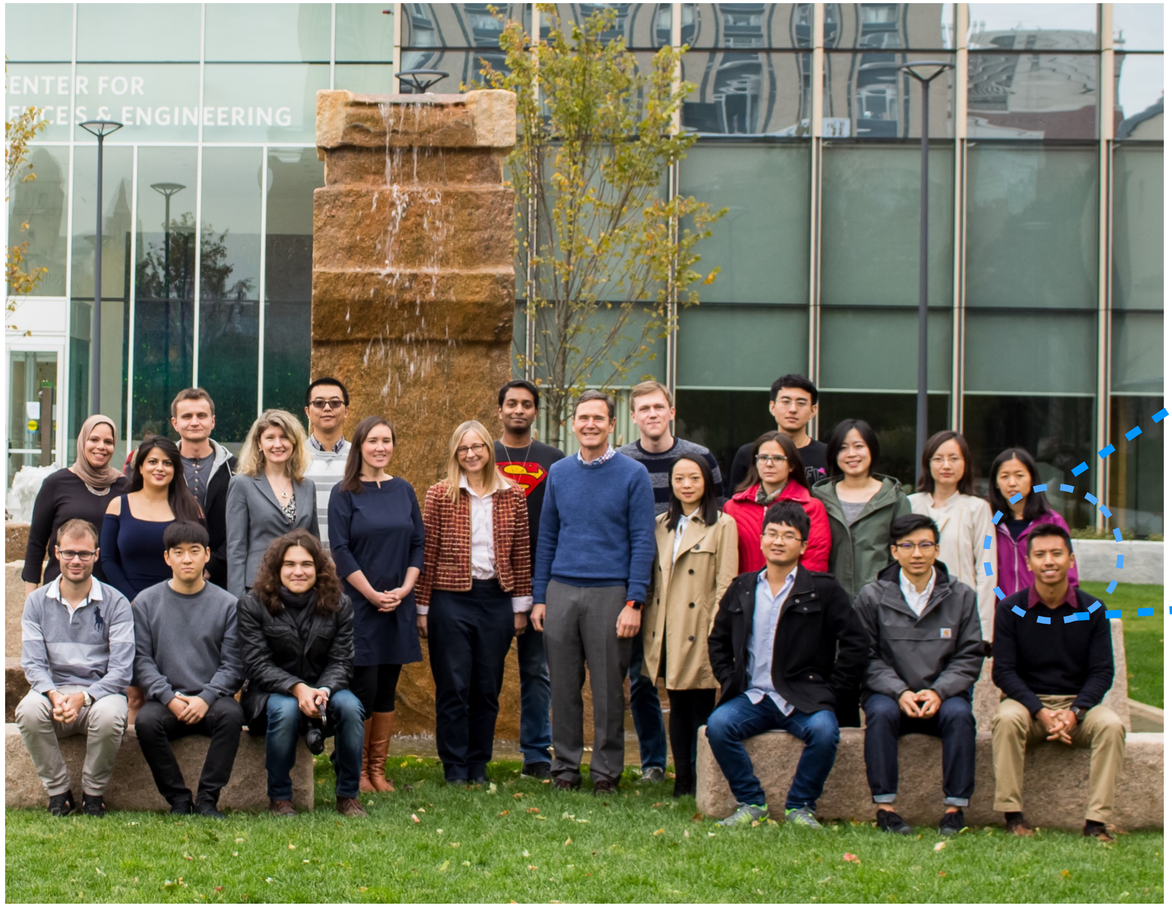
Computational Human Sensing: Applications of Face, Gesture and Affect Analysis

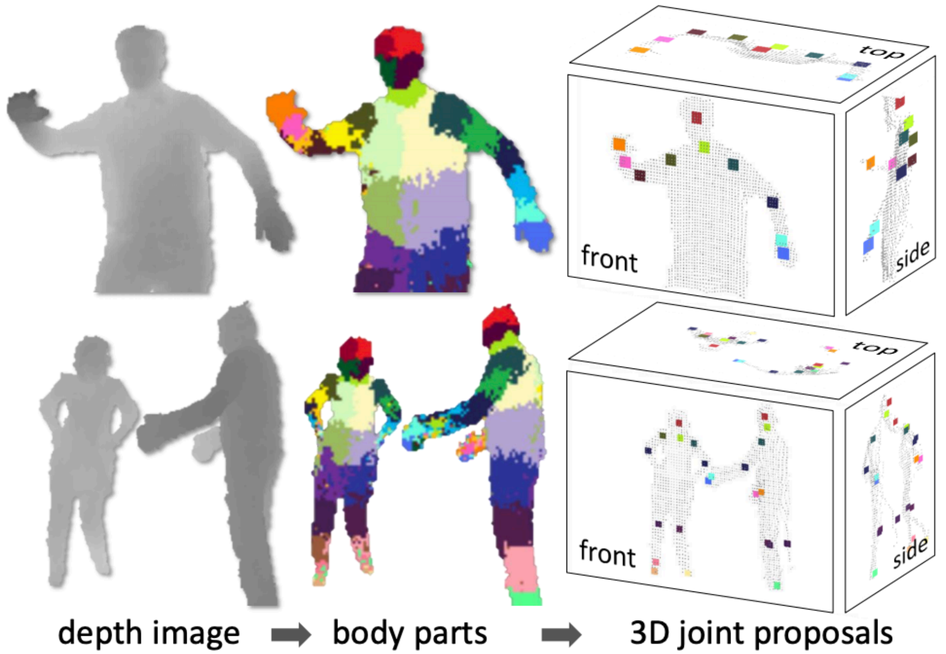
Ajjen Joshi
Affectiva

About Me



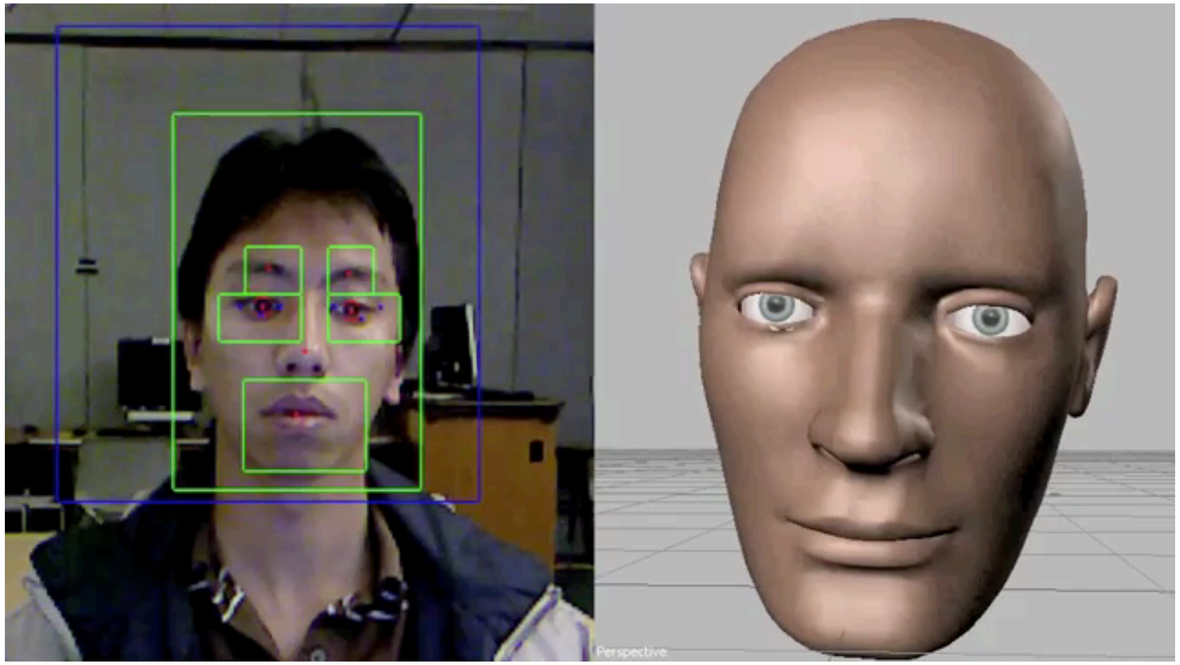
PhD Student: Boston University (2012-2018)
Research Scientist: Affectiva (2018 -)

The Kinect



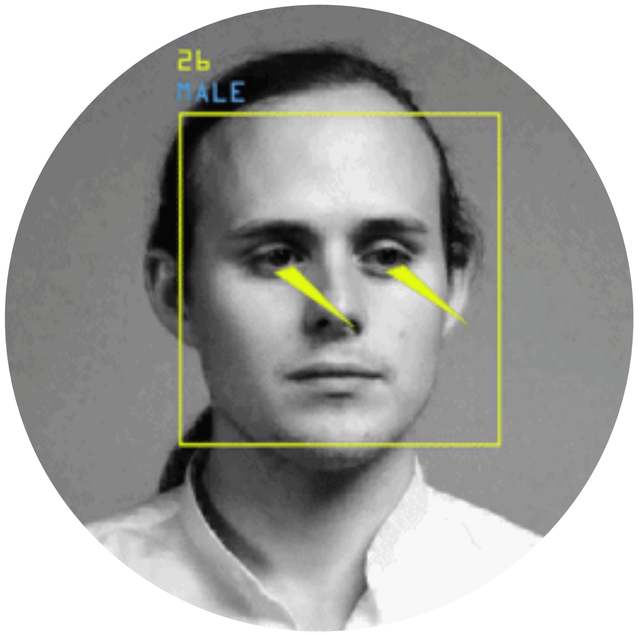
Building Interactive Installations with the Kinect
2011

Animojis



Real-time Facial Expression Imitation using a Depth Camera
2012

Human Signals



Eyegaze

Expressions

Gestures

Personalizing Gesture Recognition using Hierarchical Bayesian Neural Networks

Ajjen Joshi, Soumya Ghosh, Margrit Betke, Stan Sclaroff, Hanspeter Pfister

CVPR ‘17

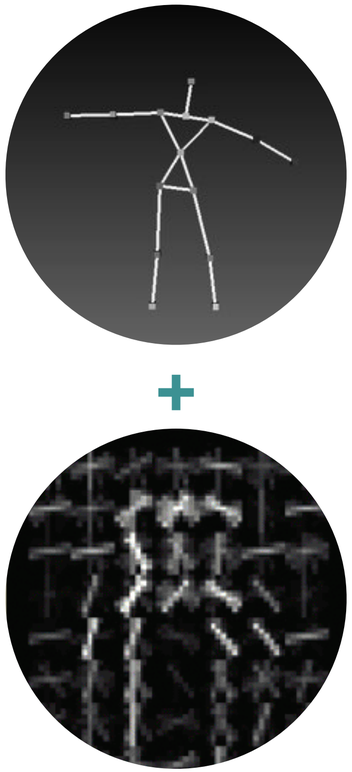
Gesture Recognition

Gestures

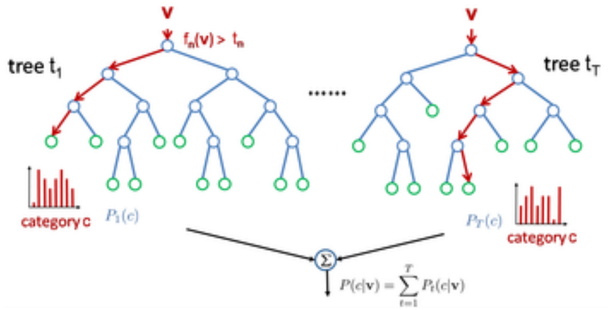
Gesture Recognition

Gesture Recognition

Input



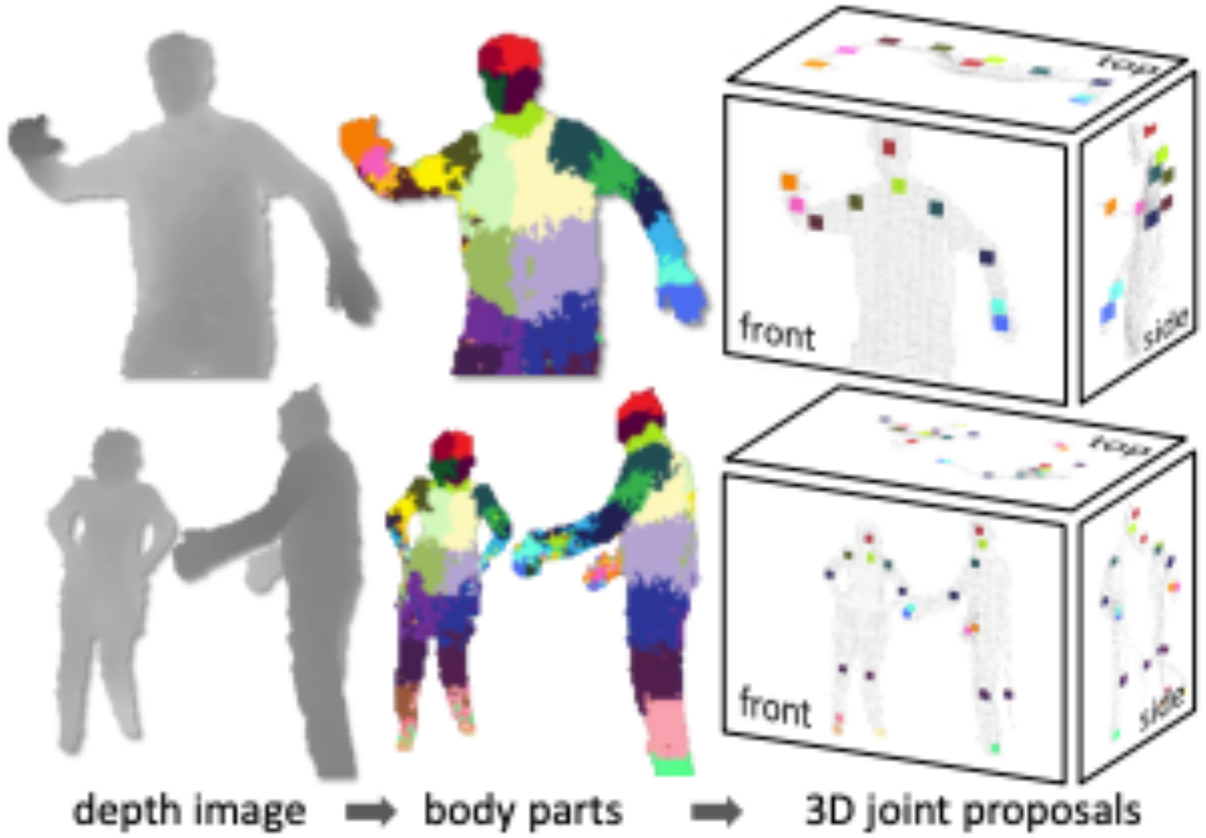
Features



Model

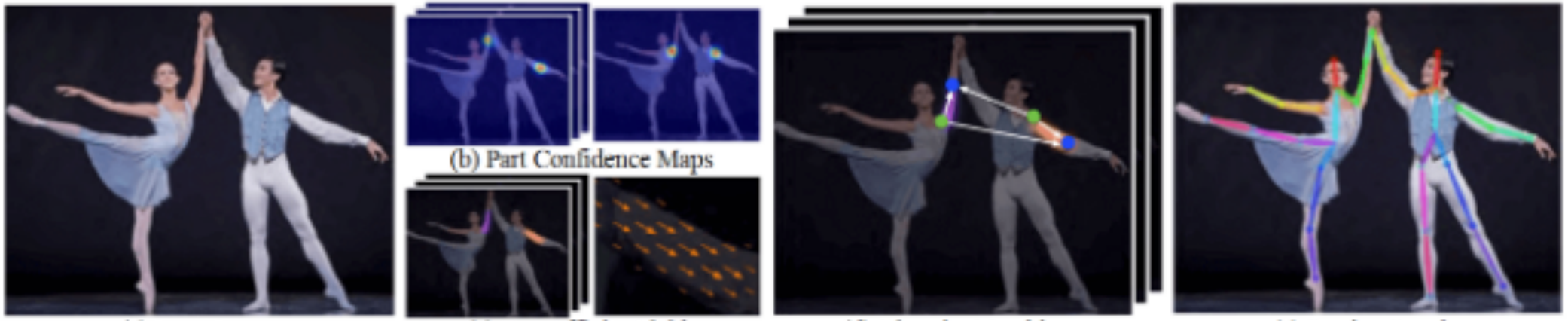
A Random Forest Approach to Segmenting and Classifying Gestures
FG ‘15

Intermediate Representation: Pose



Real-Time Human Pose Recognition in Parts from Single Depth Images
Shotton et al. CVPR ‘11

Intermediate Representation: Pose

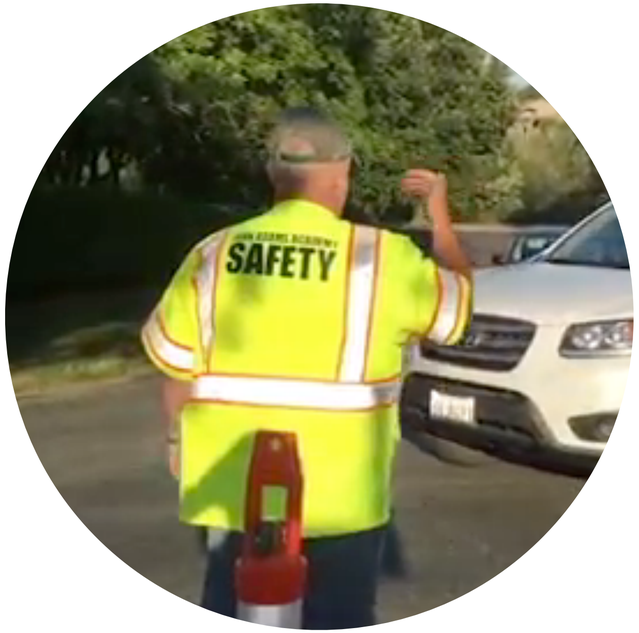


OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields
Cao et al. CVPR ‘17

Gesture Recognition: Disney Research

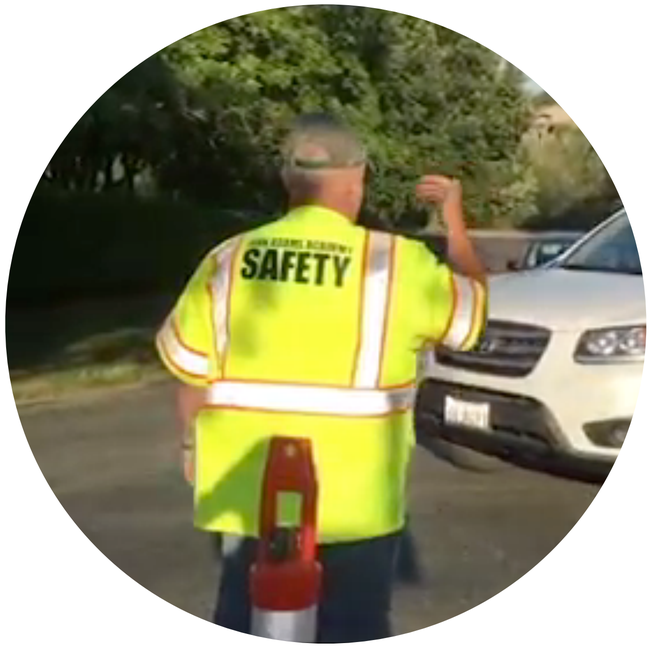
Gesture Recognition for Stormtroopers

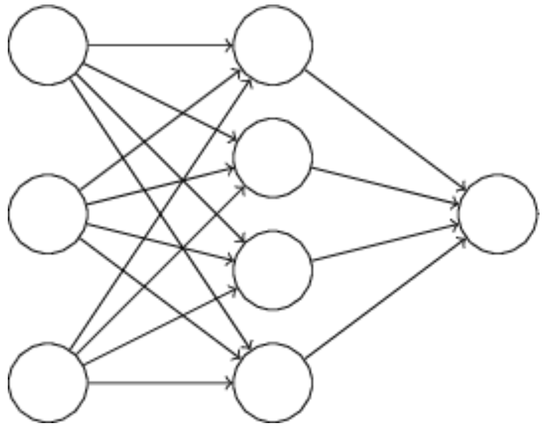
Personalization



Traffic Gestures Recognition System for Autonomous Vehicles

Personalization

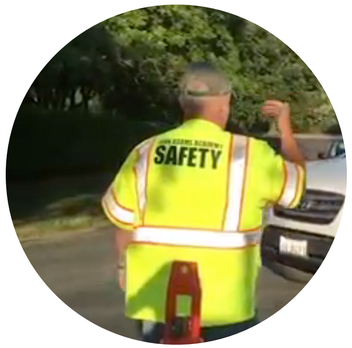


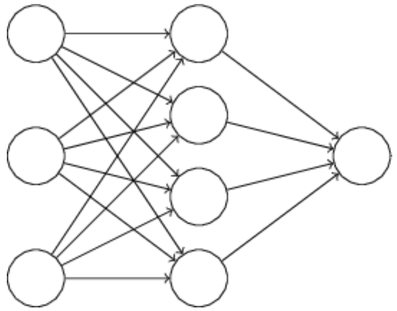


W

Traffic Gestures Recognition System for Autonomous Vehicles

Personalization

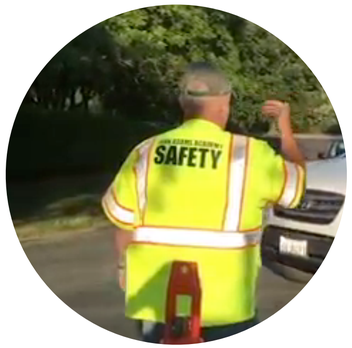


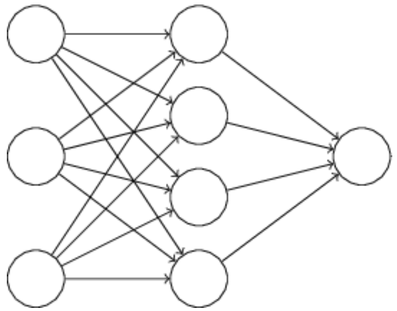


W

Traffic Gestures Recognition System for Autonomous Vehicles

Personalization

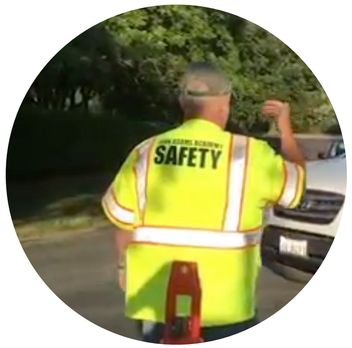


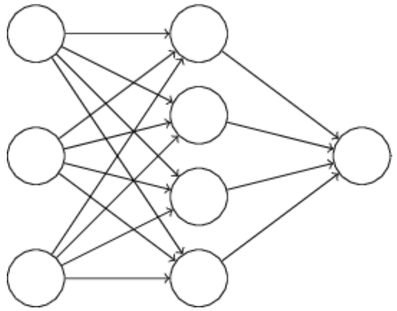


W

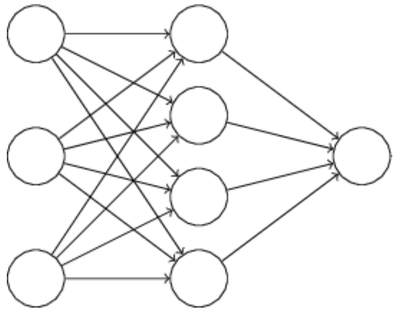
Traffic Gestures Recognition System for Autonomous Vehicles

Personalization





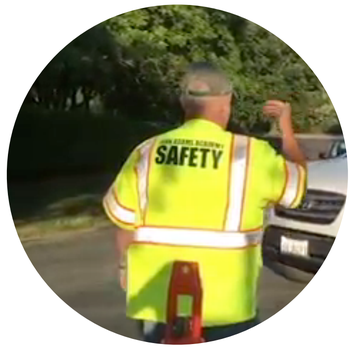
W_1

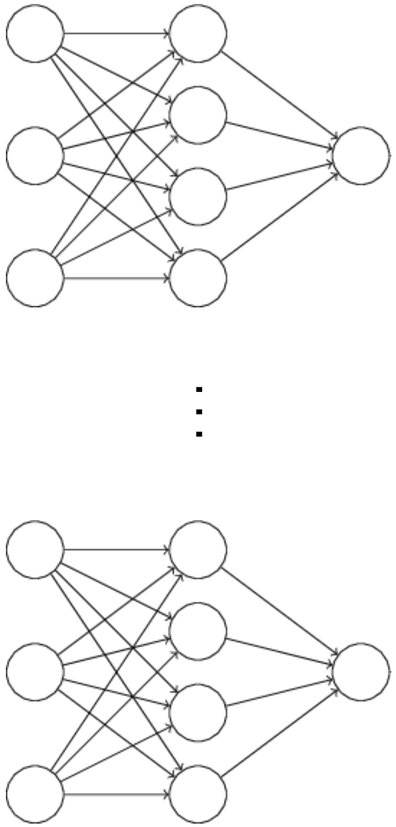


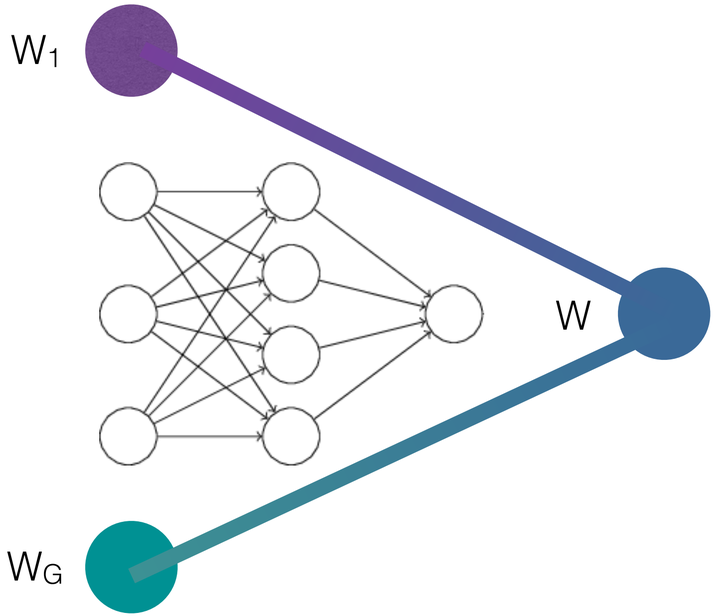
W_G

Traffic Gestures Recognition System for Autonomous Vehicles

Personalization using Hierarchical Models

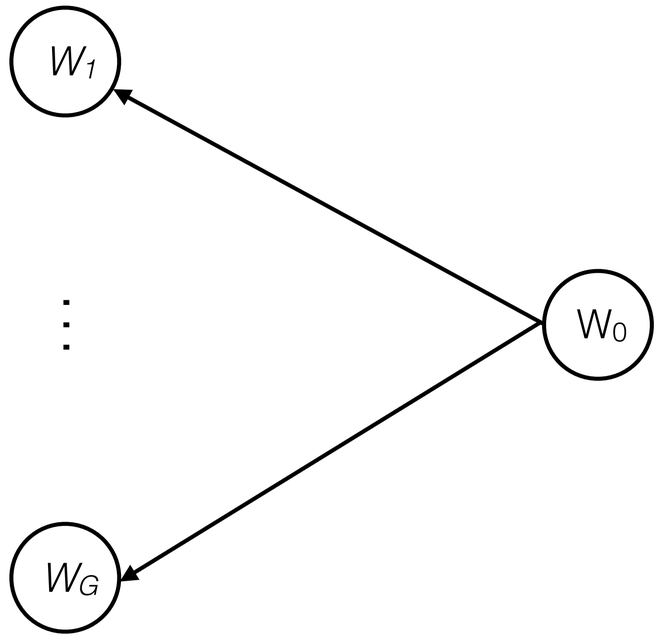






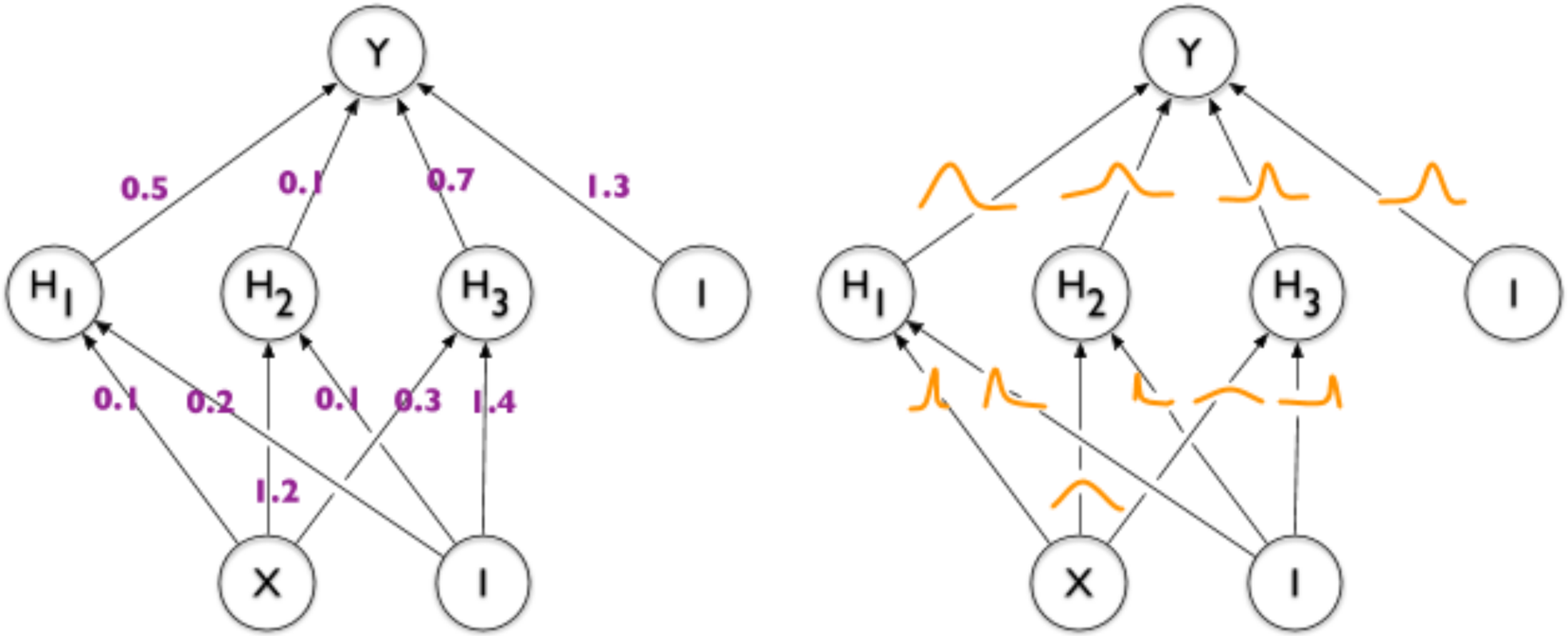
Traffic Gestures Recognition System for Autonomous Vehicles

Hierarchical Model



Hierarchical Bayesian Model

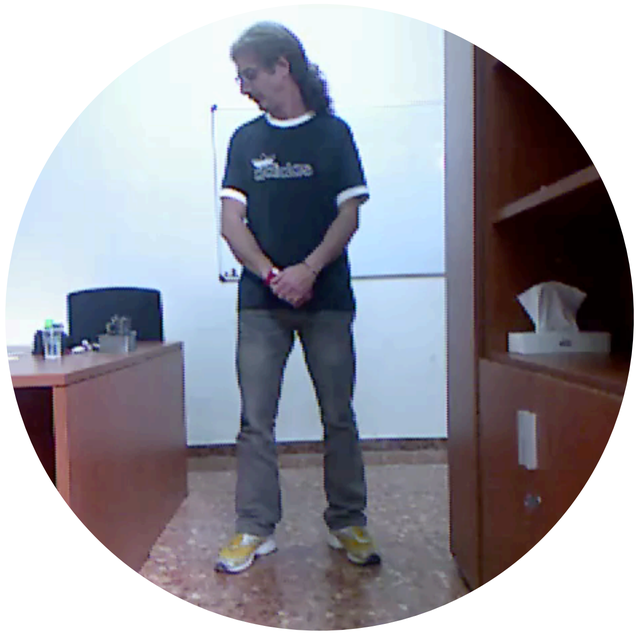
Bayesian Neural Networks



Bayesian Neural Networks

Datasets

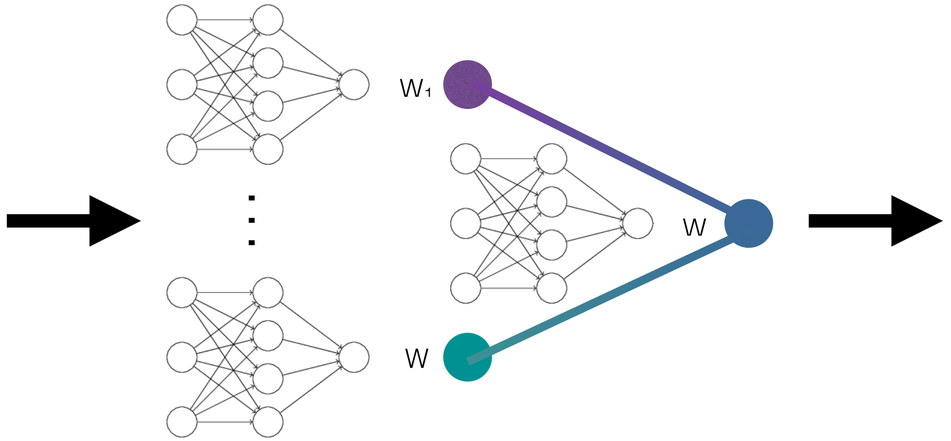
MSRC-12
12 gestures, 30 subjects
~6000 gesture instances



ChaLearn
20 gestures, 36 subjects
~11000 gesture instances

NATOPS
24 gestures, 20 subjects
~9600 gesture instances

Experiments



Gesture Label

Input:
600-d joint/appearance-based feature representation

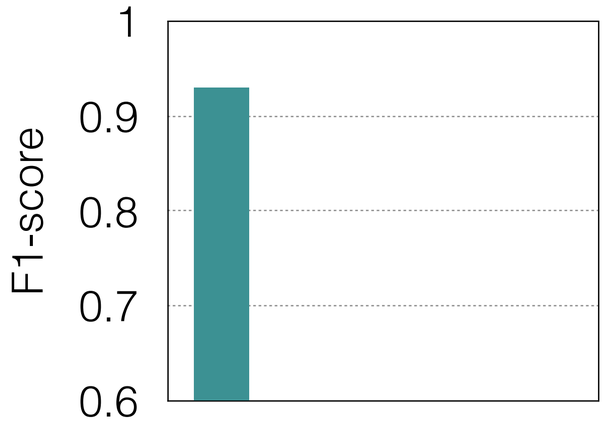
HBNN with 2 HL, each with 400 activation nodes, trained with RMSprop

Output:
Gesture classes

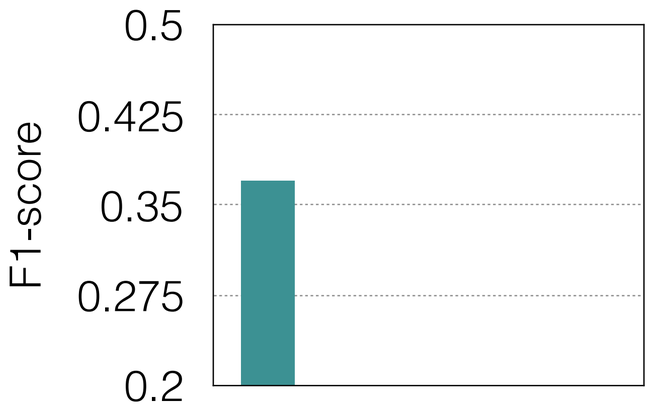
Train/Test: Random 75-25 split

Experiments

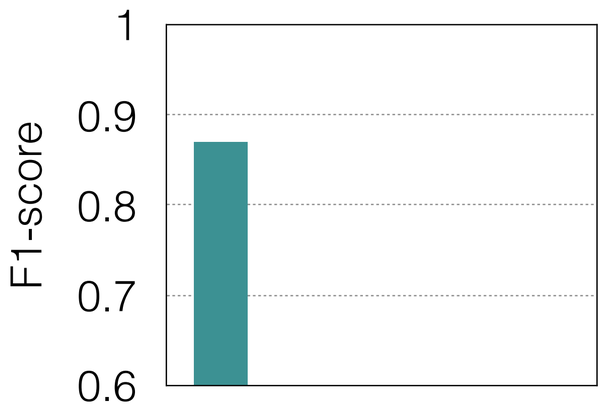
MSRC-12



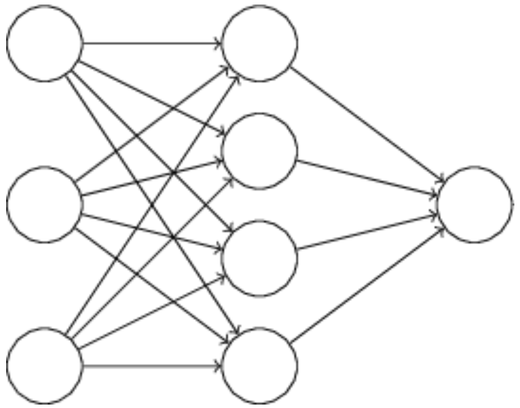
ChaLearn



NATOPS



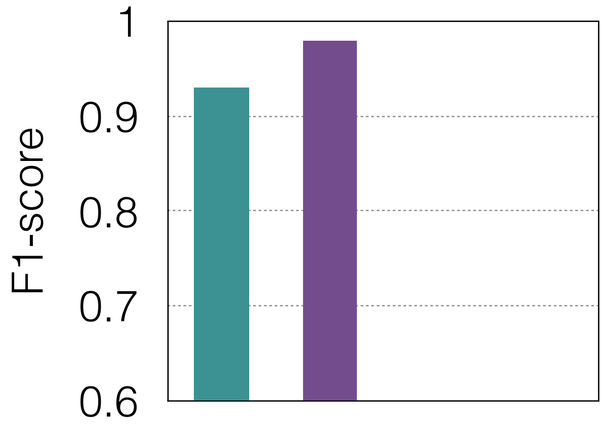
Bayesian Neural Network
Pooled



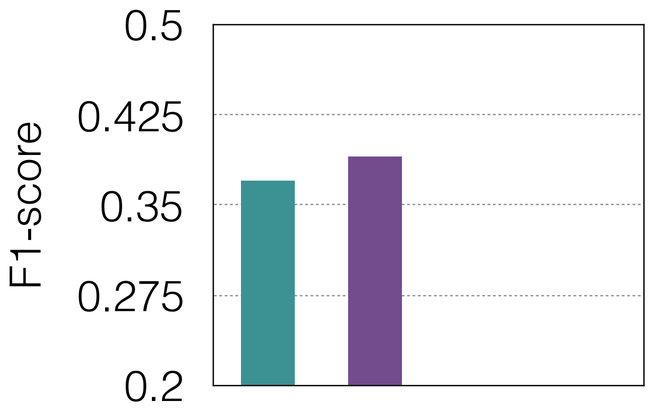
W

Experiments

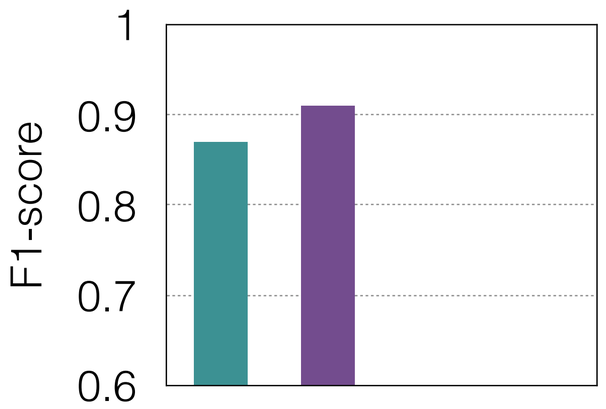
MSRC-12



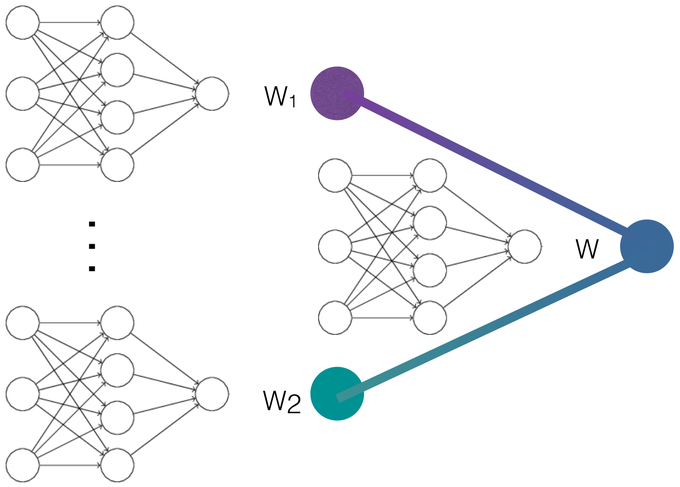
ChaLearn



NATOPS

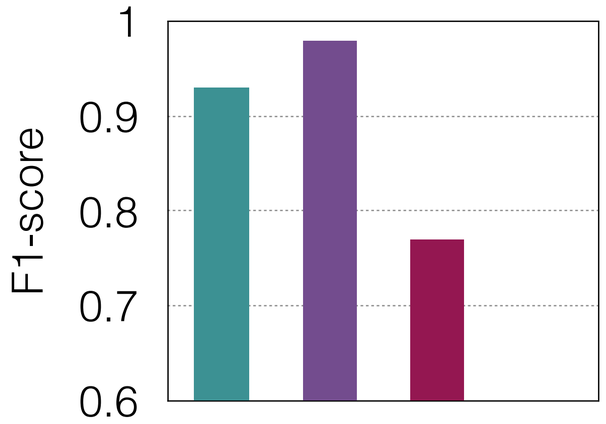


Hierarchical Bayesian Neural Network
Subject-specific

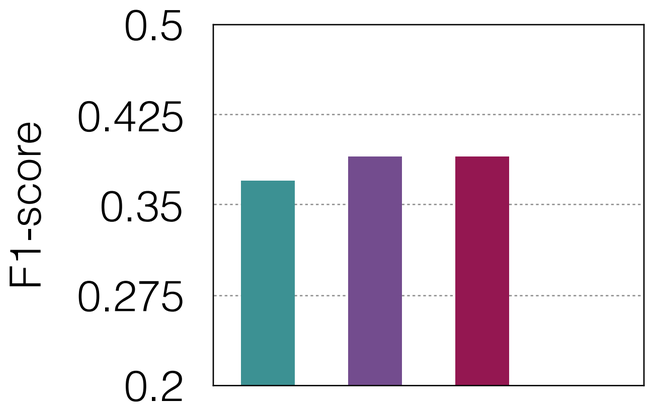


Experiments

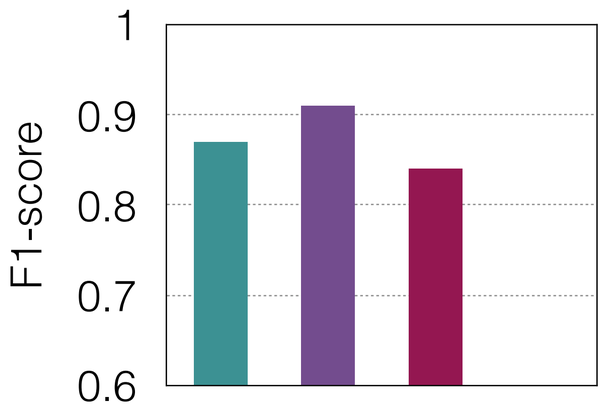
MSRC-12



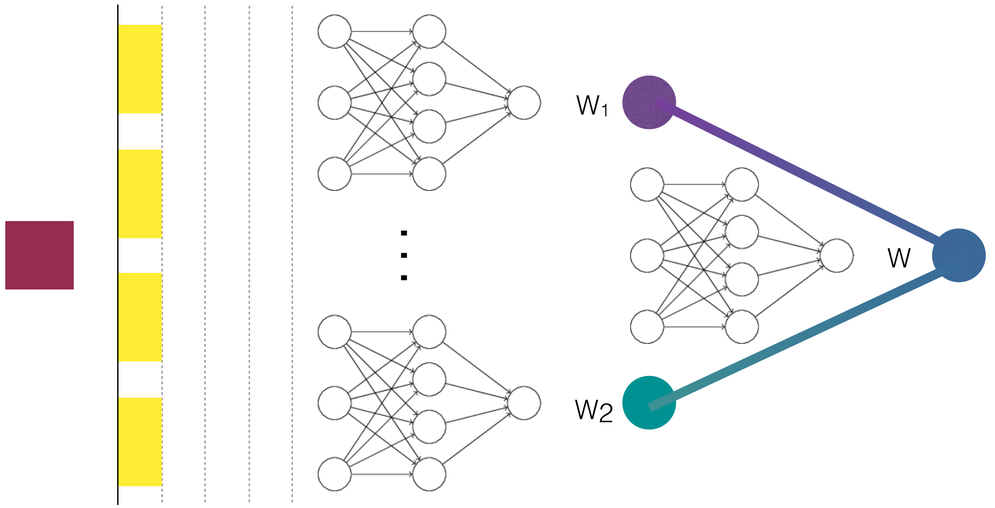
ChaLearn



NATOPS

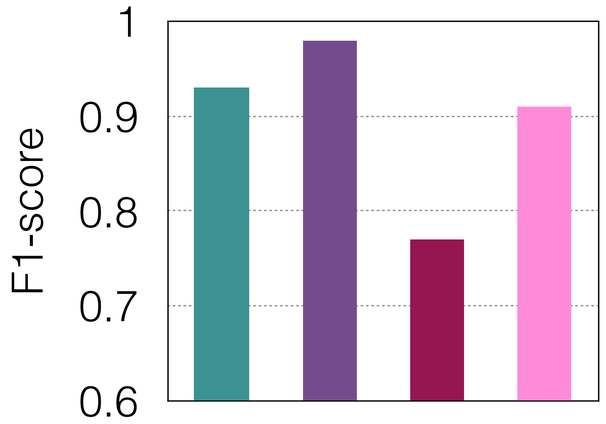


Hierarchical Bayesian Neural Network
Naive Bayesian Model Averaging

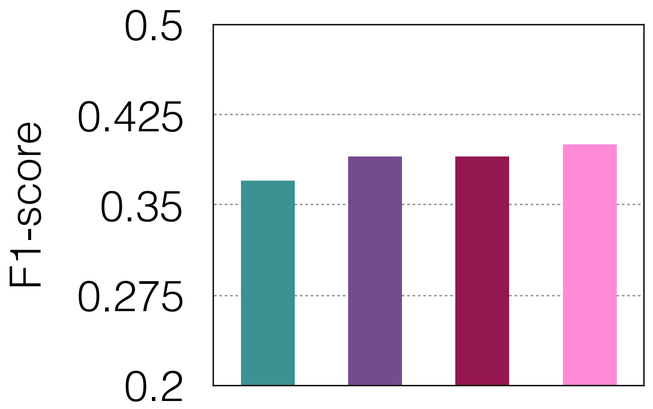


Experiments

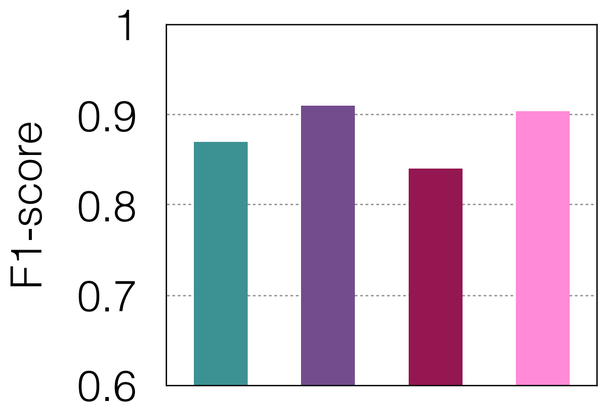
MSRC-12



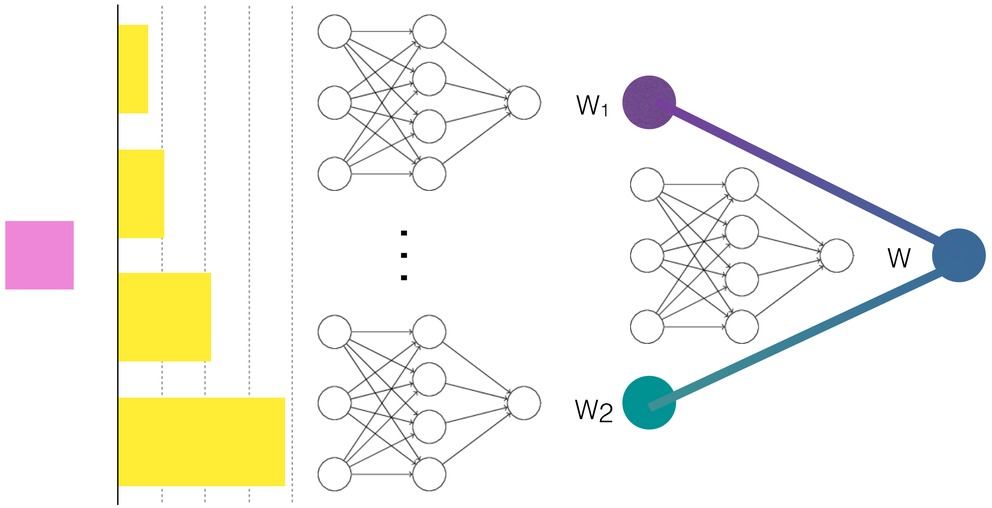
ChaLearn



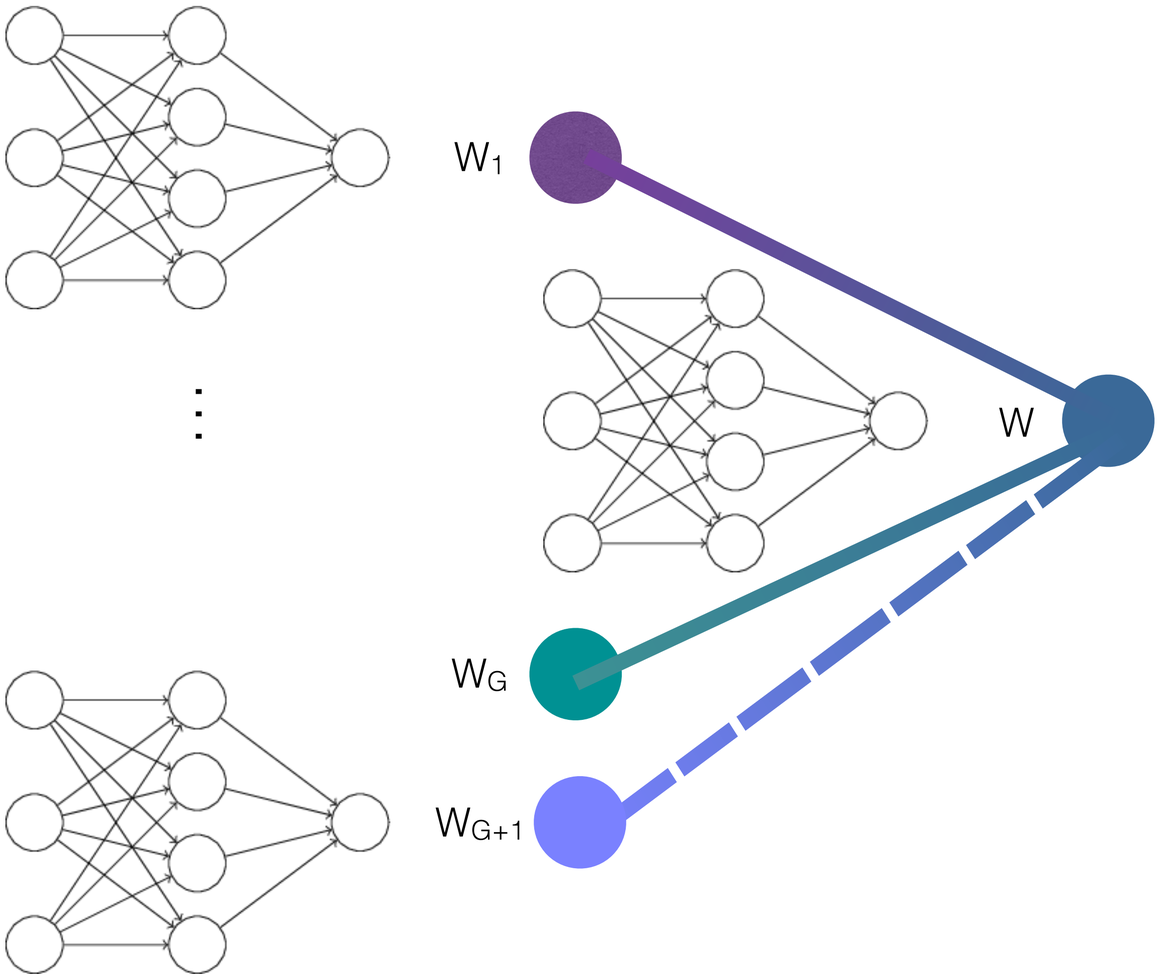
NATOPS



Hierarchical Bayesian Neural Network
Weighted Bayesian Model Averaging

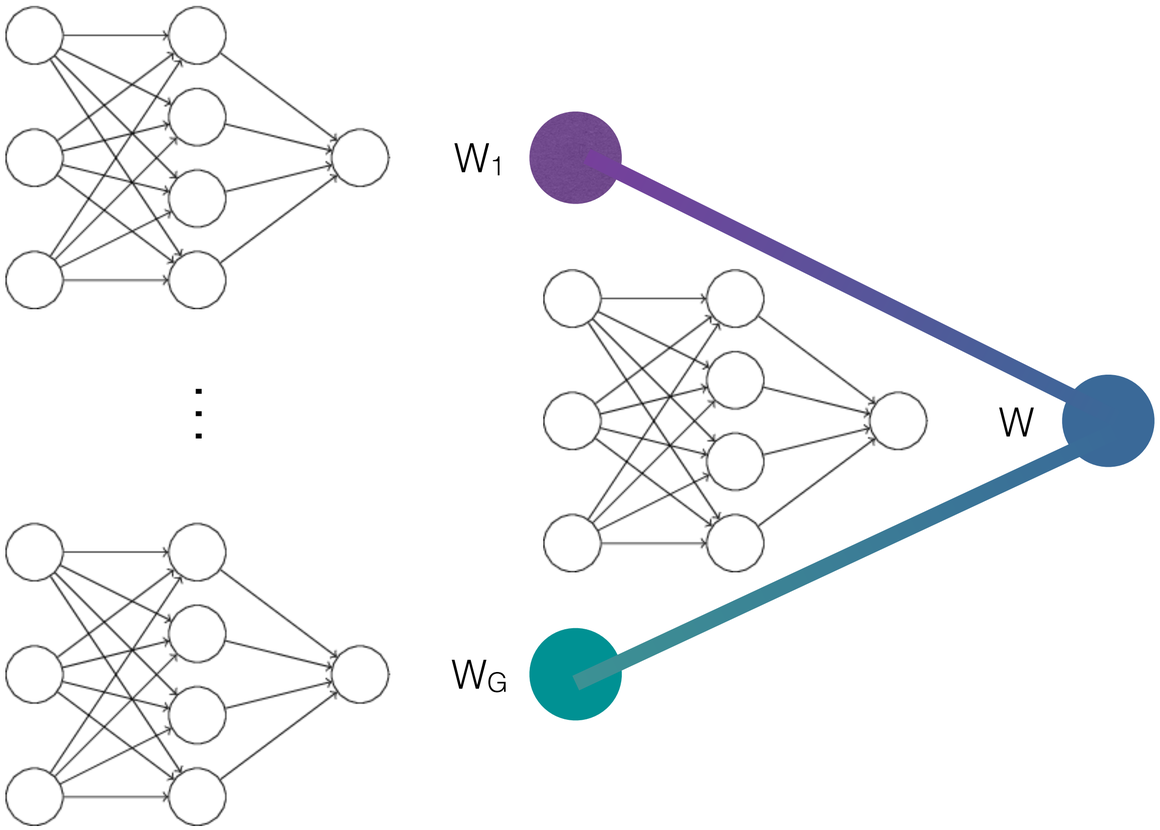


Personalization



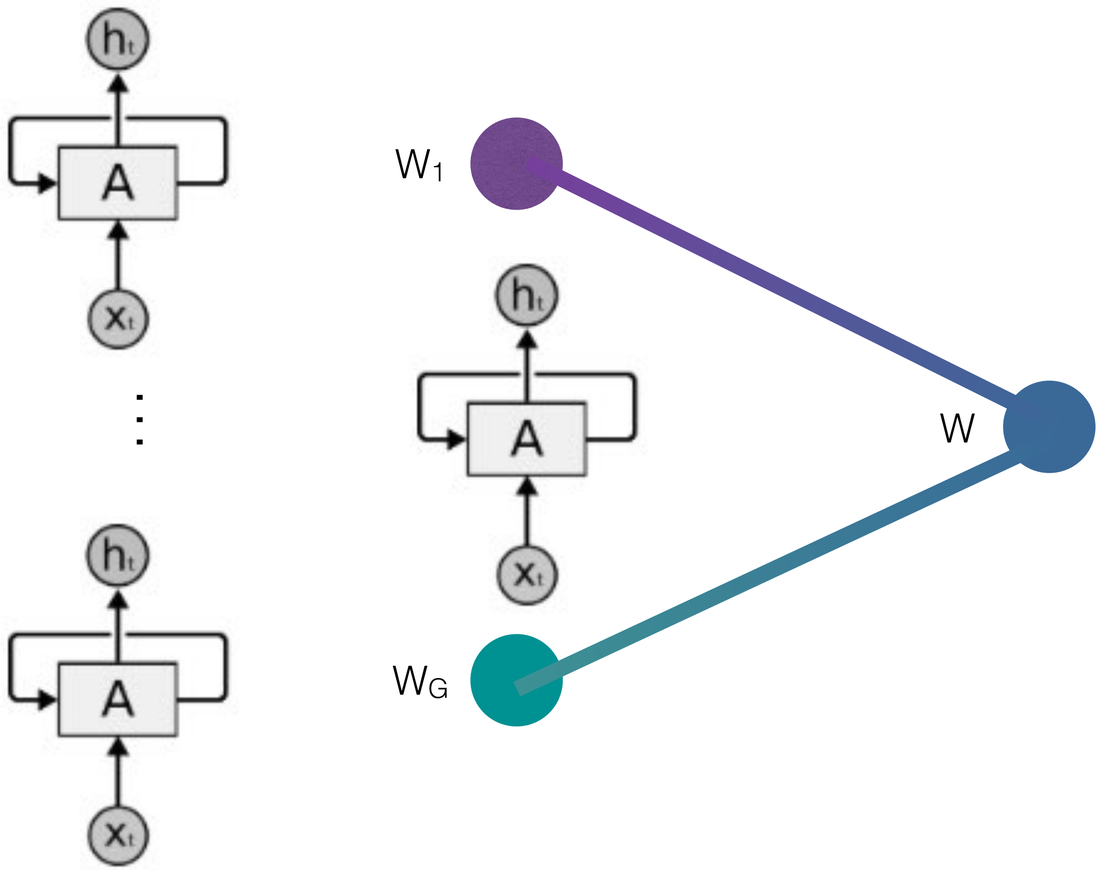
Personalizing model to new subject

Hierarchical Bayesian Neural Networks



What is an appropriate architecture to model temporal signals?

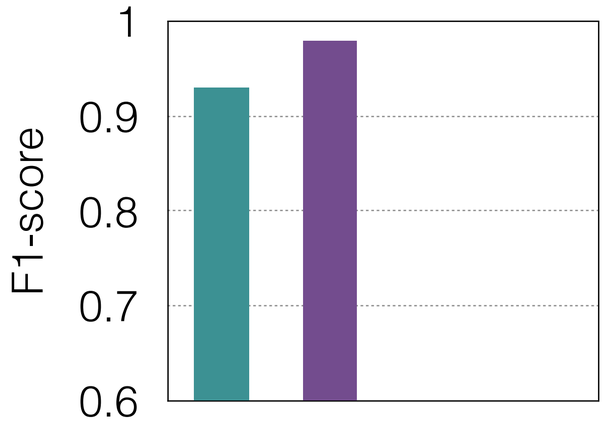
Hierarchical Bayesian Recurrent Neural Networks



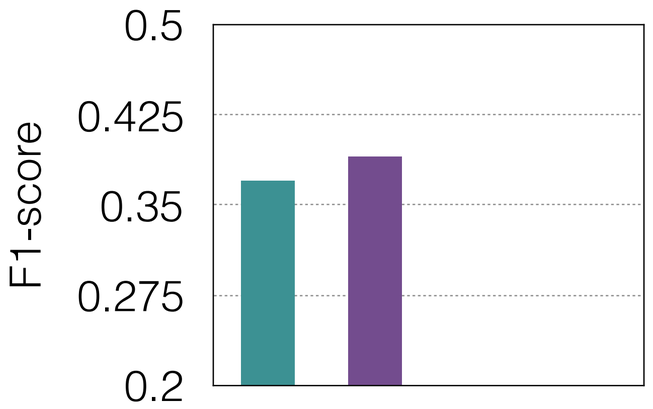
Recurrent Neural Networks

Experiments

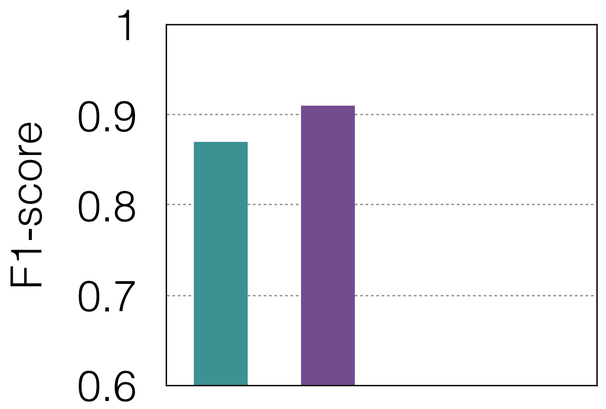
MSRC-12



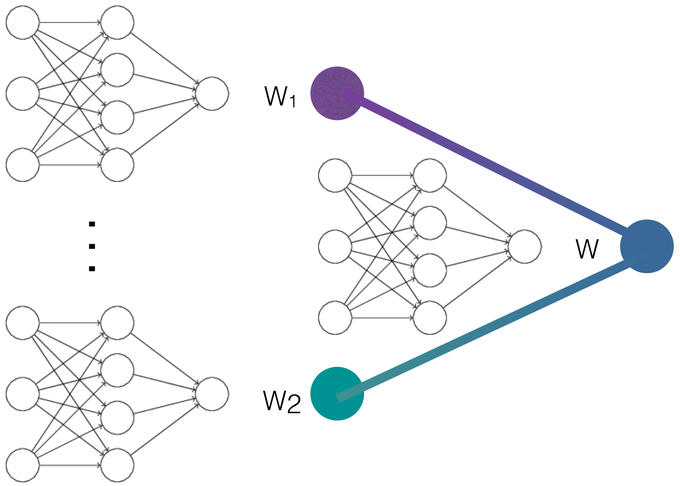
ChaLearn



NATOPS

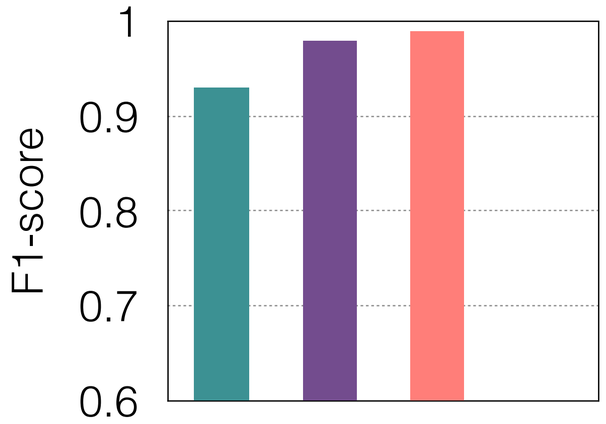


Hierarchical Bayesian Neural Network
Subject-specific

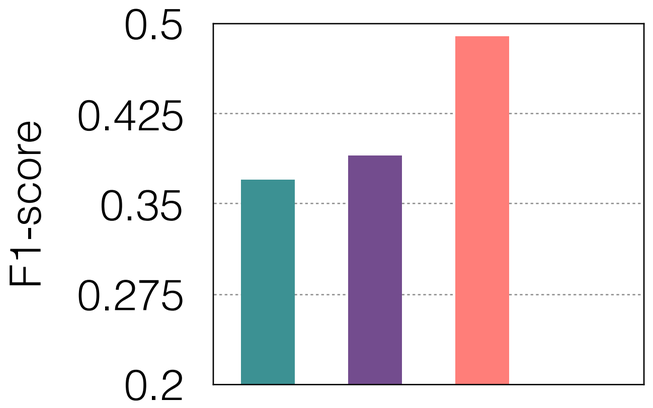


Experiments

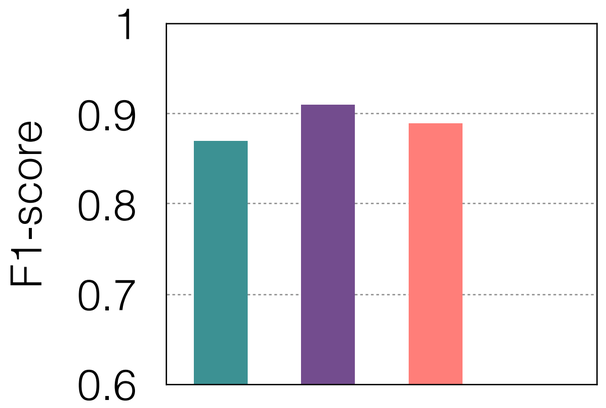
MSRC-12



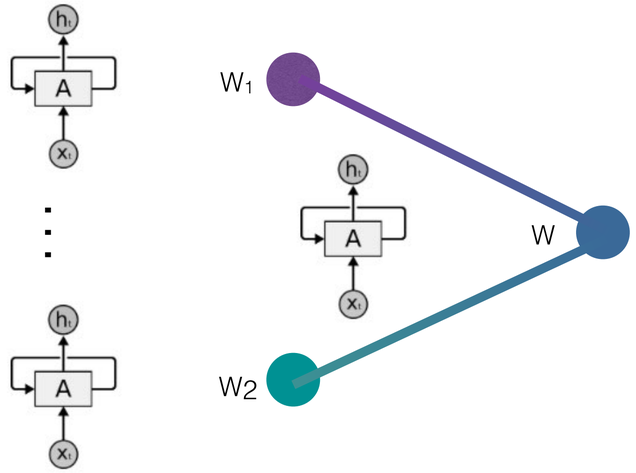
ChaLearn

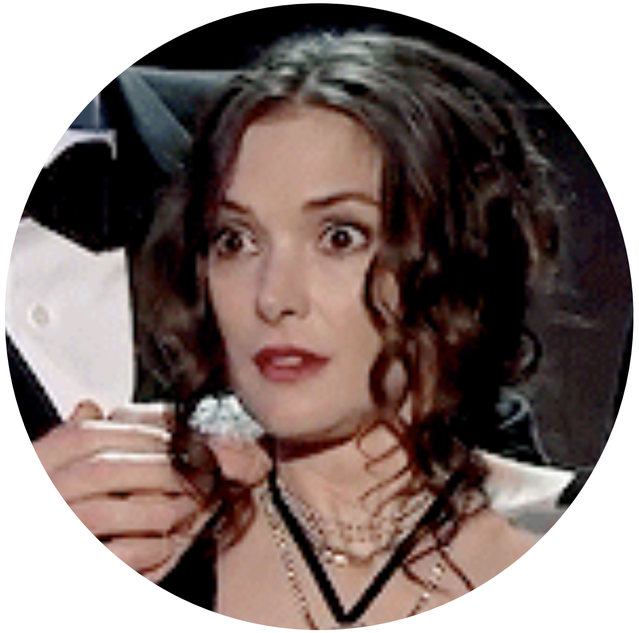


NATOPS



Hierarchical Bayesian Recurrent Neural Network
Subject-specific





Context-Sensitive Prediction of Facial Expressivity Using Multimodal Hierarchical Bayesian Neural Networks

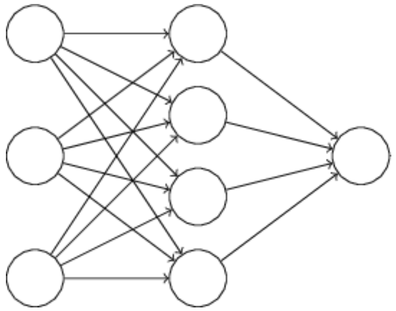
Ajjen Joshi, Soumya Ghosh, Sarah Gunnery, Linda Tickle-Degnen, Stan Sclaroff, Margrit Betke

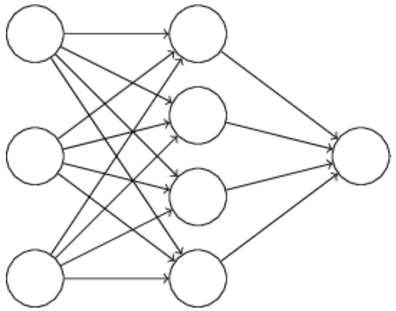
FG ‘18

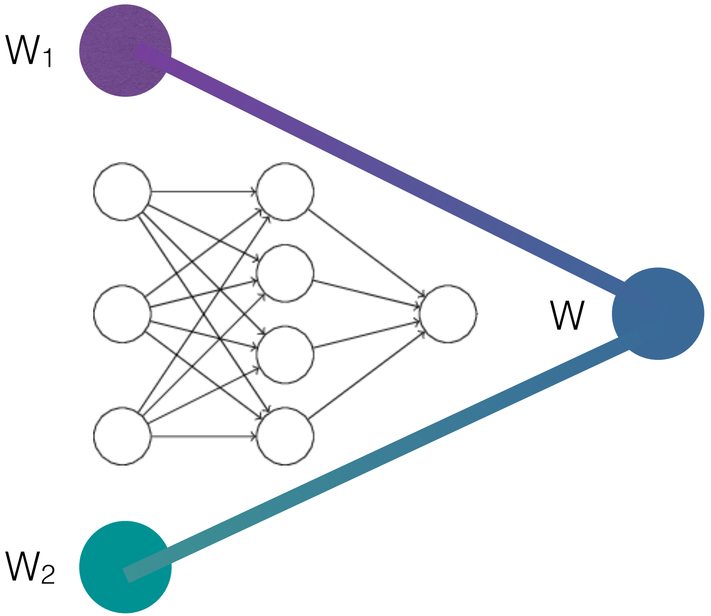
Facial Expressivity Prediction

Expressivity

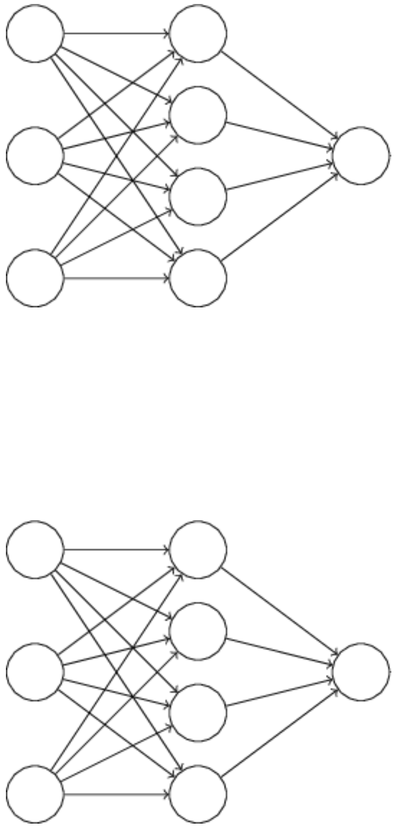
Gender-specific Facial Expressivity Prediction







Context-sensitive Facial Expressivity Prediction



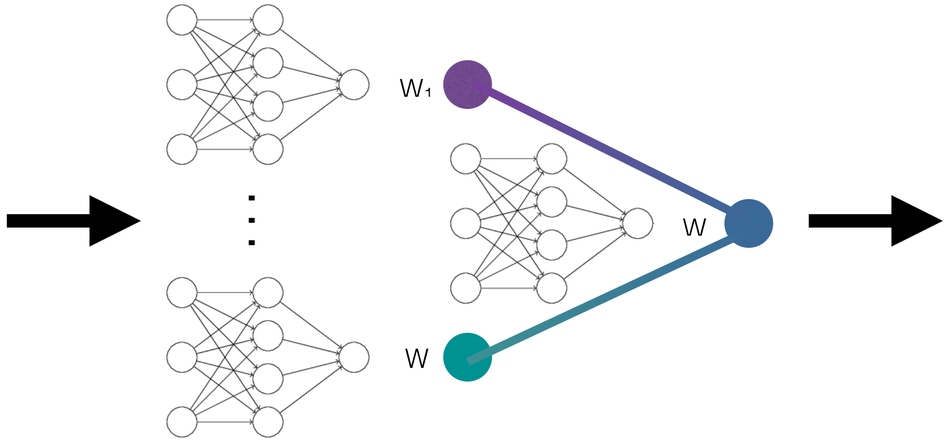
Datasets

802 video interviews of PD patients with occupational therapists

117 participants

Expertly-coded Likert-scale ratings of Active Expressivity in the Face

Experiments



Facial Expressivity Level

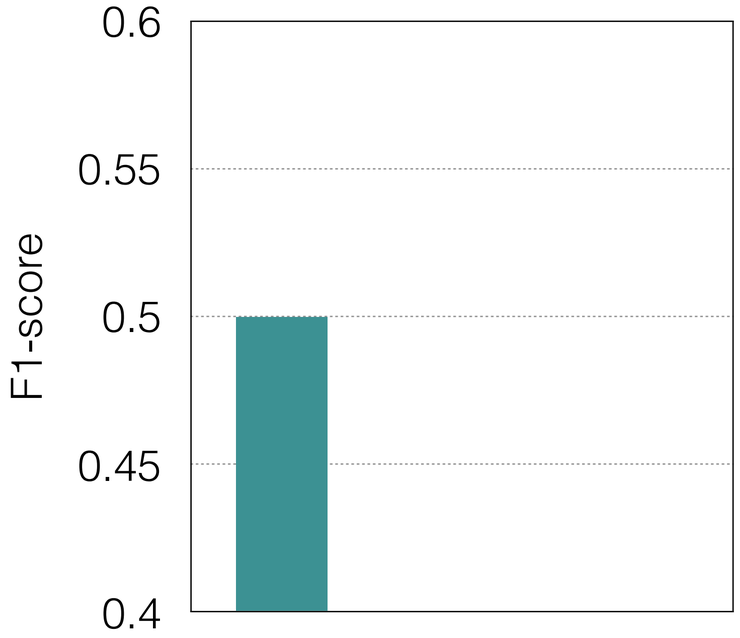
Input:
300-d multimodal (audio + video) feature representation

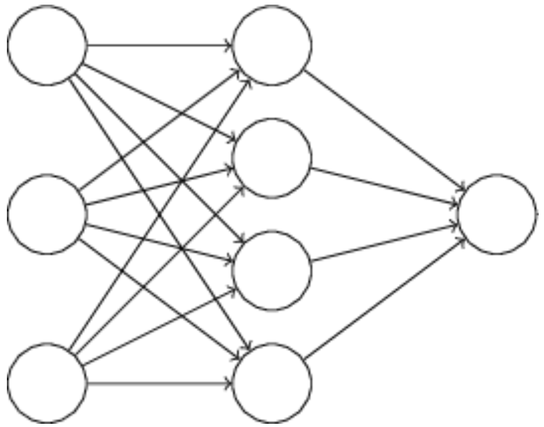
HBNN with 2 HL, each with 400 activation nodes, trained with RMSprop

Output:
Facial Expressivity class

Train/Test: 9-fold cross-validation

Experiments

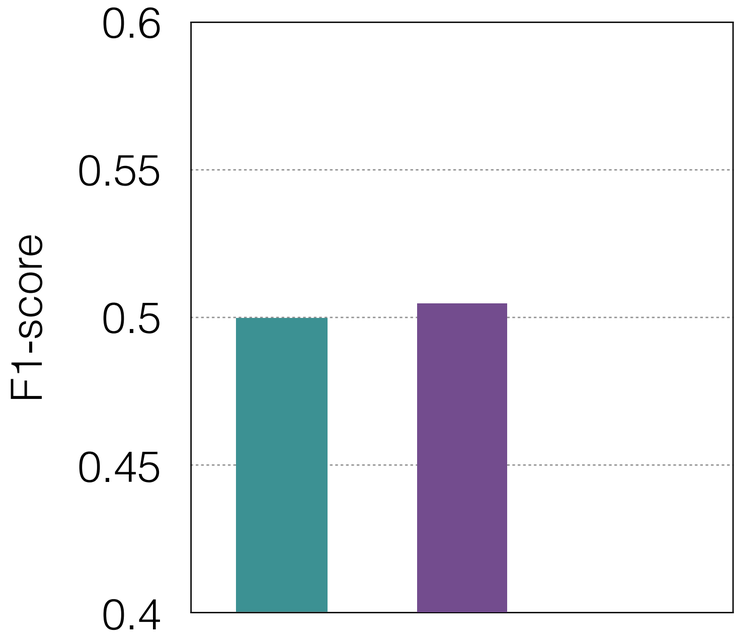


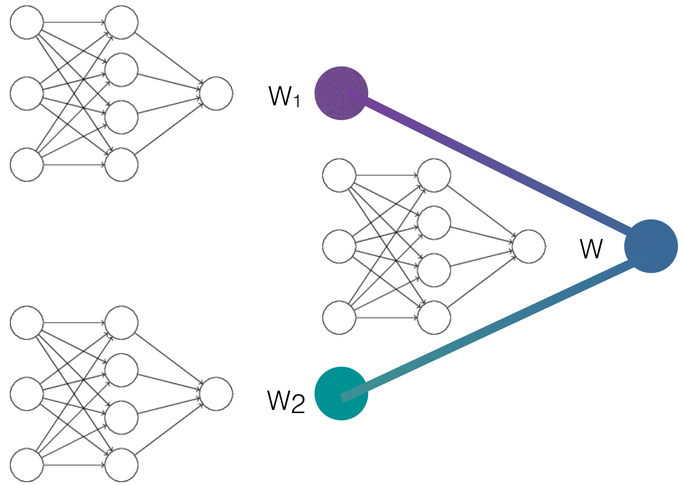


W

Bayesian Neural Network
Pooled

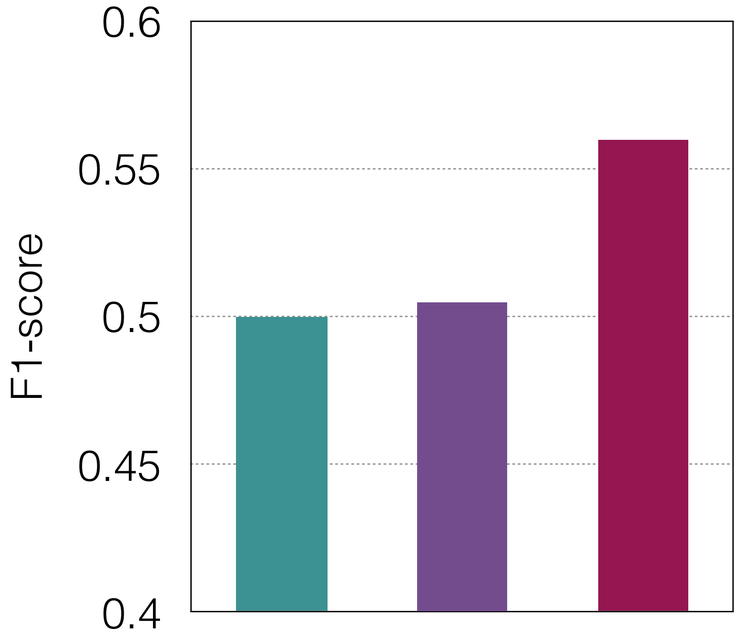
Experiments

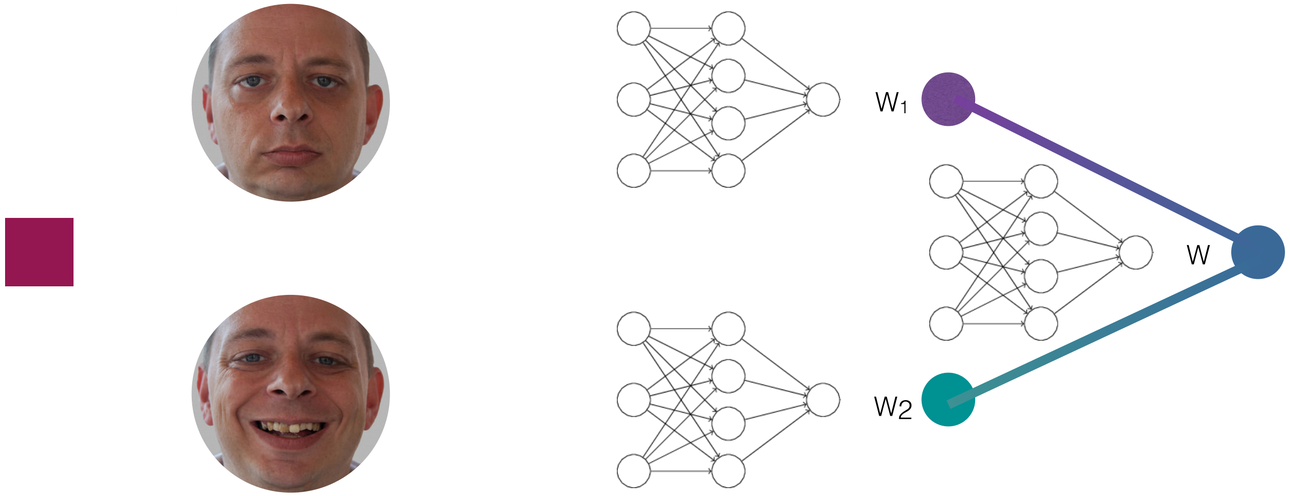




Hierarchical Bayesian Neural Network
Gender

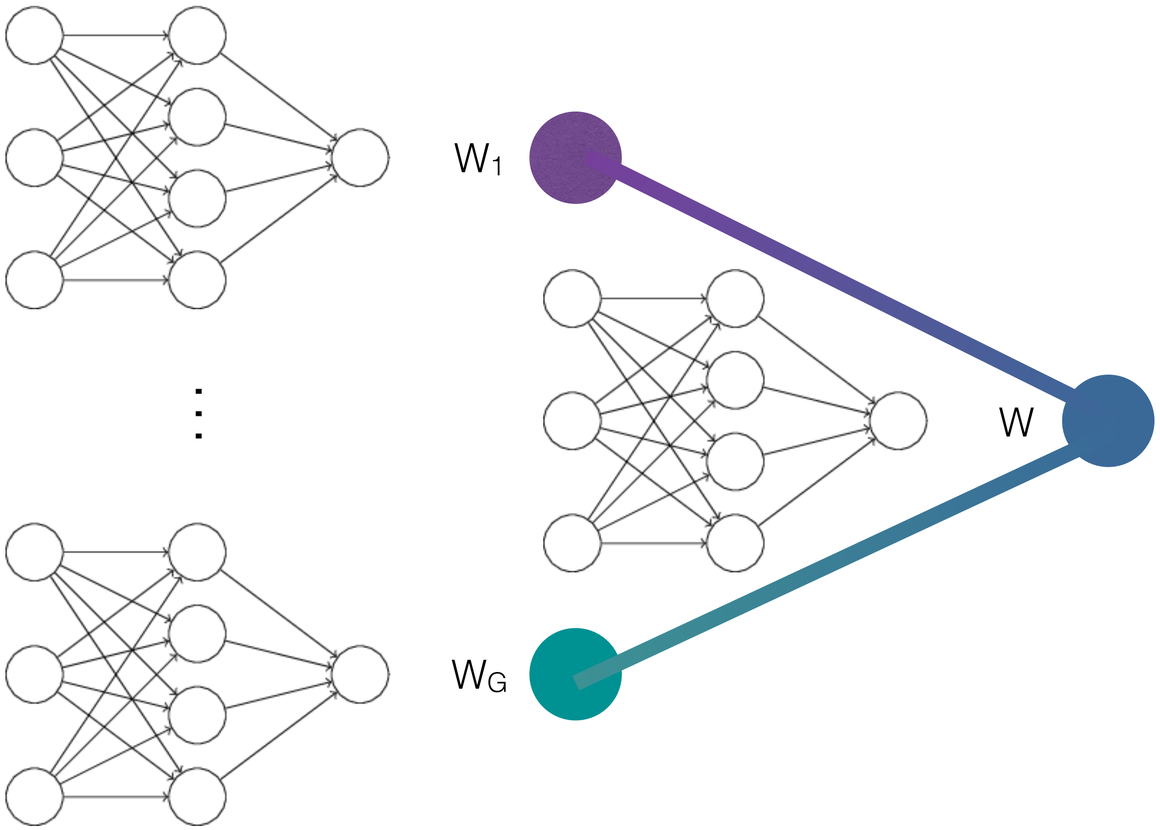
Experiments



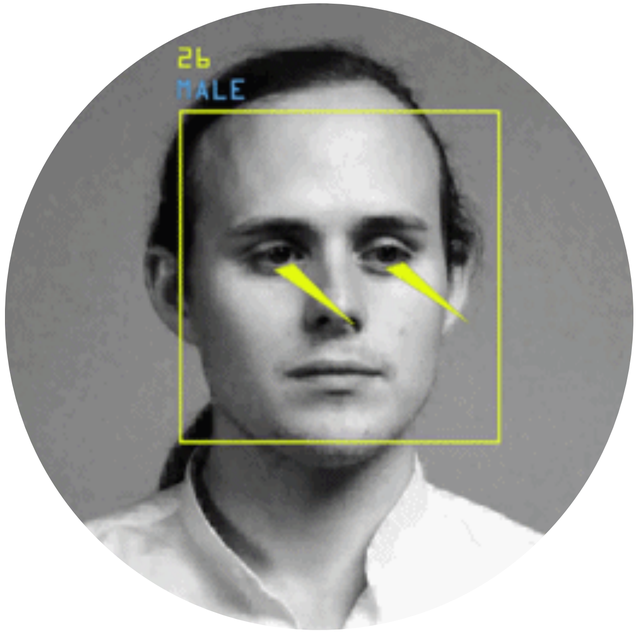


Hierarchical Bayesian Neural Network
Sentiment

Conclusion



Embrace subject and group-specific variances

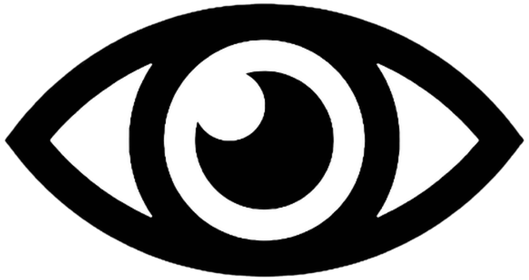


Eyeswipe: Towards Fast and Comfortable Text Entry using Gaze Paths

Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, Margrit Betke

CHI ’16

Handsfree Text Entry

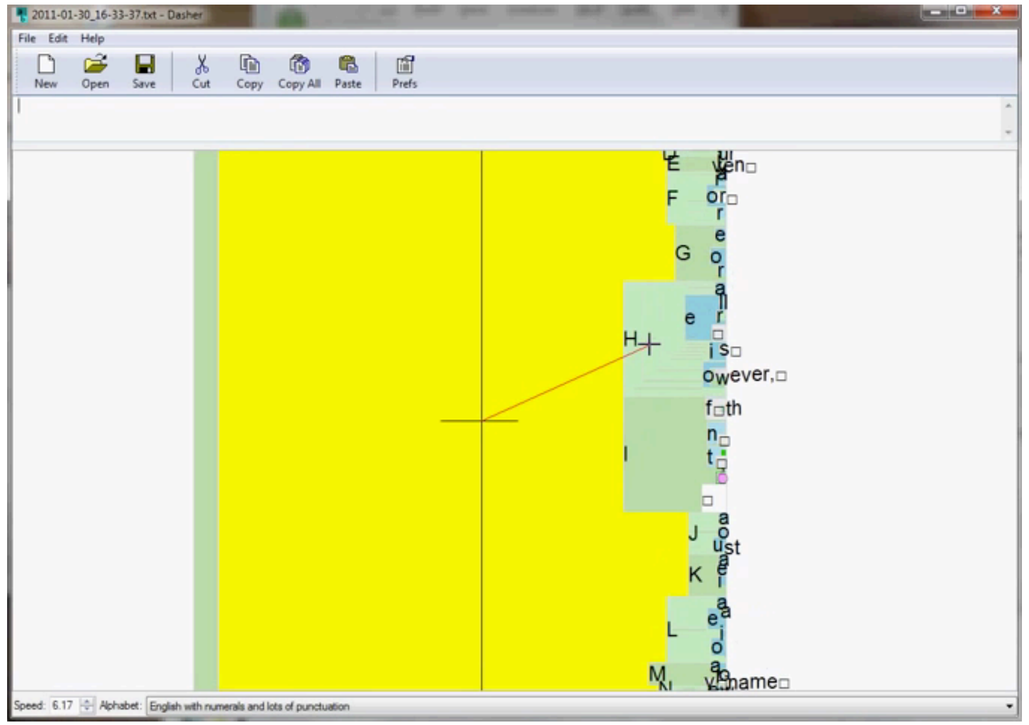


Accessibility

VR/AR

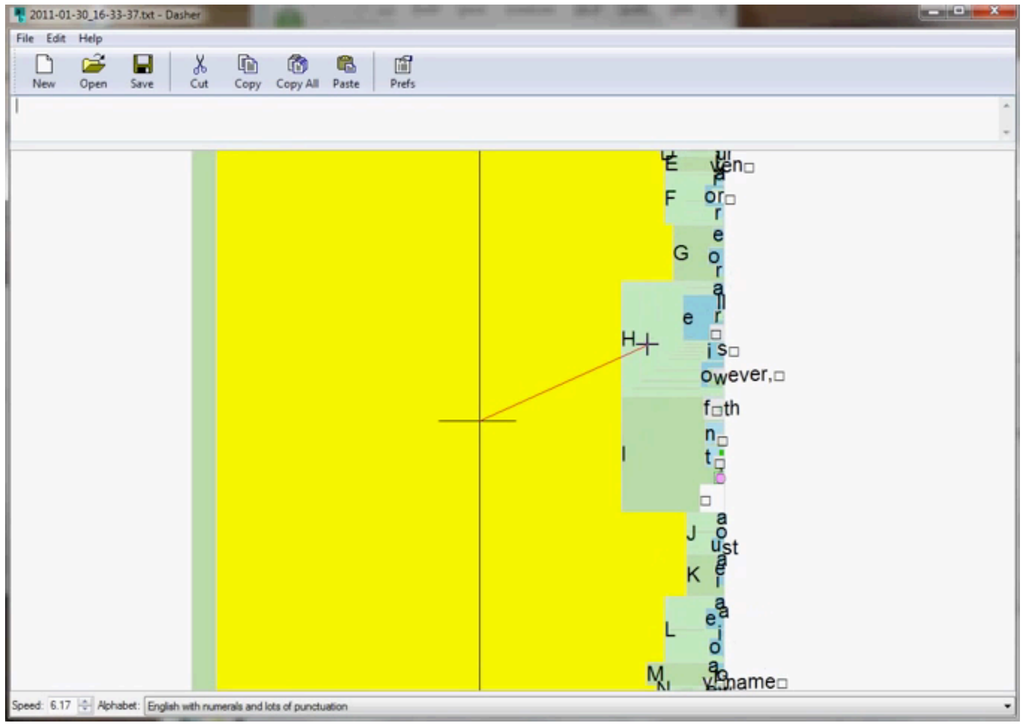
Text Entry using Gaze

Dwell-time based
Majaranta et al. (2009)



Dasher
Ward and MacKay (2002)

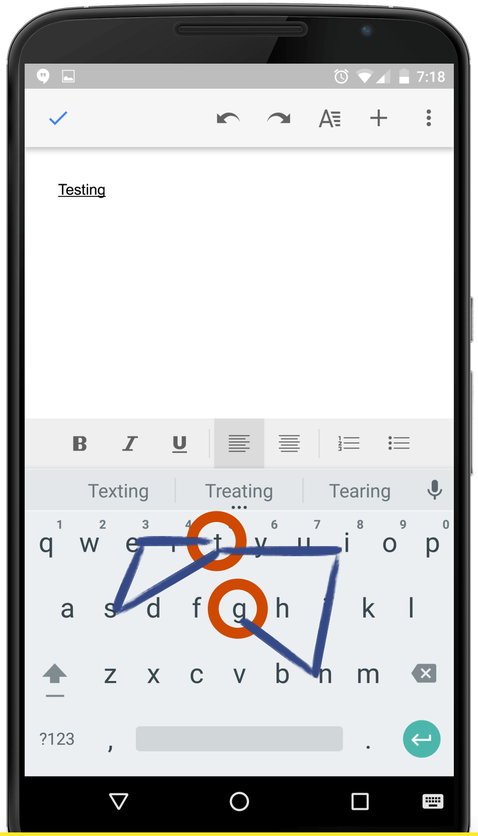
Text Entry using Gaze



Intuitive interface
Slow

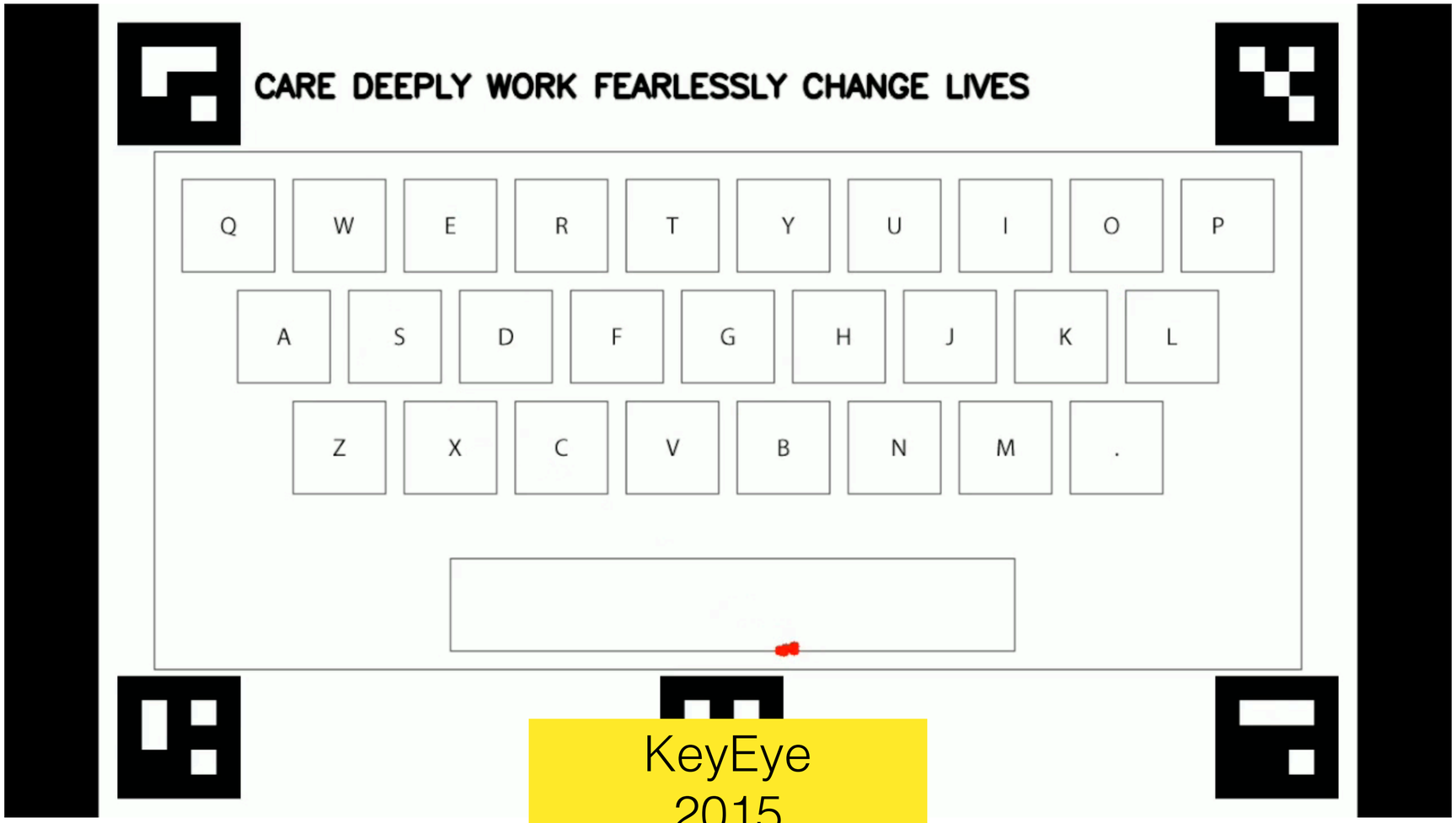
Complex interface
Fast

Inspiration



Swipe-based Keyboard
Zhai and Kristensson (2003)

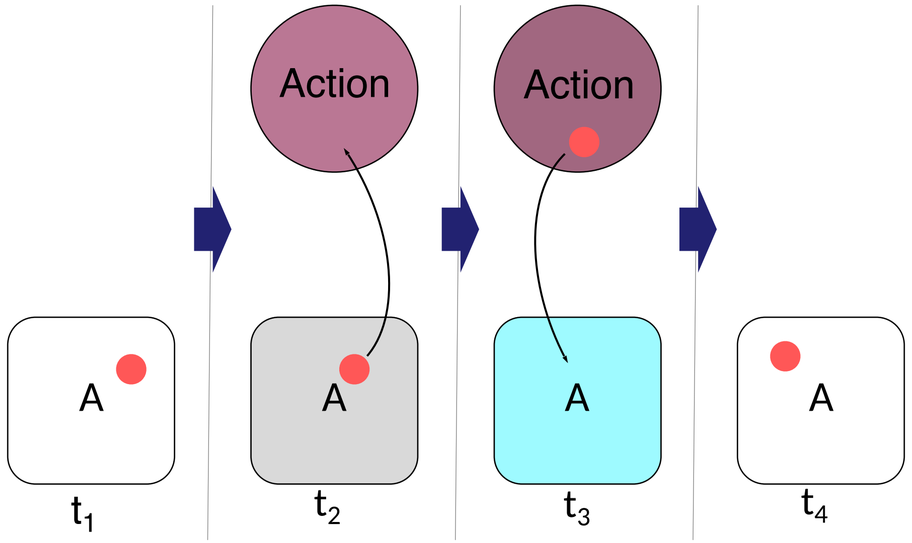
Prototype


CARE DEEPLY WORK FEARLESSLY CHANGE LIVES
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M .
KeyEye
2015

Challenge

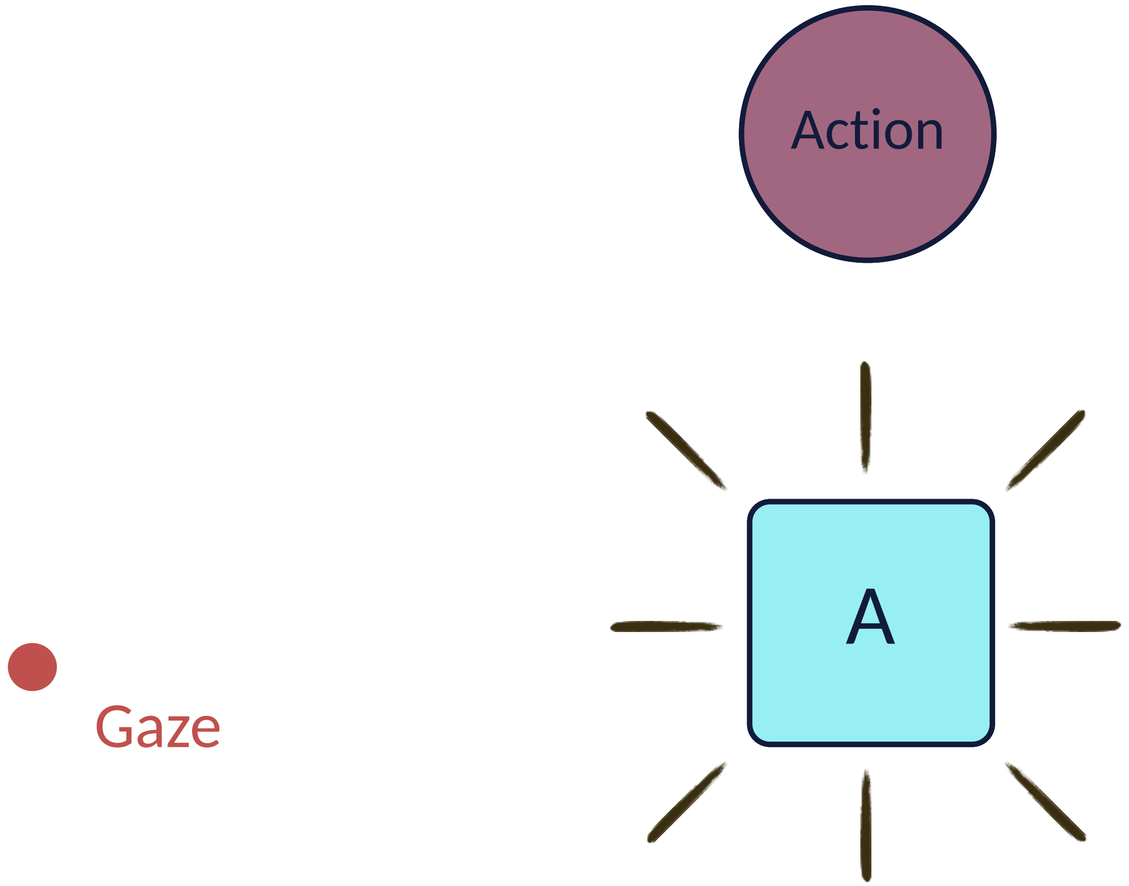
Midas Touch

Reverse Crossing

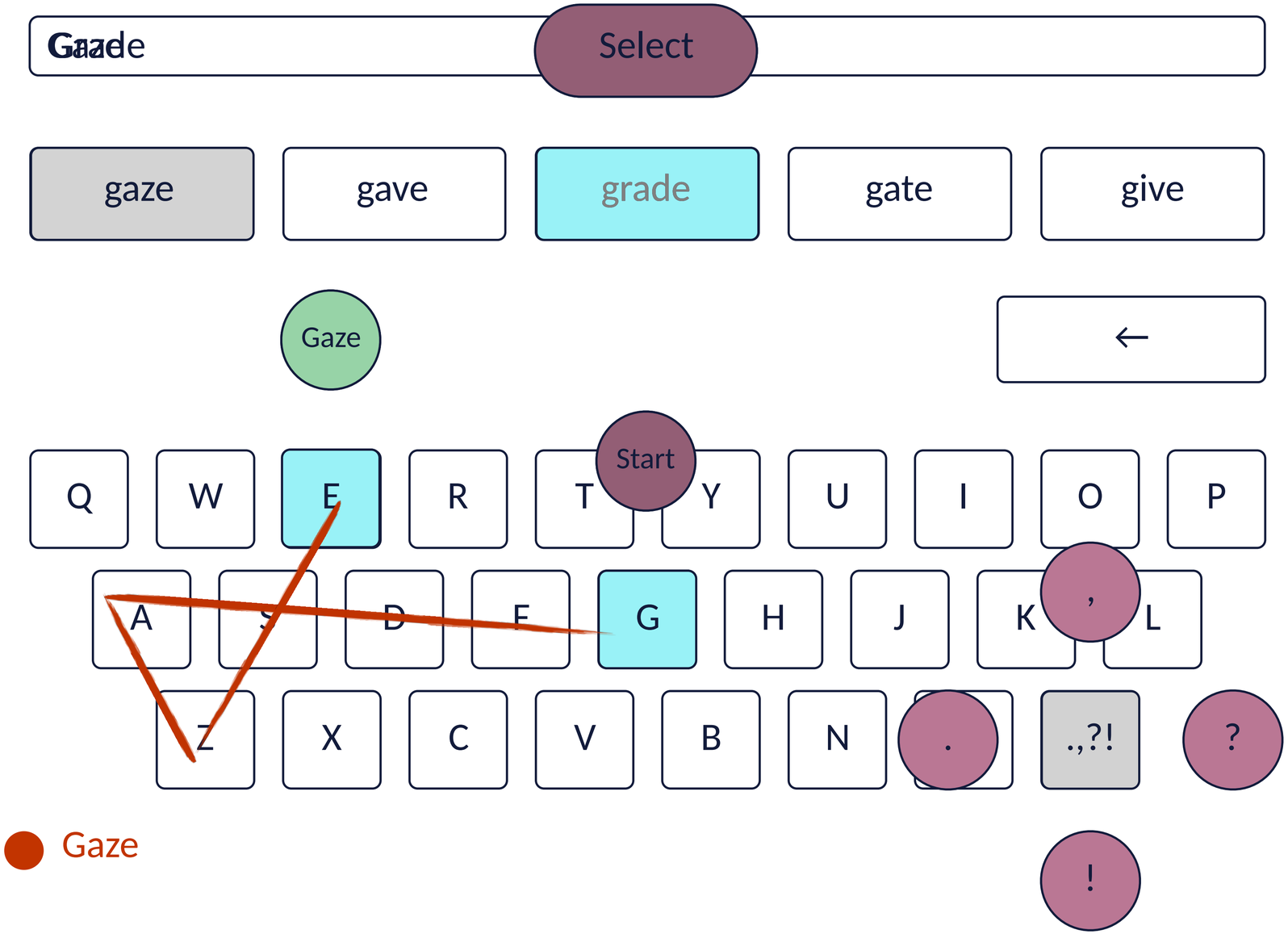


Reverse Crossing

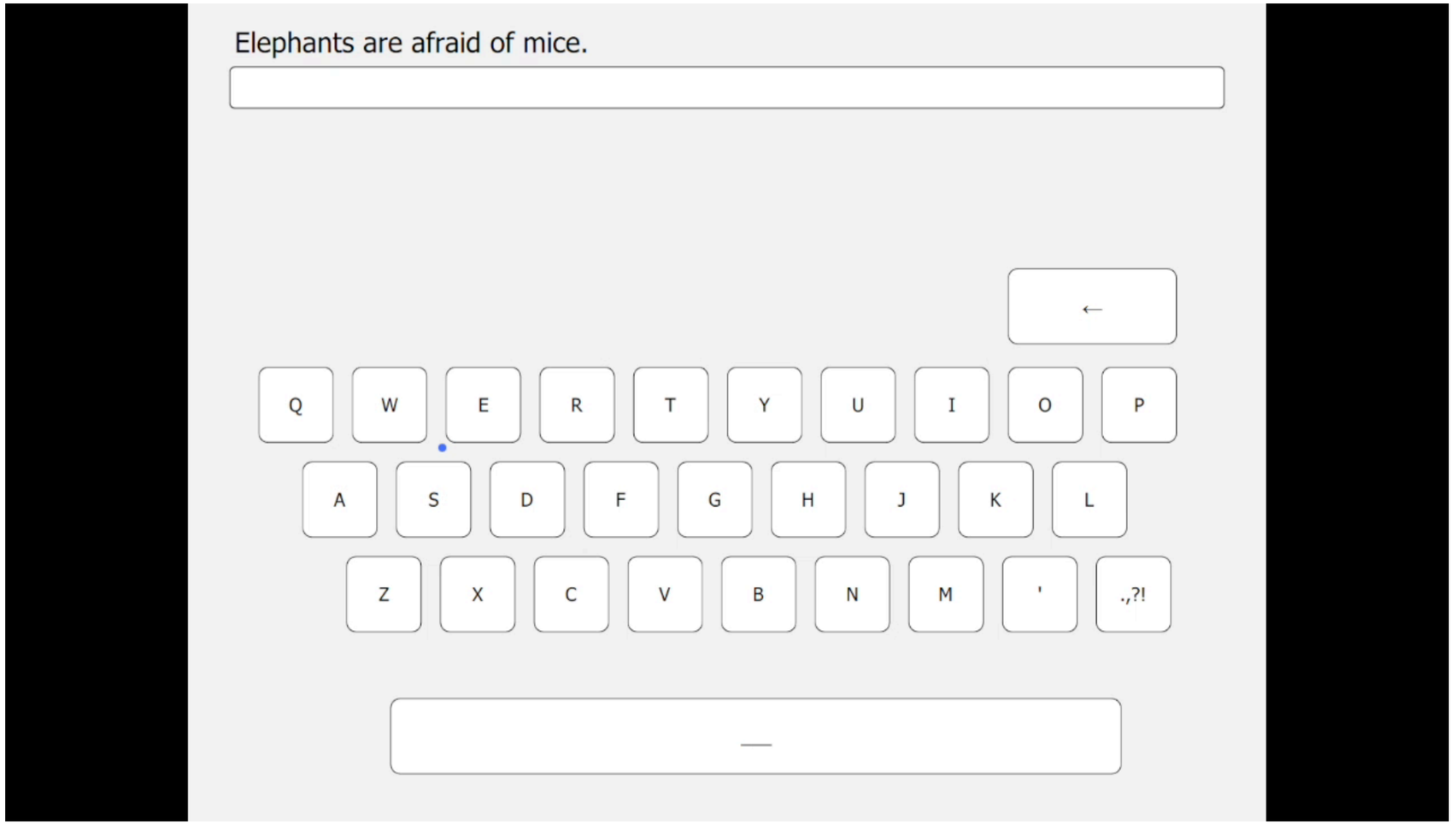
Reverse Crossing



Interface

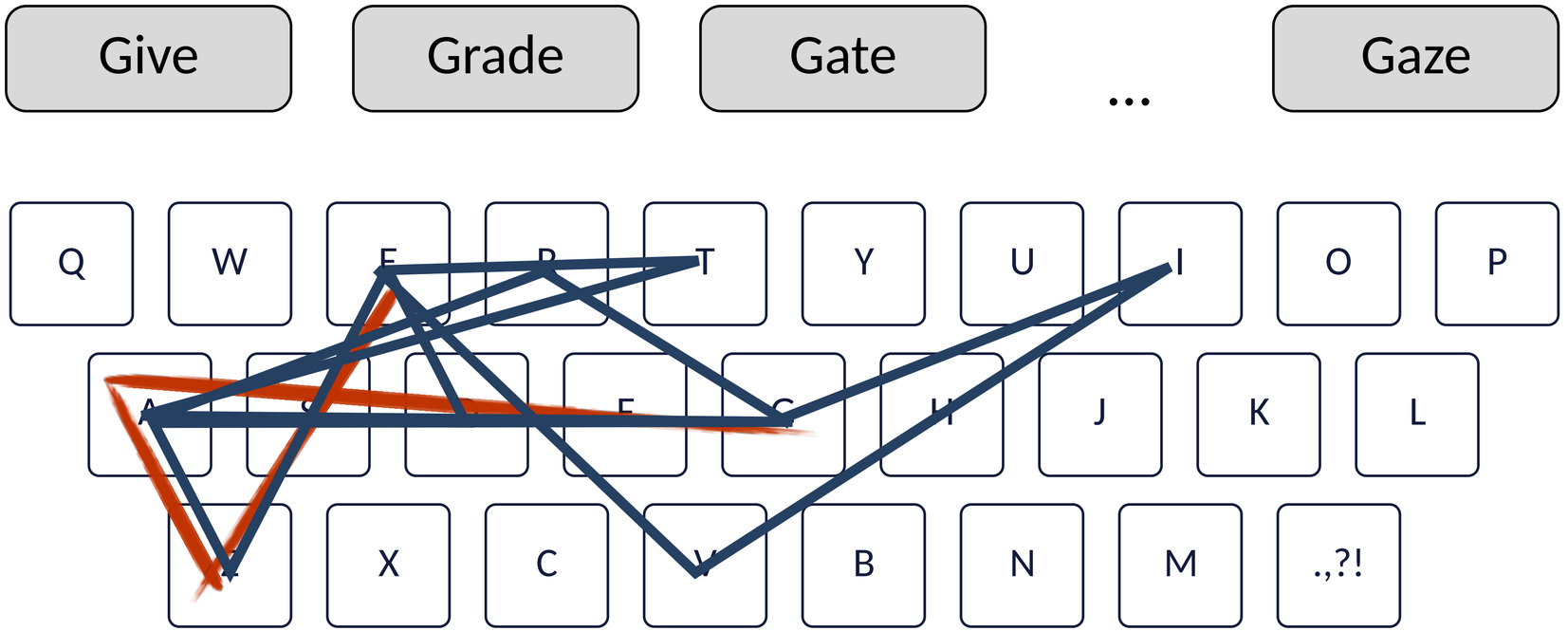

Gaze
Select
gaze
gave
grade
gate
give
Gaze
←
Start
Q
W
E
R
T
Y
U
I
O
P
A
S
D
F
G
H
J
K
L
,
Z
X
C
V
B
N
.
.,?!
?
!
Gaze

Interface


Elephants are afraid of mice.
←
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M ' .,?!
—

Candidate Selection

Find candidates in lexicon beginning in G and ending in E and sort by a weighted combination of DTW distance and a simple language model.



Experiments

Text Entry Method: EyeSwipe | Dwell-time

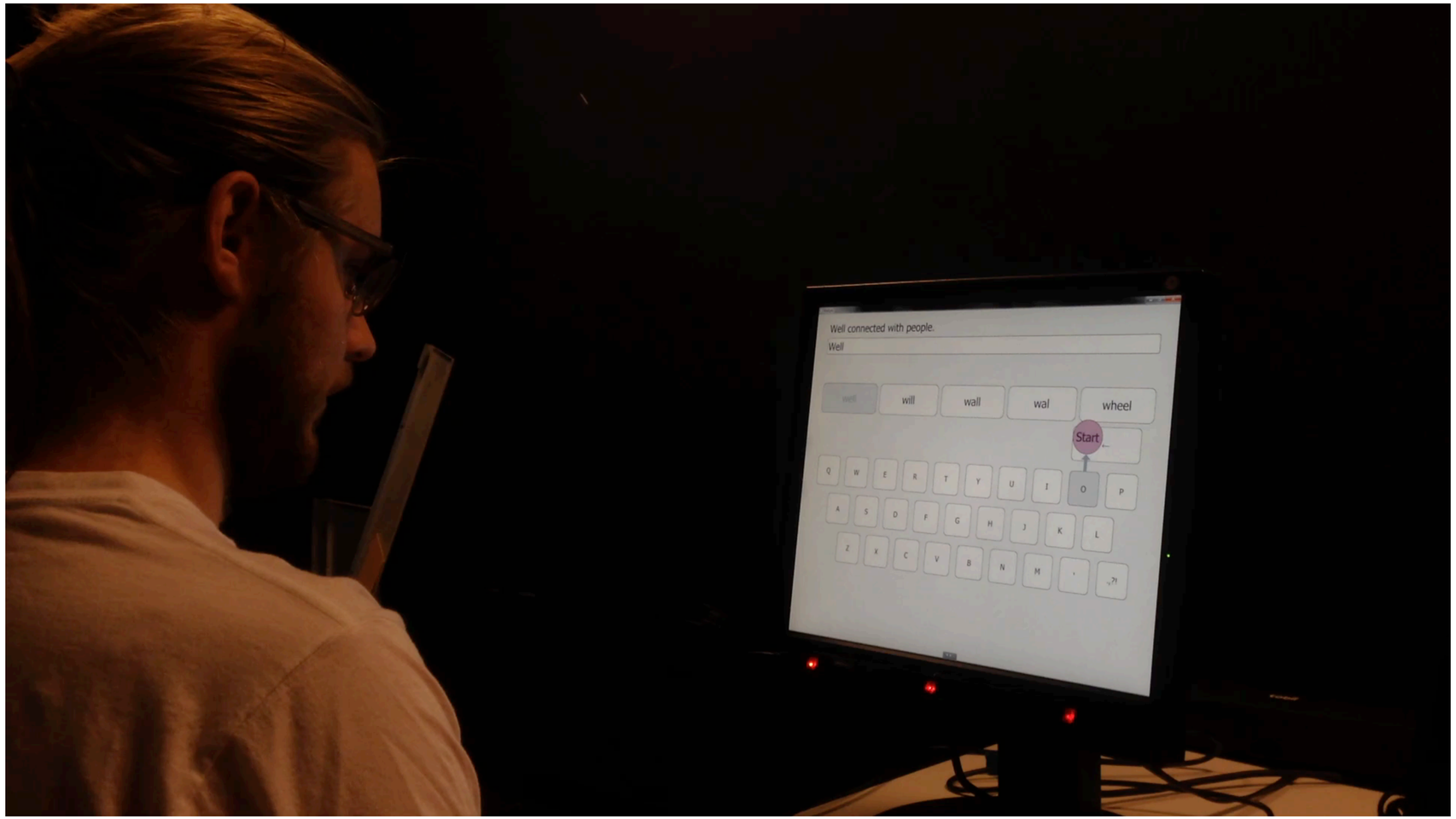
Participants: 10 (5 male, 5 female) | No previous eye-tracking experience

Apparatus: Laptop running Windows 7 | 19” LCD monitor (1024 x 768)

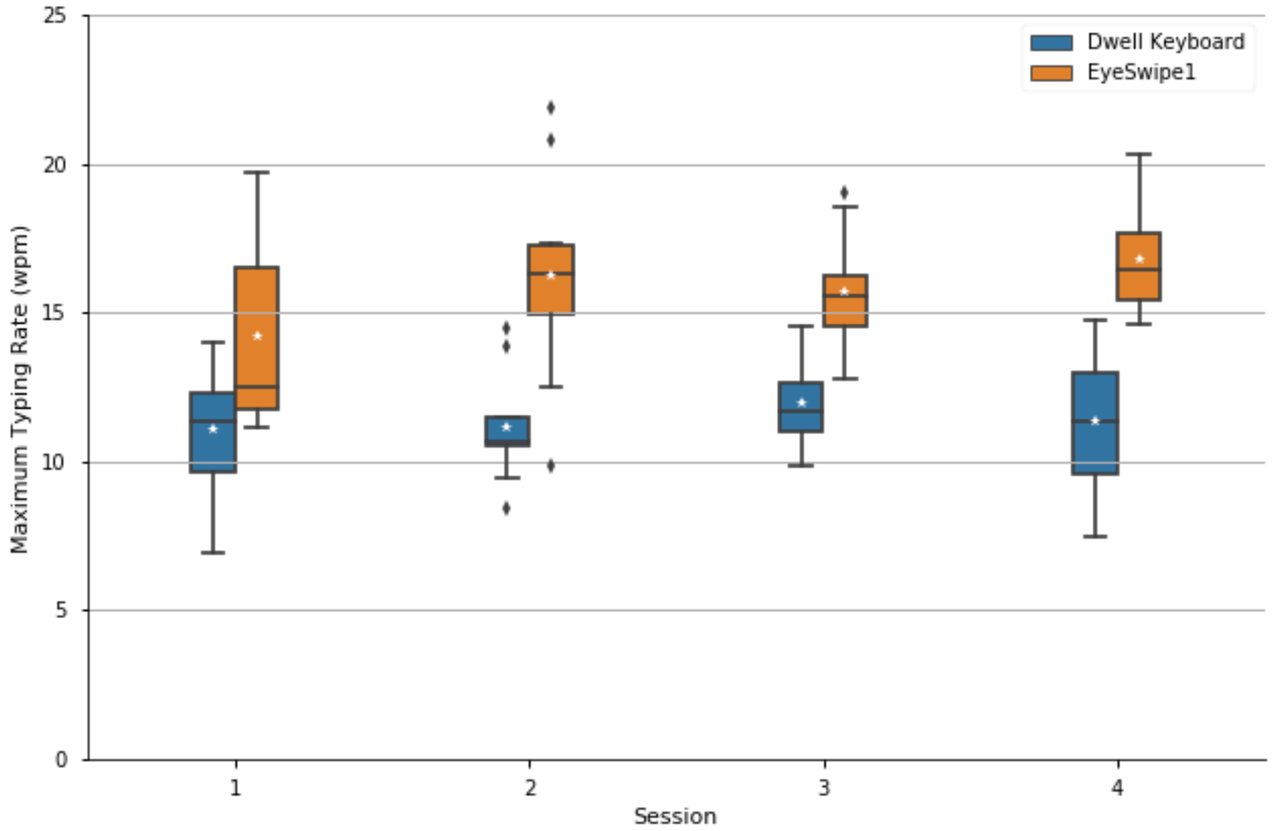
Dataset: MacKenzie and Soukoreff

Procedure: 2 days x 2 sessions x 2 methods | 10 minutes per session

Experiments

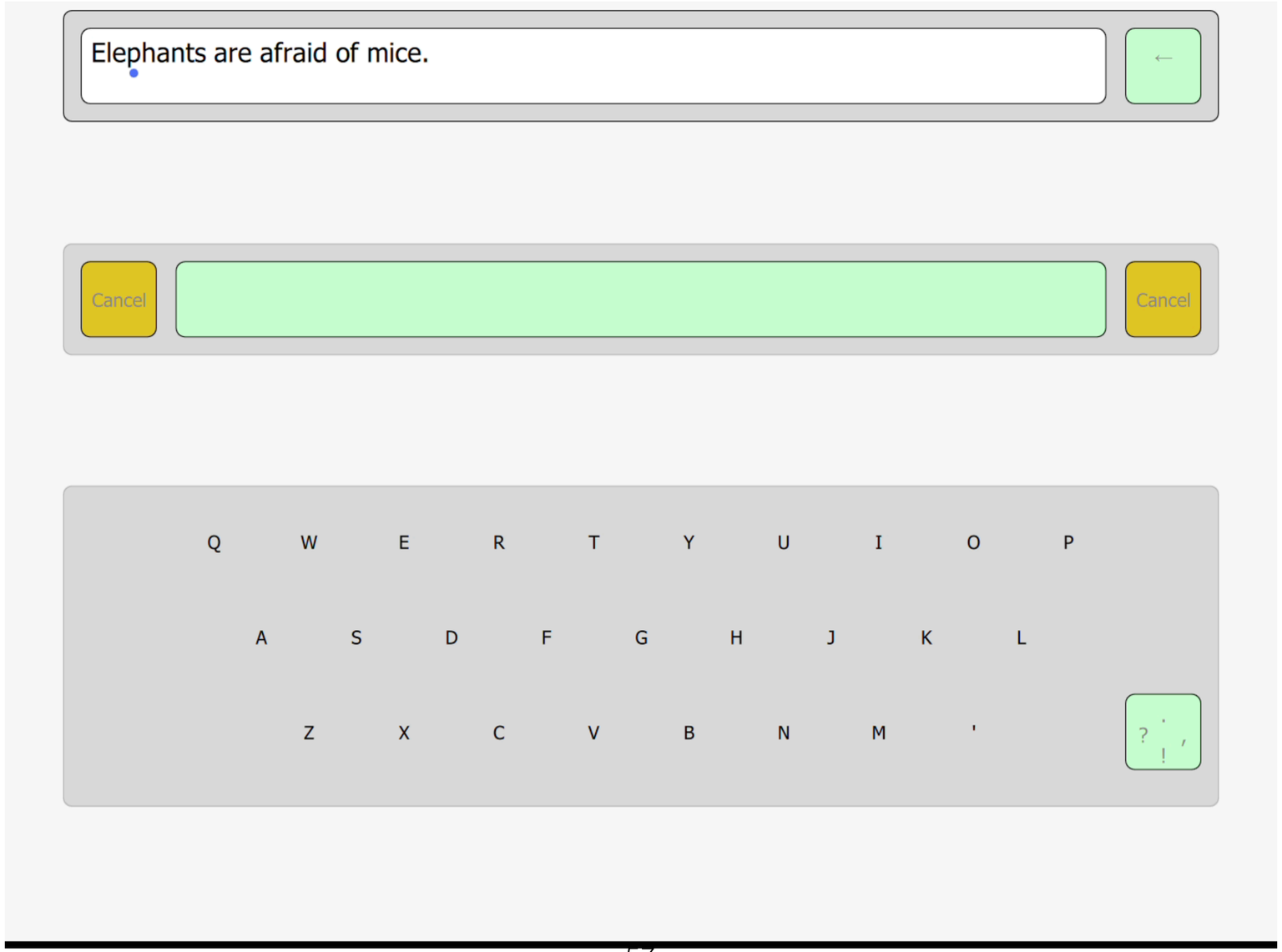


Text Entry Rate



Text Entry Rate

Swipe&Switch



Experiments

Text Entry Method: EyeSwipe1 | Swipe&Switch

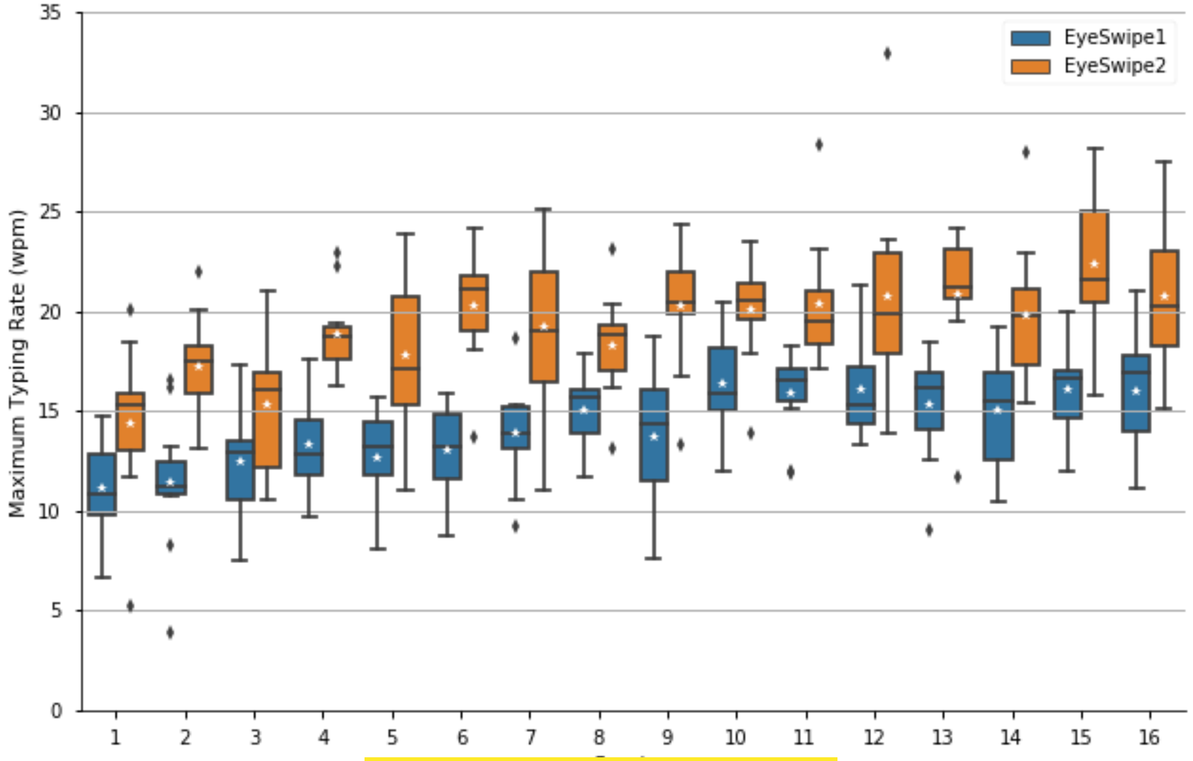
Participants: 11 (4 male, 7 female) | No previous eye-tracking experience

Apparatus: Laptop running Windows 7 | 19” LCD monitor (1024 x 768)

Dataset: MacKenzie and Soukoreff

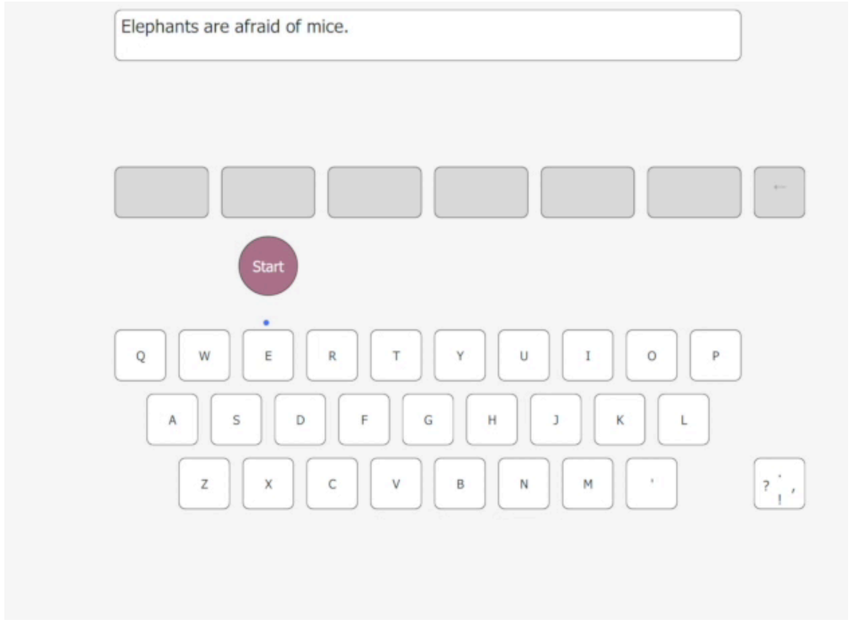
Procedure: 4 days x 4 sessions x 2 methods | 5 minutes per session

Text Entry Rate



Text Entry Rate

Swipe&Switch



Dwell-time

EyeSwipe1

Swipe&Switch*

*In submission

Affectiva

Today's technology is increasingly interacting with humans the way humans do – through conversation, perception and more

The problem: technology has high IQ, no EQ
Emotional intelligence is the missing element

Affectiva builds artificial emotional intelligence:
Emotion AI

Emotion AI will be ubiquitous

Affectiva

Data

Multi-Modal Data Acquisition

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

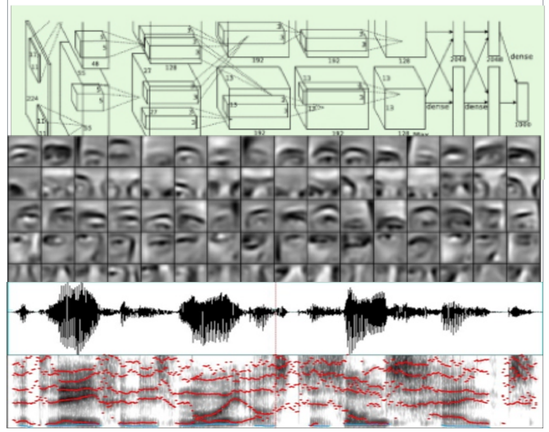
Data Annotation Infrastructure

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

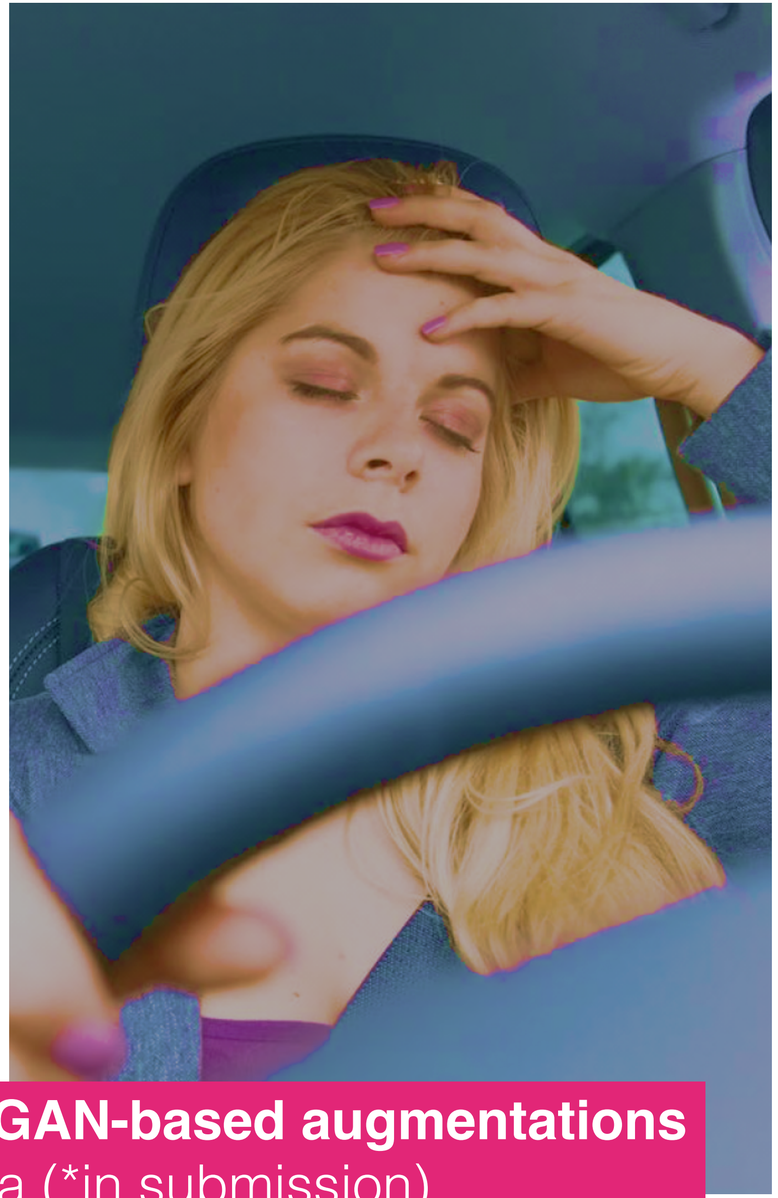
Affectiva Automotive AI

AI-based in-cabin sensing solution that detects nuanced cognitive and emotional states from face and voice.

- Improve road safety by monitoring **driver state**.
- Deliver optimal transportation experience by measuring **occupant mood and reaction**.

Drowsiness

It is conservatively estimated that between **2.3% to 2.5%** of all police-reported fatal crashes between 2011-2015 involved drowsy driving, resulting in more than **4,000 deaths**.

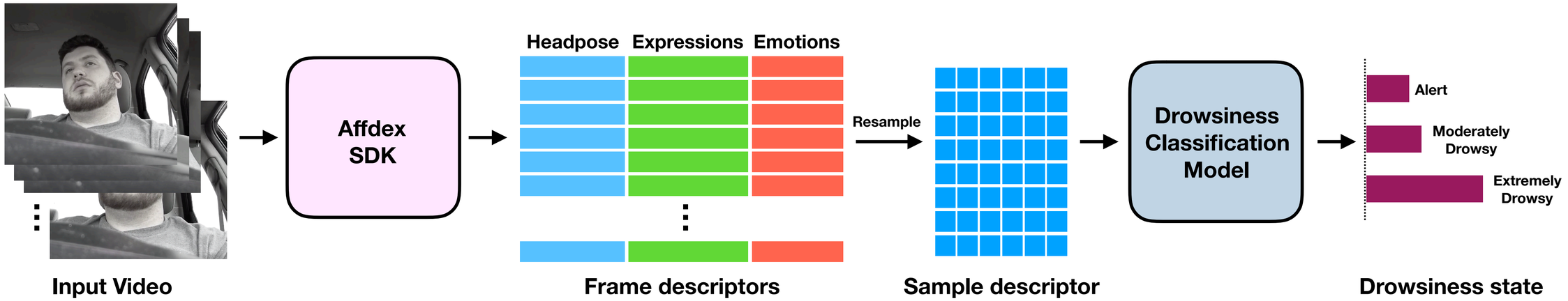


In-the-wild Drowsiness Detection from Facial Expressions using GAN-based augmentations

Ajjen Joshi[1], Sandipan Banerjee[1], Survi Kyal[1], Taniya Mishra (*in submission)

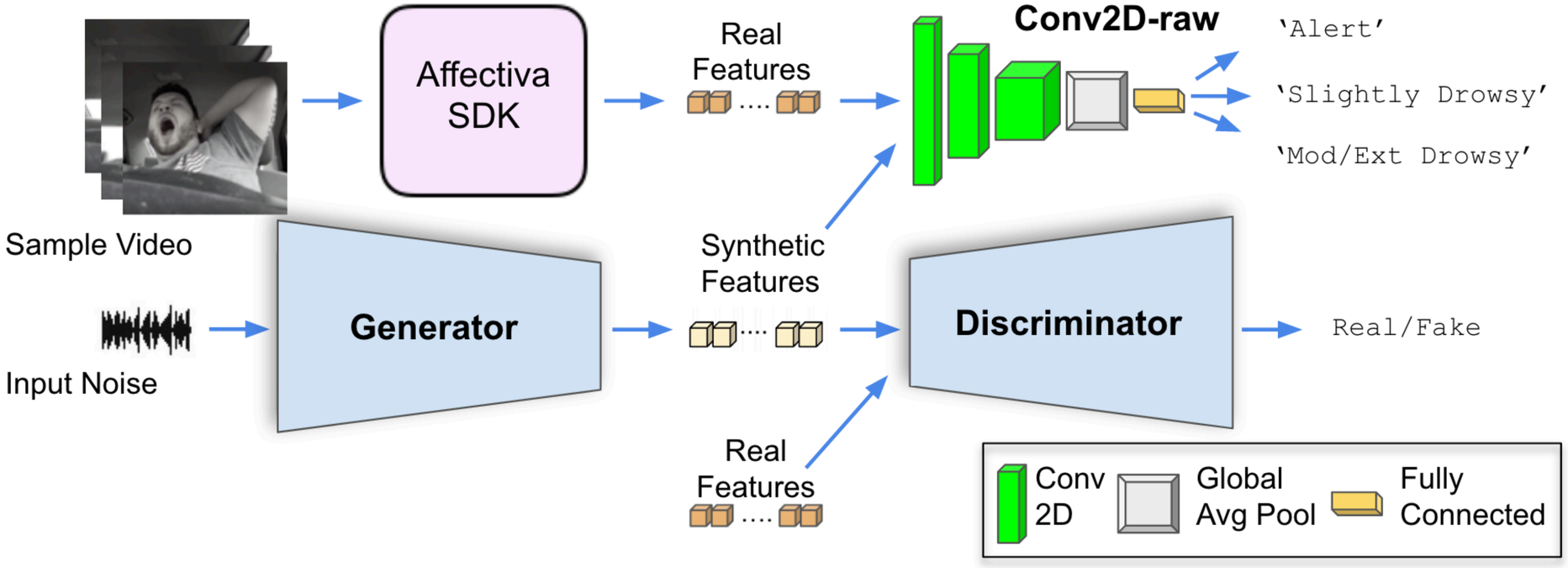
[1]Equal contribution

Drowsiness Classification Schema



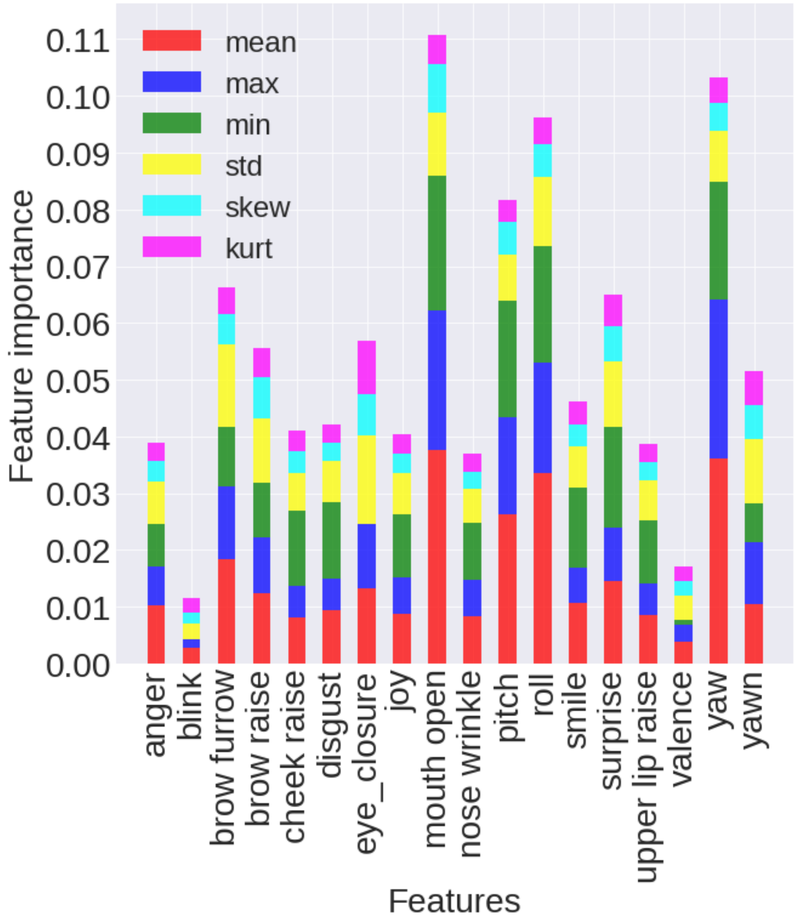
Drowsiness Classification Schema

GAN-based Data Augmentation



Synthesizing examples of sparsely occurring classes

Experimental Results



Feature Importance

Model	AUC	Acc	Pre	Rec	F1
RF-baseline	0.72	0.42	0.39	0.42	0.39
MLP-stats	0.71	0.58	0.66	0.58	0.59
MLP-raw	0.71	0.53	0.64	0.53	0.53
MLP-enc	0.74	0.50	0.65	0.50	0.52
Conv1D-raw	0.75	0.59	0.64	0.59	0.57
Conv2D-raw	0.78	0.63	**0.69**	0.63	**0.63**
LSTM-raw	0.77	0.57	0.64	0.57	0.54
Conv2D-raw + SMOTE	0.75	0.64	**0.69**	0.64	**0.63**
Conv2D-raw + GAN	**0.80**	**0.65**	0.66	**0.65**	**0.63**

Model Performance

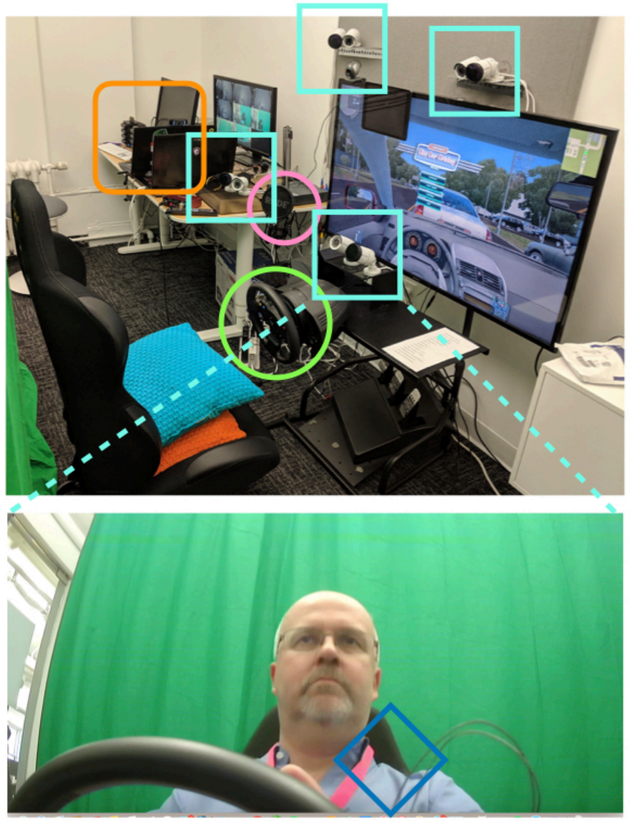
Frustration

In the context of driving, frustration occurs when the goal of achieving mobility is impeded, for example, by red-light signals, slow moving vehicles, or blocked path by other vehicles or pedestrians [Malta et al., 2011].

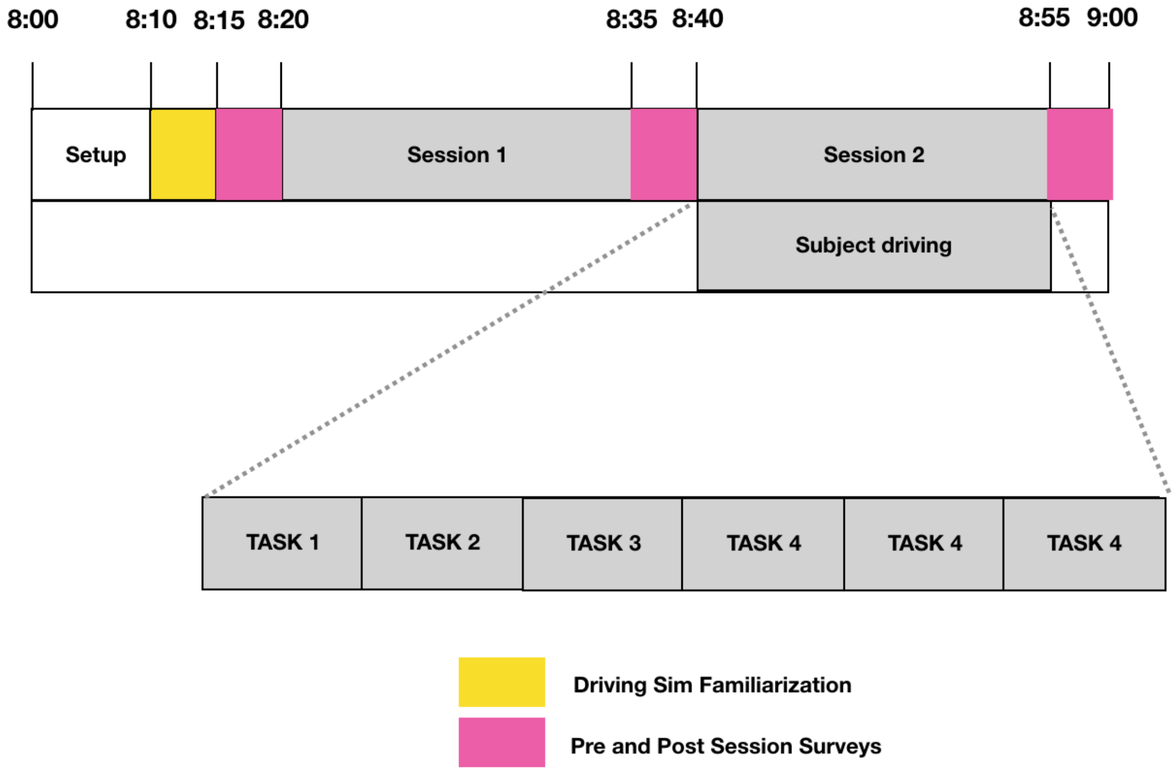
Driver frustration can lead to aggressive driving behavior.

Protocol for Eliciting Frustration in an In-vehicle Environment
Ajjen Joshi, Yousseff Attia, Taniya Mishra (ACII ’19)

Collecting Frustration Data

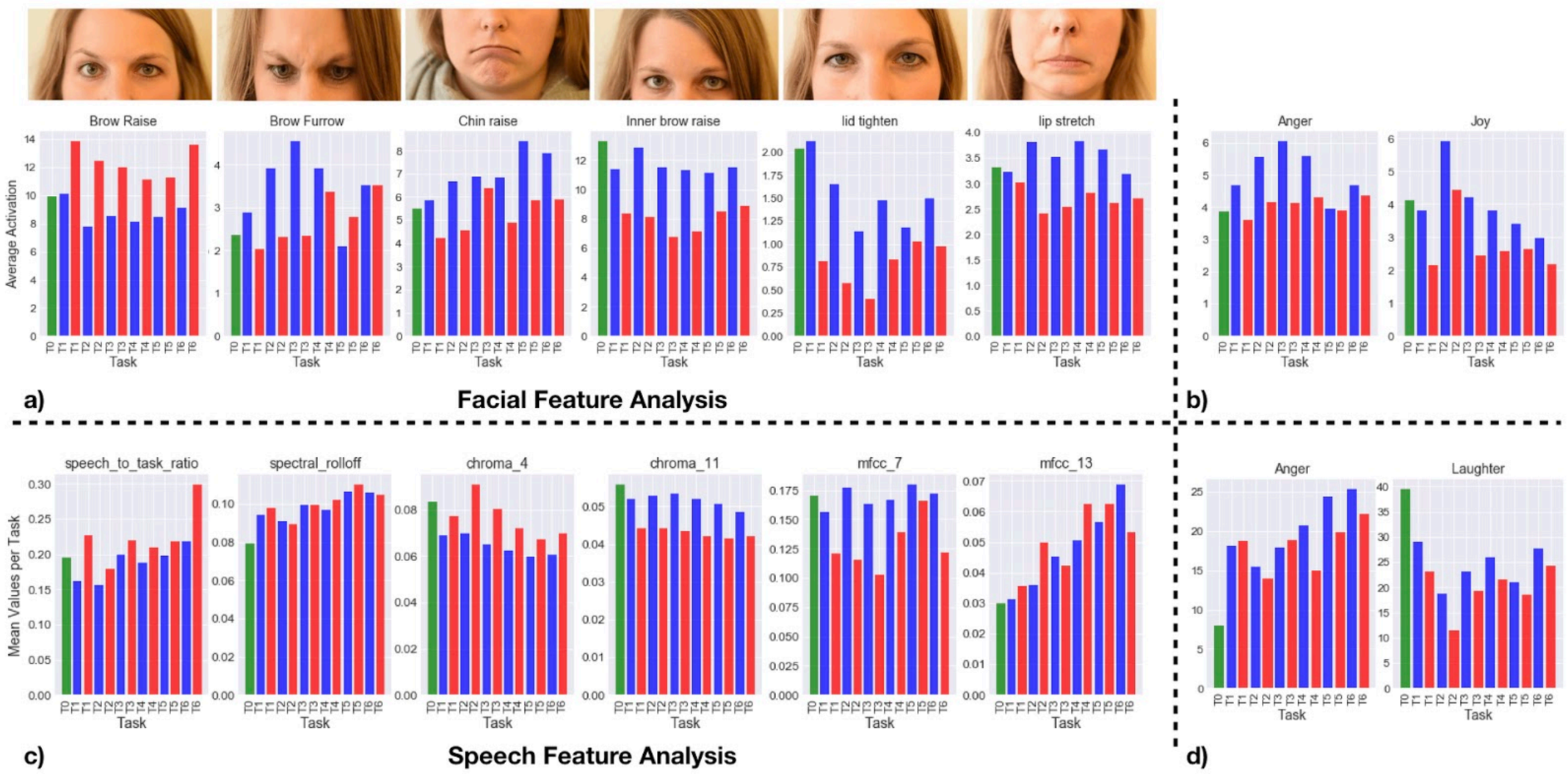


- Driving Simulator
- RGB/nIR cameras
- High Quality Microphone
- Integration Platform
- Physiological Sensors (ECG and GSR)



Data Collection Protocol

Frustration: Audio-Video Feature Manifestations



Boredom

Mind-wandering is a phenomenon that is estimated to occur as much as 50% of the time depending on the individual, task, and environment [Schroeter, Ronald et al (2015)]

Boredom can lead to distraction.

Mitigating Boredom using an Empathetic Conversational Agent
Samiha Samrose, Kavya Anbarasu, Ajjen Joshi, Taniya Mishra (CHI CUI 2020)

Inducing Boredom

Start

Left: press z

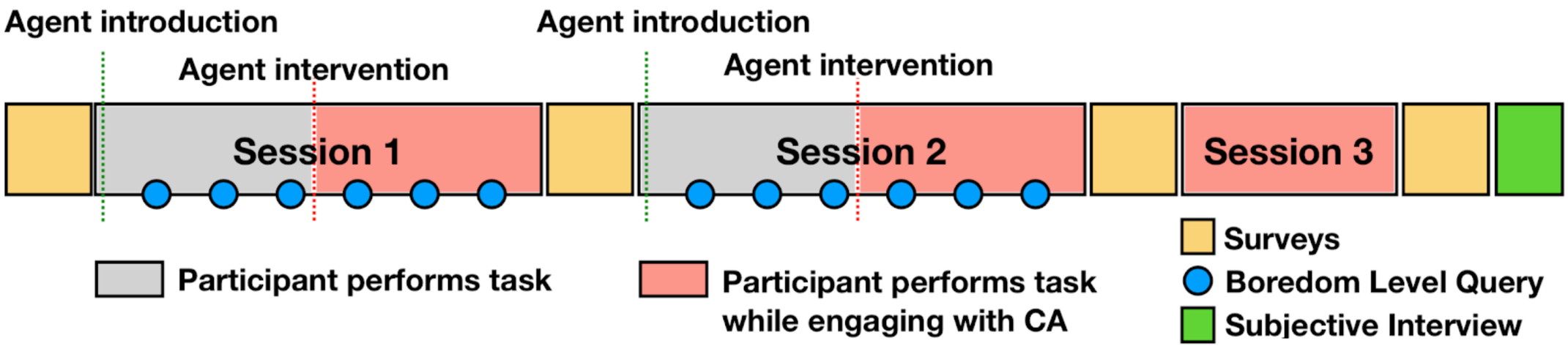
Right: press m

On a scale of 0-9, how bored are you right now? (Press Key)

0 1 2 3 4 5 6 7 8 9

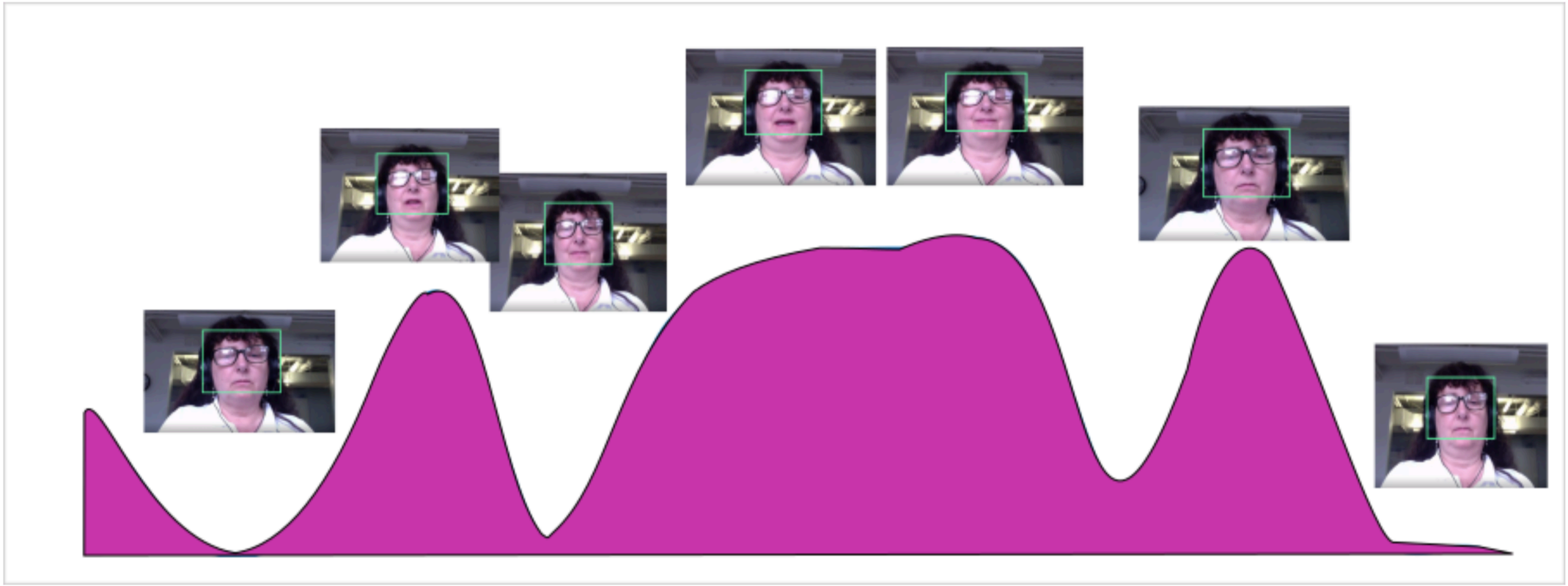
Participant Task UI

Boredom Collection Protocol

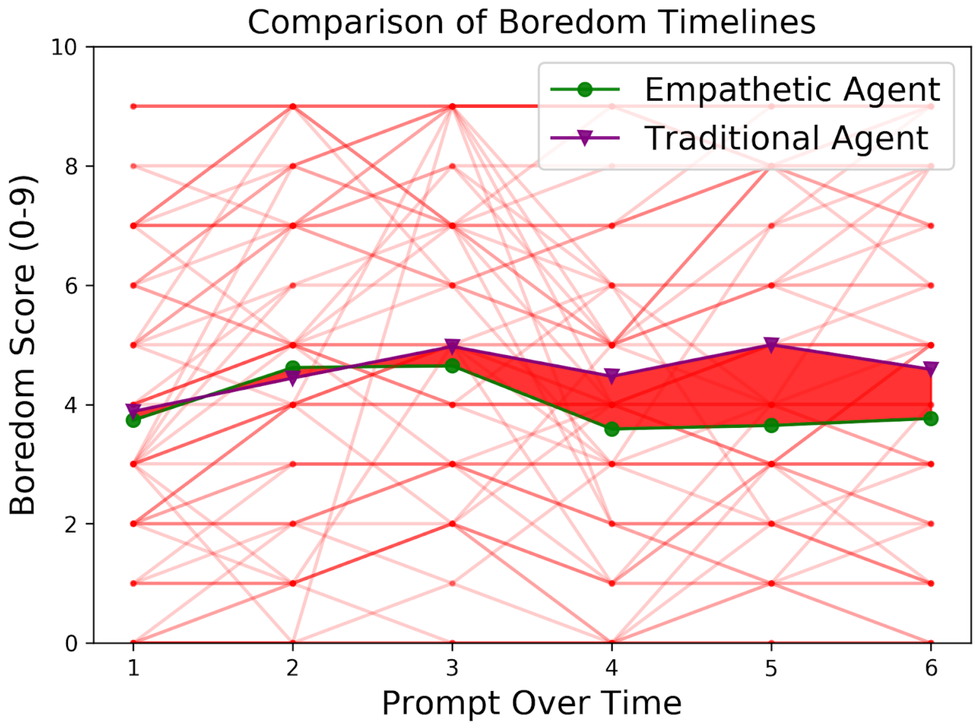


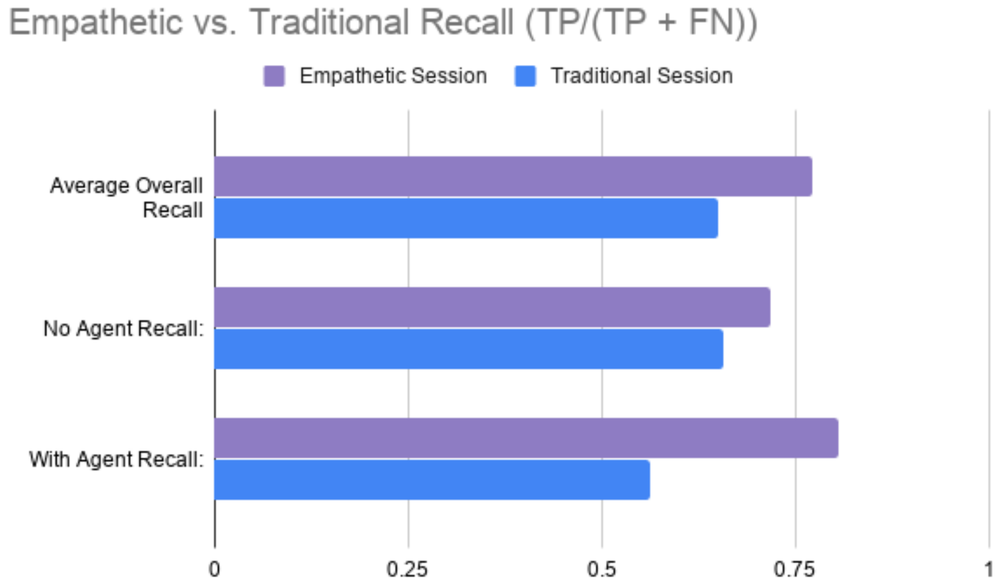
Data Collection Protocol

Evolution of Perceived Engagement



Experimental Results

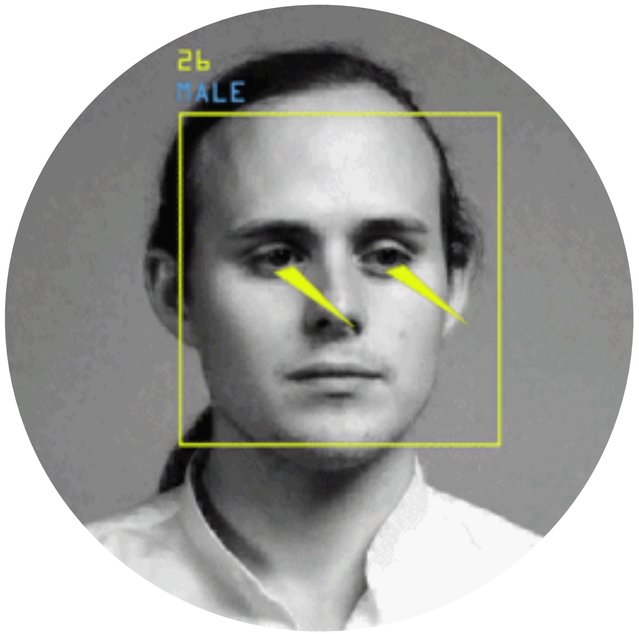




Boredom Mitigation

Task Performance

Summary



Eyegaze

Expressions

Gestures

Questions

Thank you!