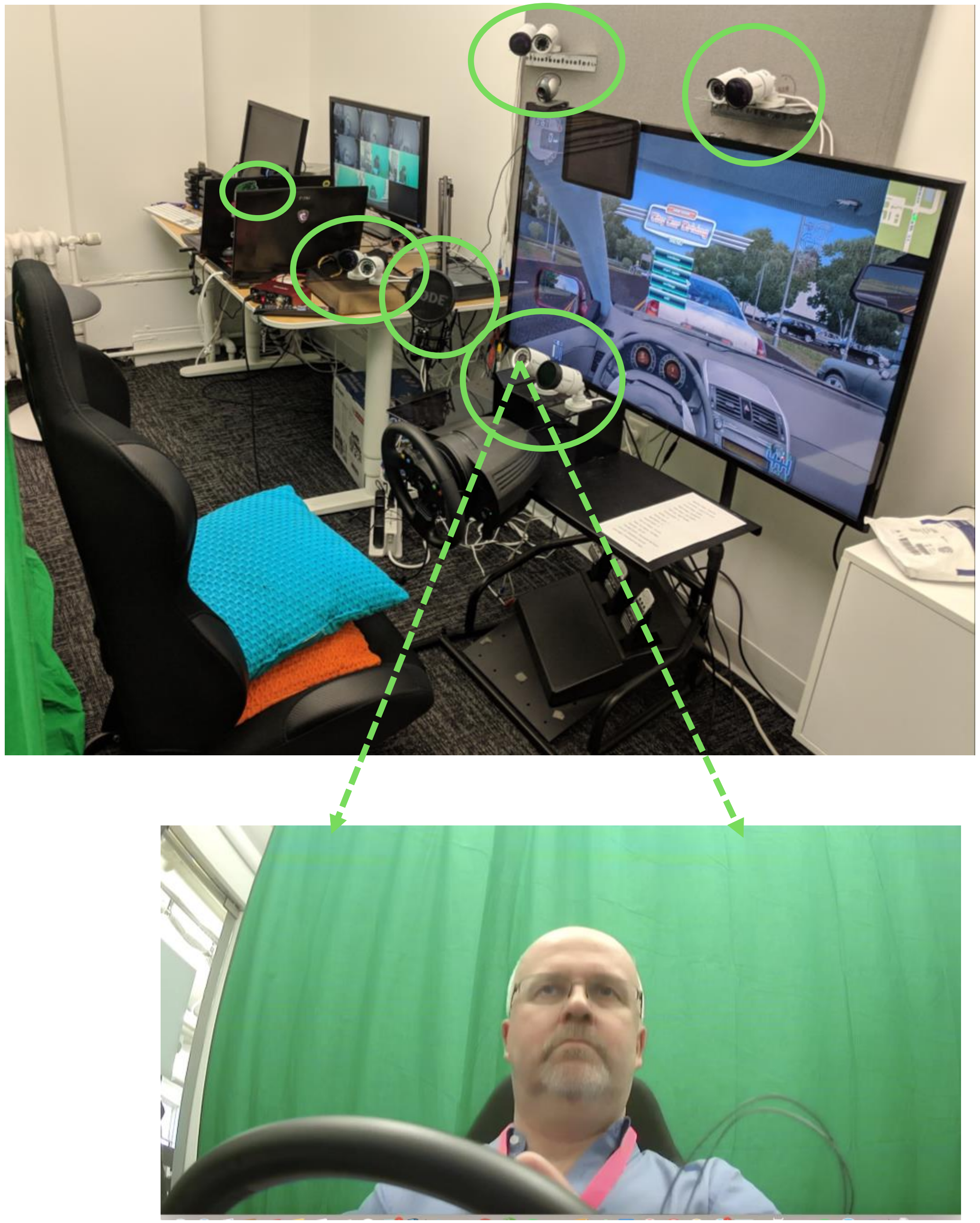
- Multi-cameras and audio setup (4 pairs of NIR and RGB cameras, 2 additional cameras, 3 microphones): The multi-camera audio-video setup was used to capture multiple views of the participant as well as their audio stream.

- ECG: Subjects were asked to put 4 ECG sensors on their body to measure heart rate.

- GSR: Subjects also wore a skin conductance sensor.

- Integration platform: An software platform that allowed study admin to see and hear the participant, their vitals, and their performance on the driving sim, so that pre-recorded voice responses could be played appropriately to simulate HMI.

- Total: 24 pieces of hardware and matching software.



# Challenges of data collection

Setup and syncing multiple sensors.

- 24 pieces of hardware and matching software
- Individually not difficult to set up
- But setup and sync non-trivial

Eliciting “real” frustration in participants.

- Engagement constraint: Frustration had to be managed. Some tasks purposely frustrating but not all --- otherwise people would give up; some tasks had to be easy to accomplish so people could win at it and stay engaged.

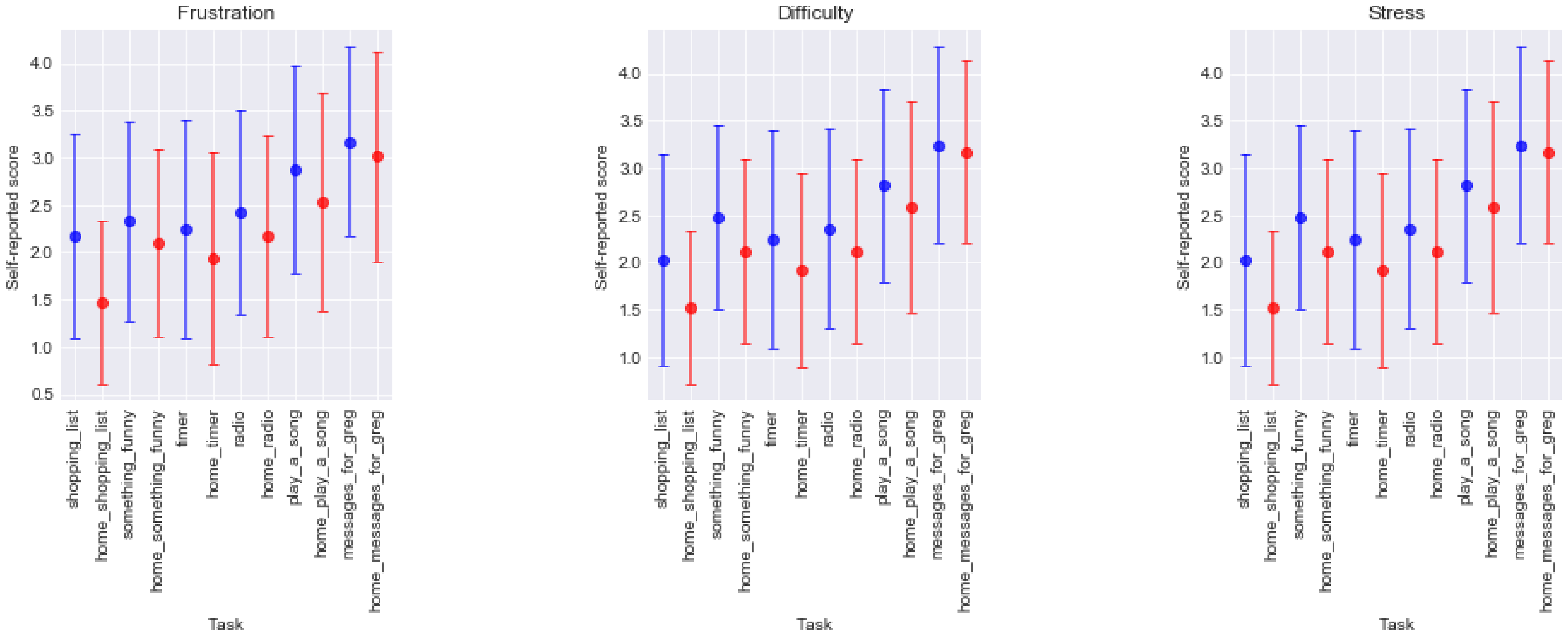
- Believability constraint: Requests and responses in scenarios had to be believable/acceptable yet frustrating.

# Example: Frustrated due to difficulty getting radio to play

# Analysis of frustration from face and voice

# Self report: Is multitasking more frustrating?

## Multitasking defined as driving + HMI interaction

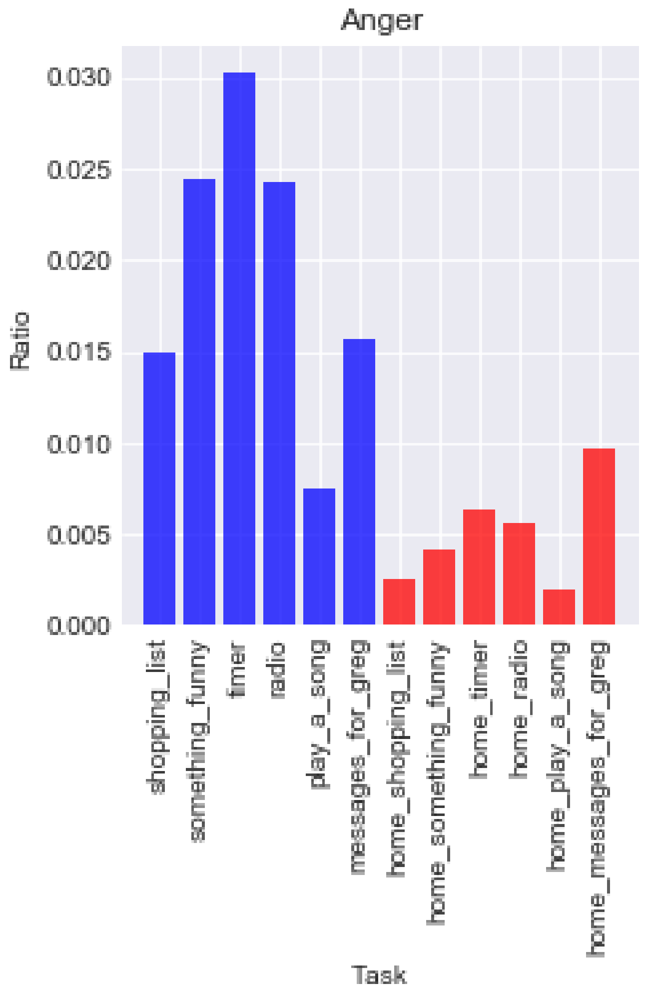


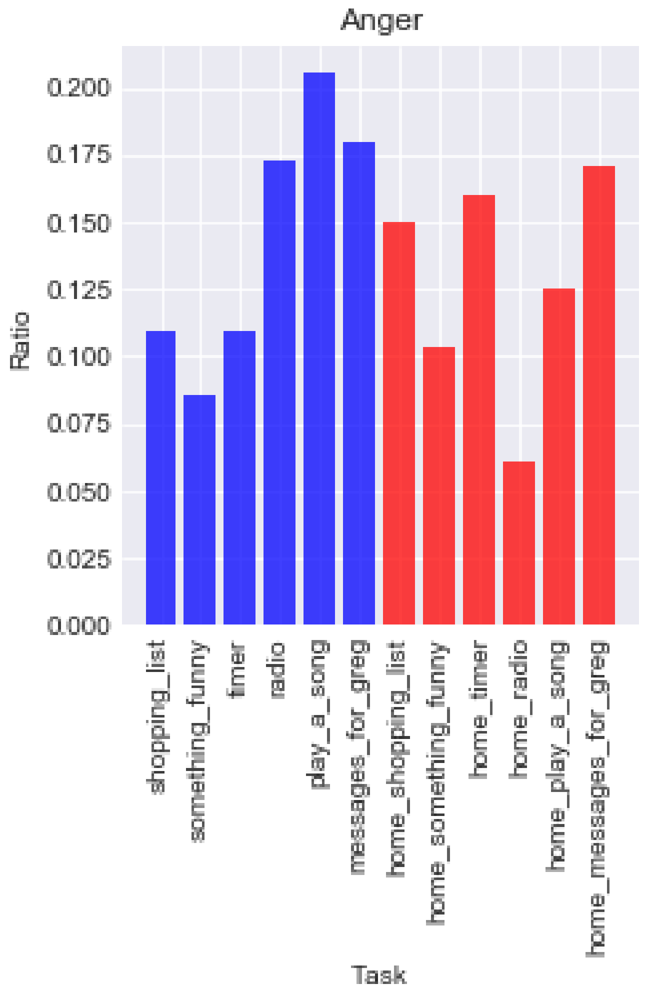

Self-reported Frustration, Difficulty and Stress Values for each task

# Automatic analysis: Is multitasking more frustrating?

**Multitasking defined as driving + HMI interaction**

**Face**



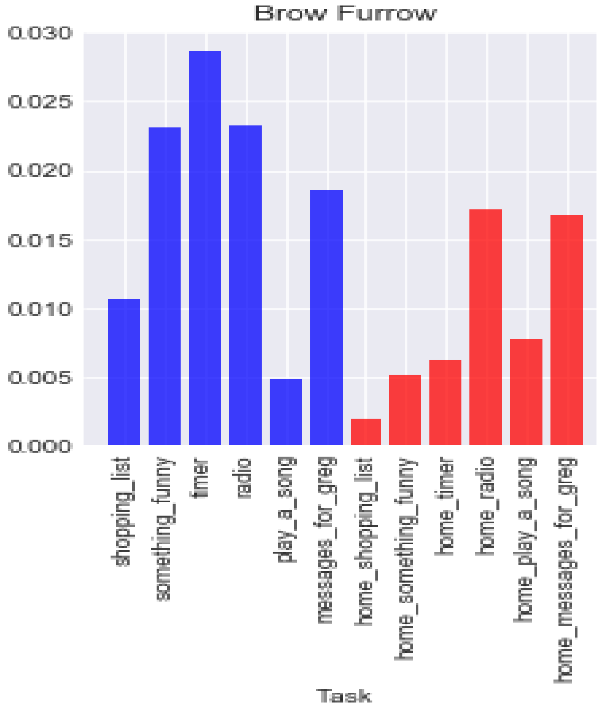



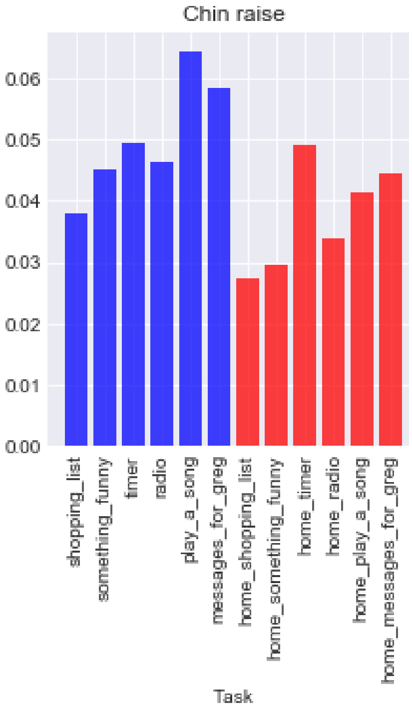

**Speech**

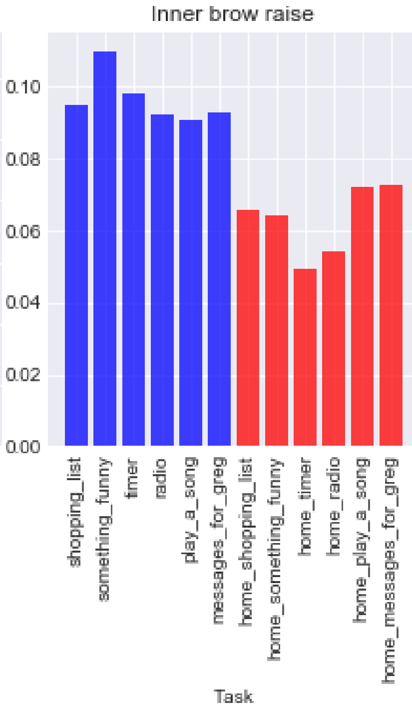
**Average percentage of anger activation for different tasks**

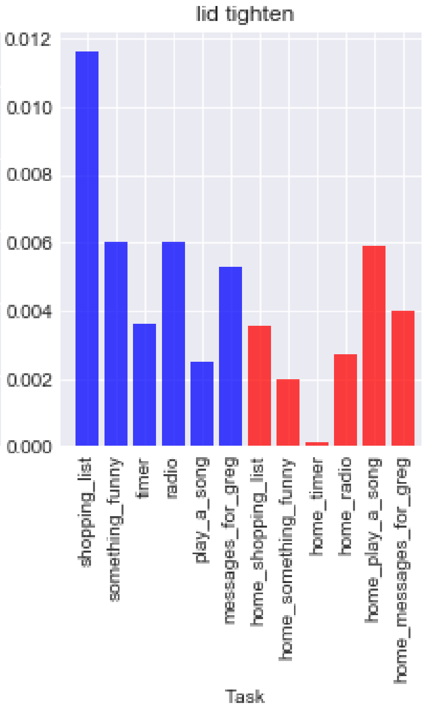
# Automatic analysis: Is multitasking more frustrating?

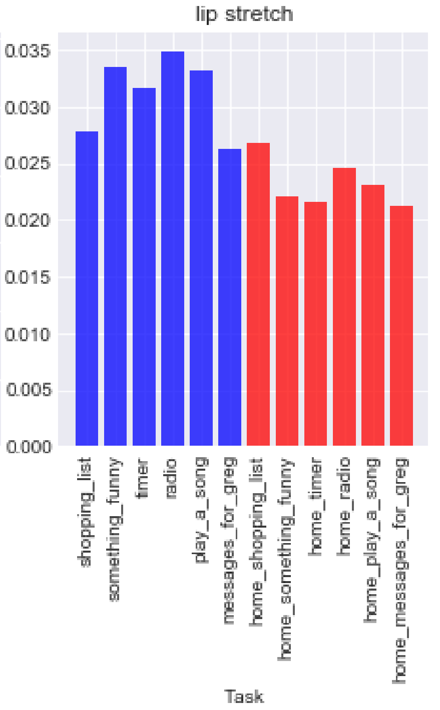
## Multitasking defined as driving + HMI interaction



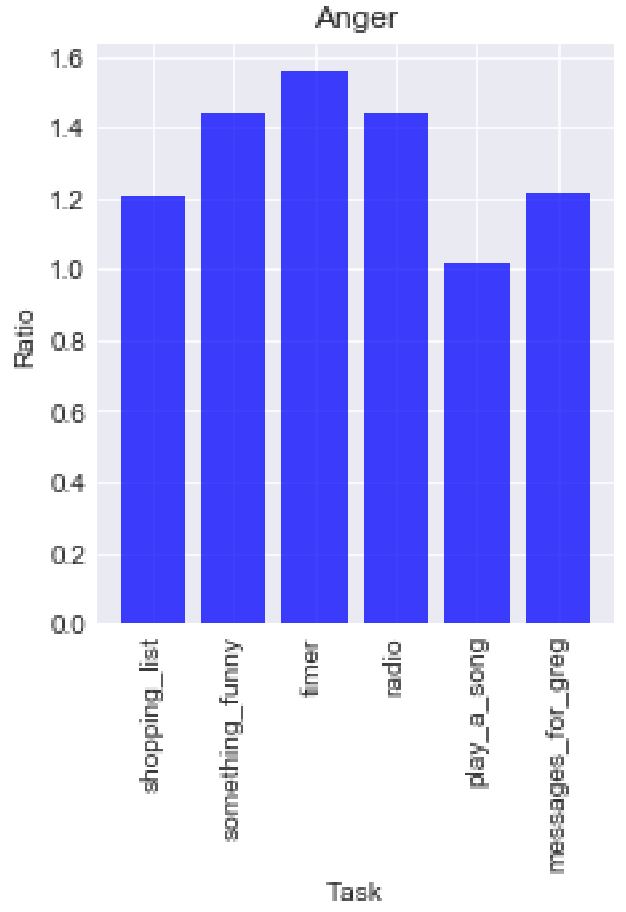



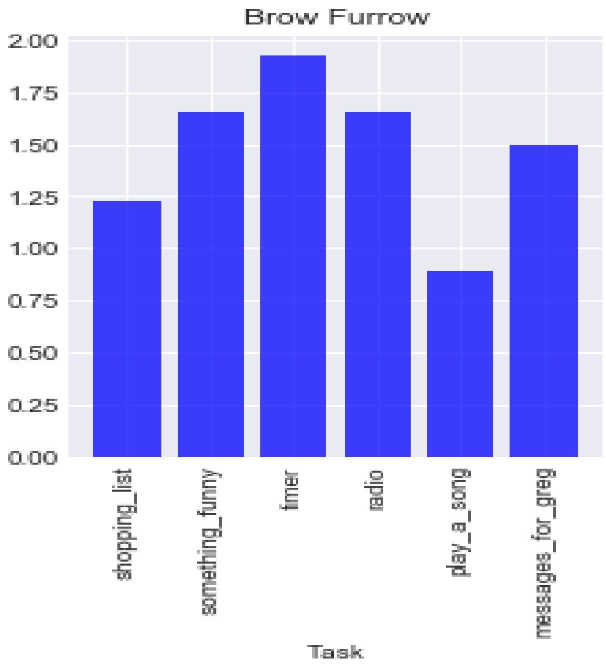



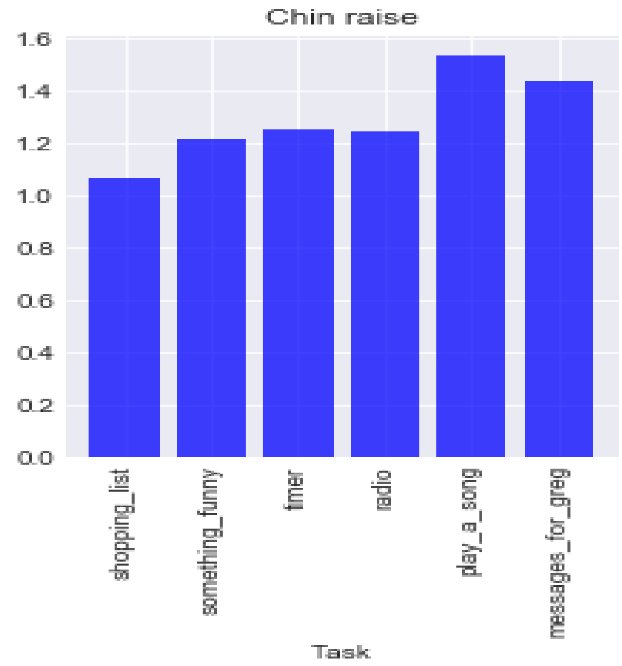




**Average percentage of activation of metrics for different tasks**

# How much more frustrating is multitasking compared to free driving?

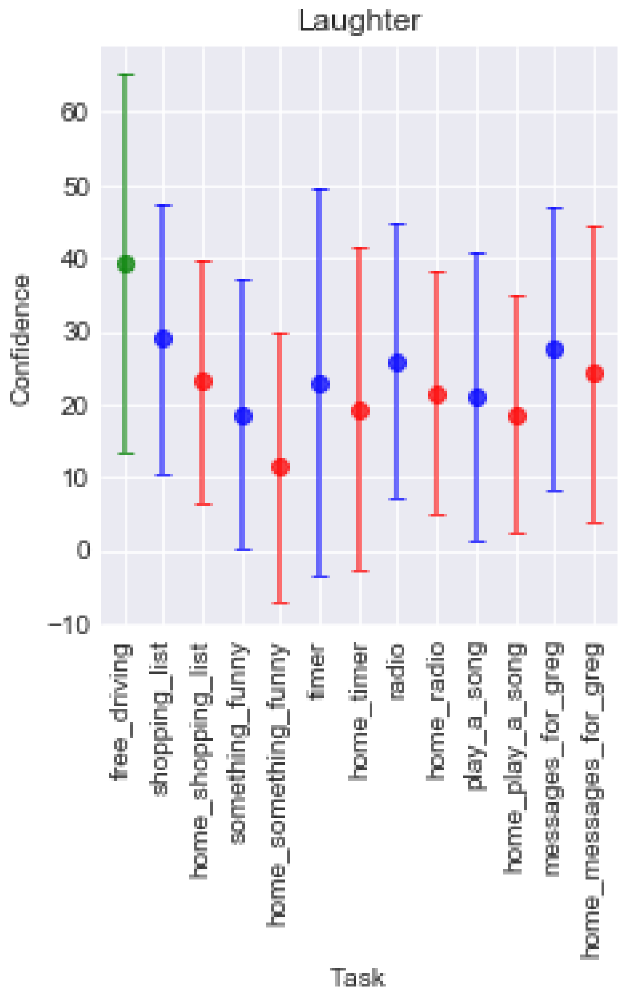



**Average ratio of facial activations for different tasks with respect to its average value for free driving**

# Unexpected observation: Laughing while frustrated




# Next steps:
# multimodal frustration detection

# Analyzing other markers of frustration

### Driving behavior

- Examine behaviors such as honking, tailgating and flashing of high beams [Hennessey and Wiesenthal, 1999], overtaking [[Kinnear et al., 2015] and pedal actuation signals [Malta et al., 2011].

### Gestures and body posture

- Hand movements provide a means for displaying frustration [Dittmann and Llewelyn, 1969]

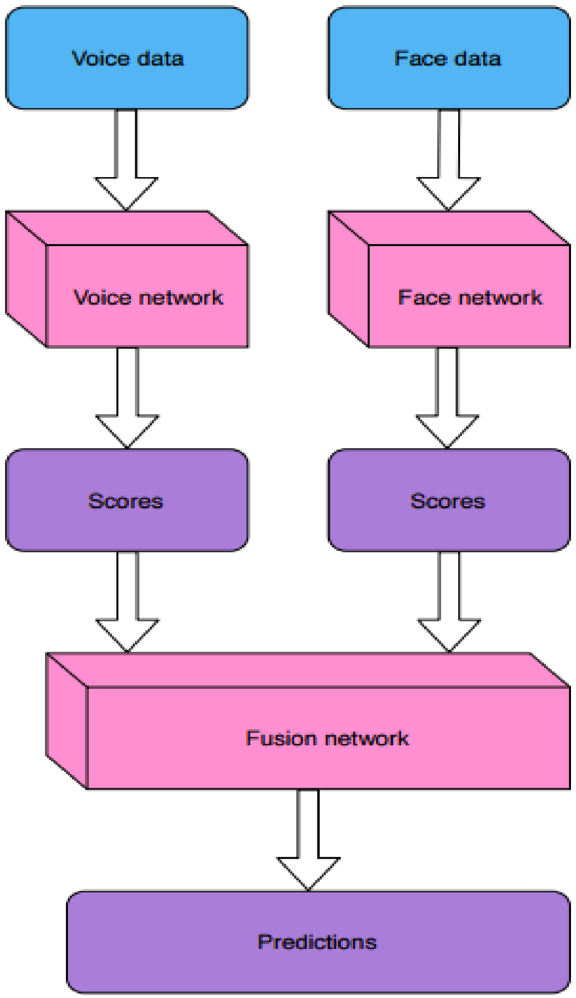
### Physiological responses

- Fernandez and Picard, 1998 showed that electrodermal response (GSR) is indicative of human frustration in interacting with systems.

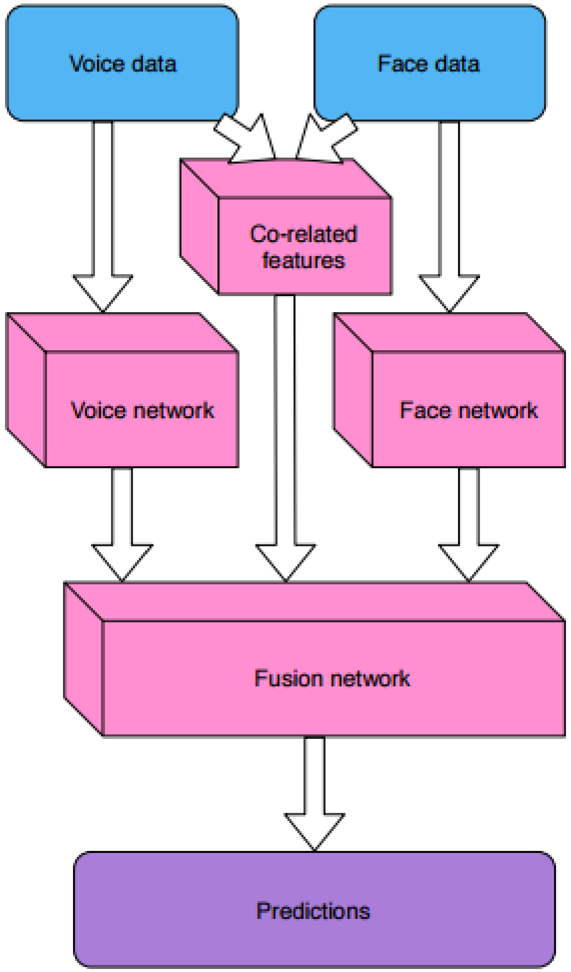
- Belle et al. 2010 analyzed ECG data of students and found that the ECG profile of person who is calm can be distinguished from a person who is frustrated.

# Multimodal Training Strategies for Frustration detection

## Decision Level Fusion




## Feature Level Fusion



# Human Perception AI fuels deep understanding of people in a vehicle

Delivering valuable services to vehicle occupants depends on a deep understanding of their current state

## Third Party Solutions

**In-Cab Context**
Infotainment content
Inanimate objects
Cabin environment

**External Context**
Weather
Traffic
Signs
Pedestrians

**Personal Context**
Identity
Likes/dislikes & preferences
Occupant state history
Calendar

+

## Affectiva Automotive AI


=

## Advanced Vehicle Services

**Safety**
Next generation driver monitoring
Smart handoff & safety drivers
Proactive intervention

**Occupant Experience**
Individually customized baseline
Adaptive environment
Personalization across vehicles

**Monetization**
Differentiation among brands
Premium content delivery
Purchase recommendations

# Learn more
# www.affectiva.com

Contact us:
Email: taniya.mishra@affectiva.com
Email: mohammad.mavadati@affectiva.com