Multimodal Interaction

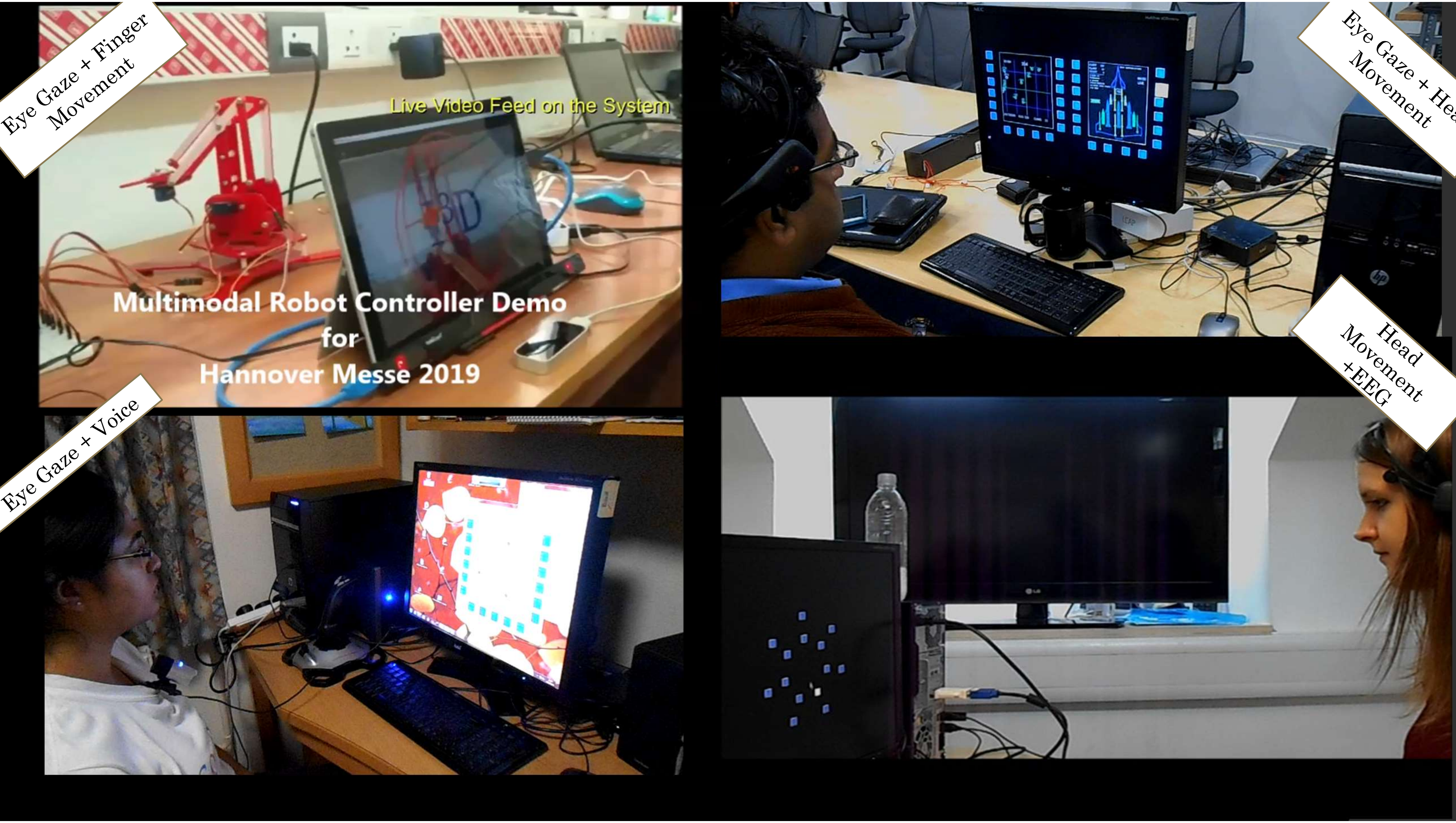
Dr Pradipta Biswas, PhD (Cantab)
Assistant Professor
Indian Institute of Science
https://cambum.net/

What is Multimodal Interaction

- More than one input or output modalities
- Combining more than one modalities together into a single input
  - Mouse plus Eye Gaze movement
- Rendering output in more than one modality
  - Screen plus Spoken Text

Applications

- Gaming Consoles
- Information Visualization
- Automotive Environment
- Aviation Environment
- Assistive Technology


Eye Gaze + Finger Movement
Live Video Feed on the System
Multimodal Robot Controller Demo
for
Hannover Messe 2019
Eye Gaze + Hea
Movement
Head Movement +EEG
Eye Gaze + Voice

Advantages over unimodal systems

- Easier to use; Less training
- Robust, flexible
- Preferred by users
- Faster, more efficient
- Supports new functionality
- Applies to many different environments and form factors that challenge GUI, especially mobile ones

Challenges

- Mismatch in latencies
- Different ranges of (in)accuracies of sensors
- Simultaneous input – setting precedence
- Quality of multimodal input or output

Fusion Strategies

- **Time of fusion** : Fusion is classified into pre, mid and post-mapping with respect to the time of mapping raw sensor data into recognizable symbols.
- **Information to be fused**: Information can be fused at raw data level, feature level or at decision level.
- **Type of cooperation among different modalities**: Different modalities can coexist simultaneously or one at a time or specific modality can be used for specific information.
- **Methods of Fusion**: Different modalities can be fused following rules or modelling them as a set of prior probabilities in a Bayesian model or using filtering techniques on multiple modalities to estimate a probability distribution function.

Case Study 1

Multimodal Gaze Control

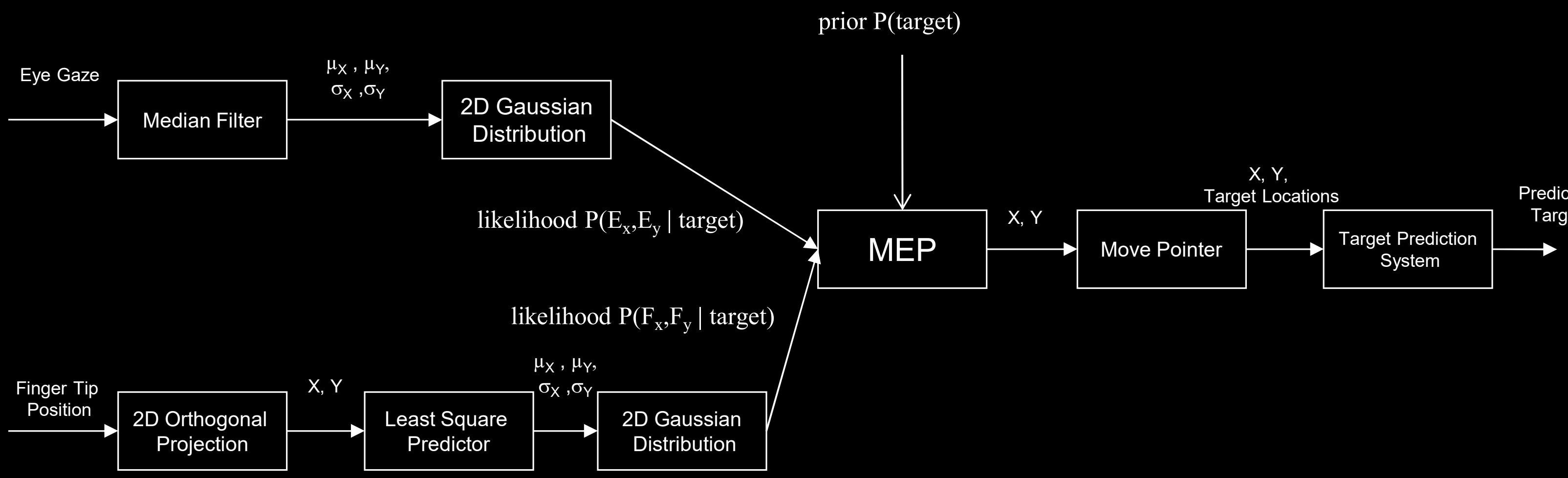
Objective

- Exploring use of non-invasive gaze-controlled interface and finger tracking technology in automotive environment

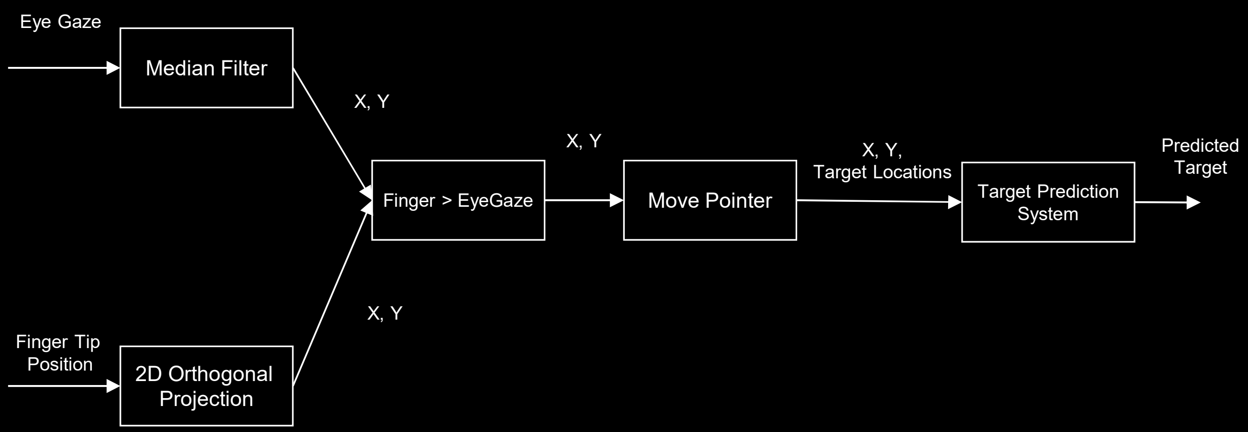
- Comparing different multimodal fusion strategies

Secondary Task- Point & Select

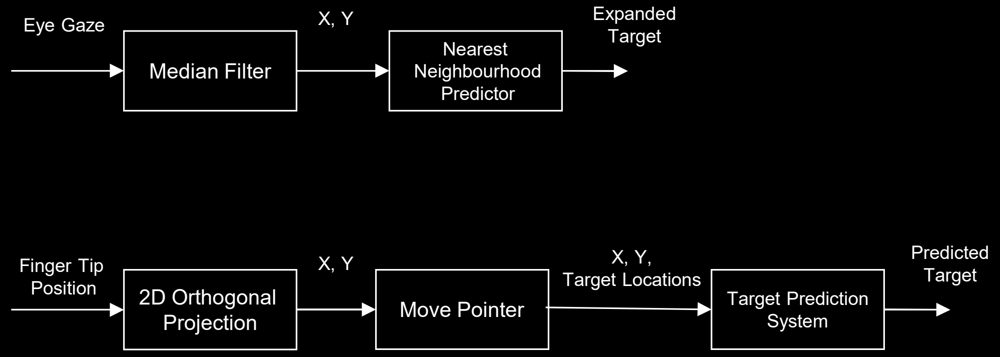
MultiModal Bayesian (MMB)



MultiModal Sequential



Finger Tracking

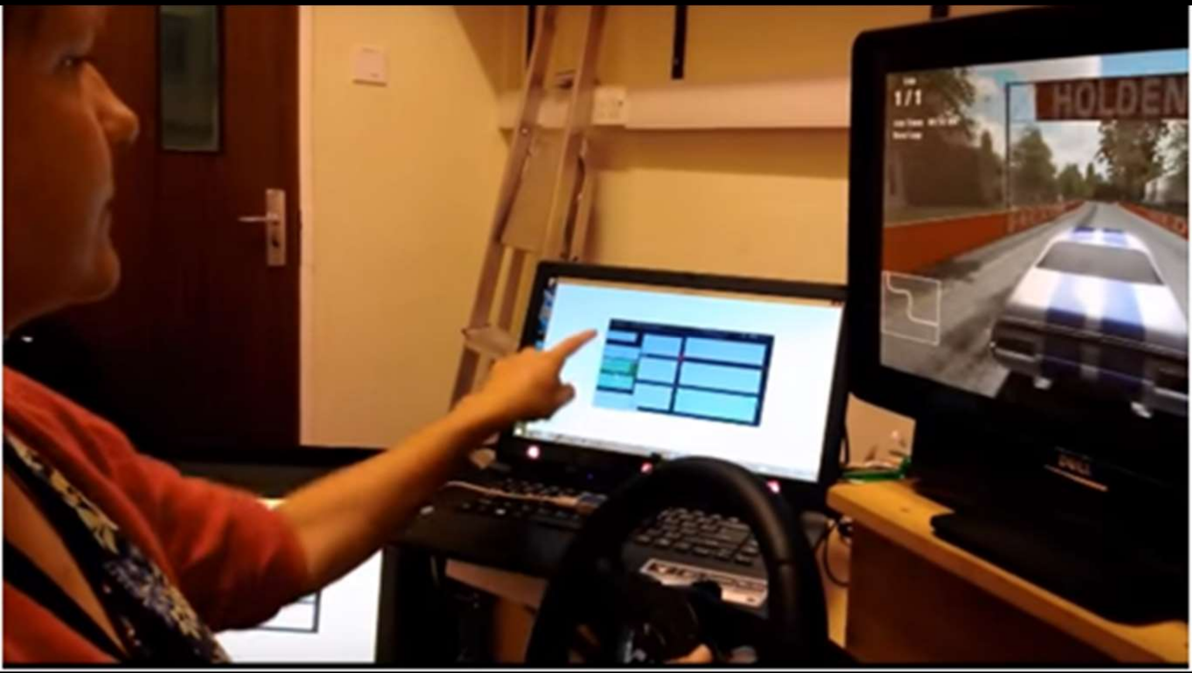


Design

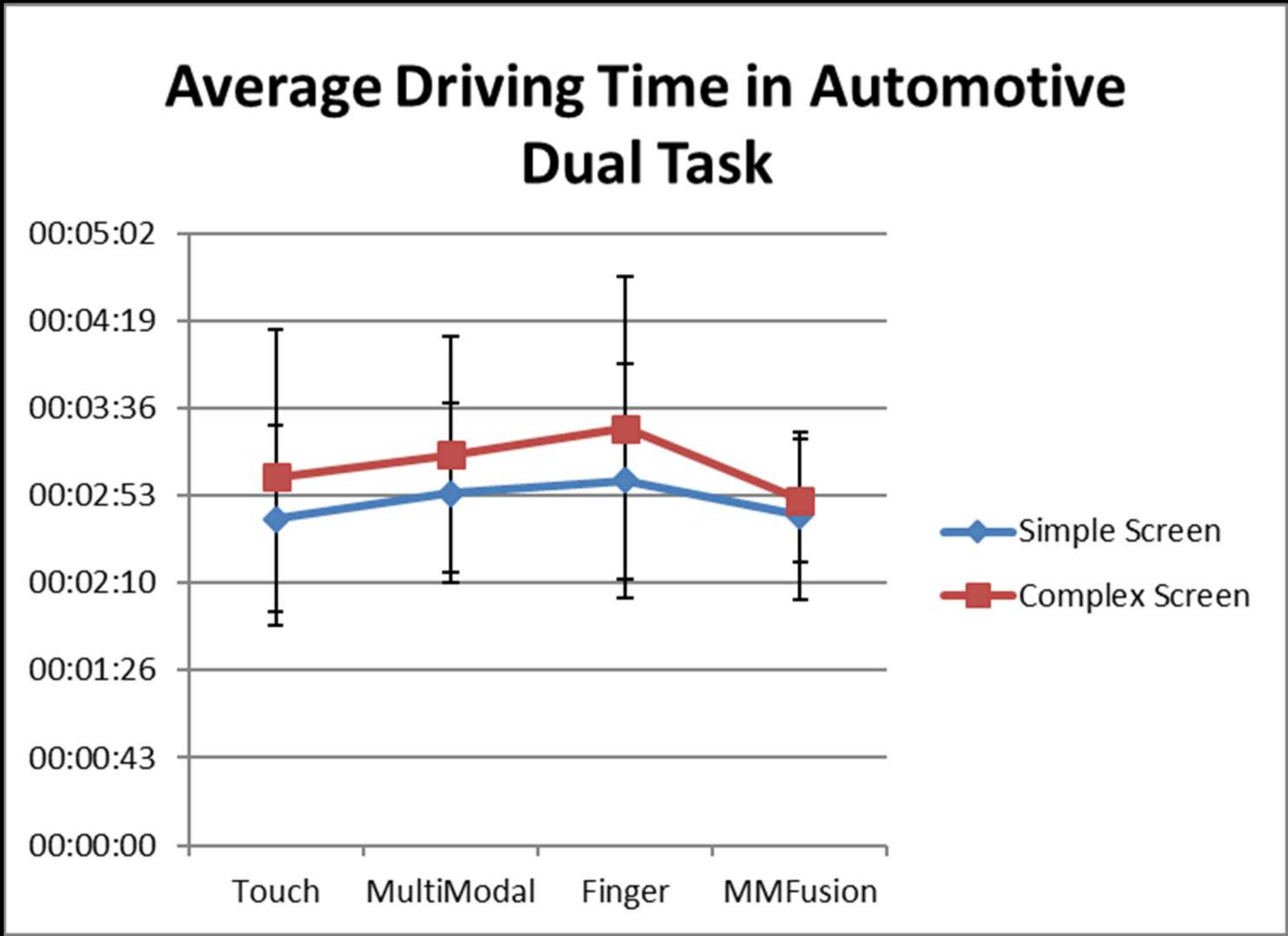
- Dual Task Study
  - Primary task was driving
  - Secondary task was point and selection

- Repeated Measure 2 × 4 Design
  - Two screen layouts
  - Four different interaction strategies
    - Touchscreen and 3 different finger / eye-gaze tracking

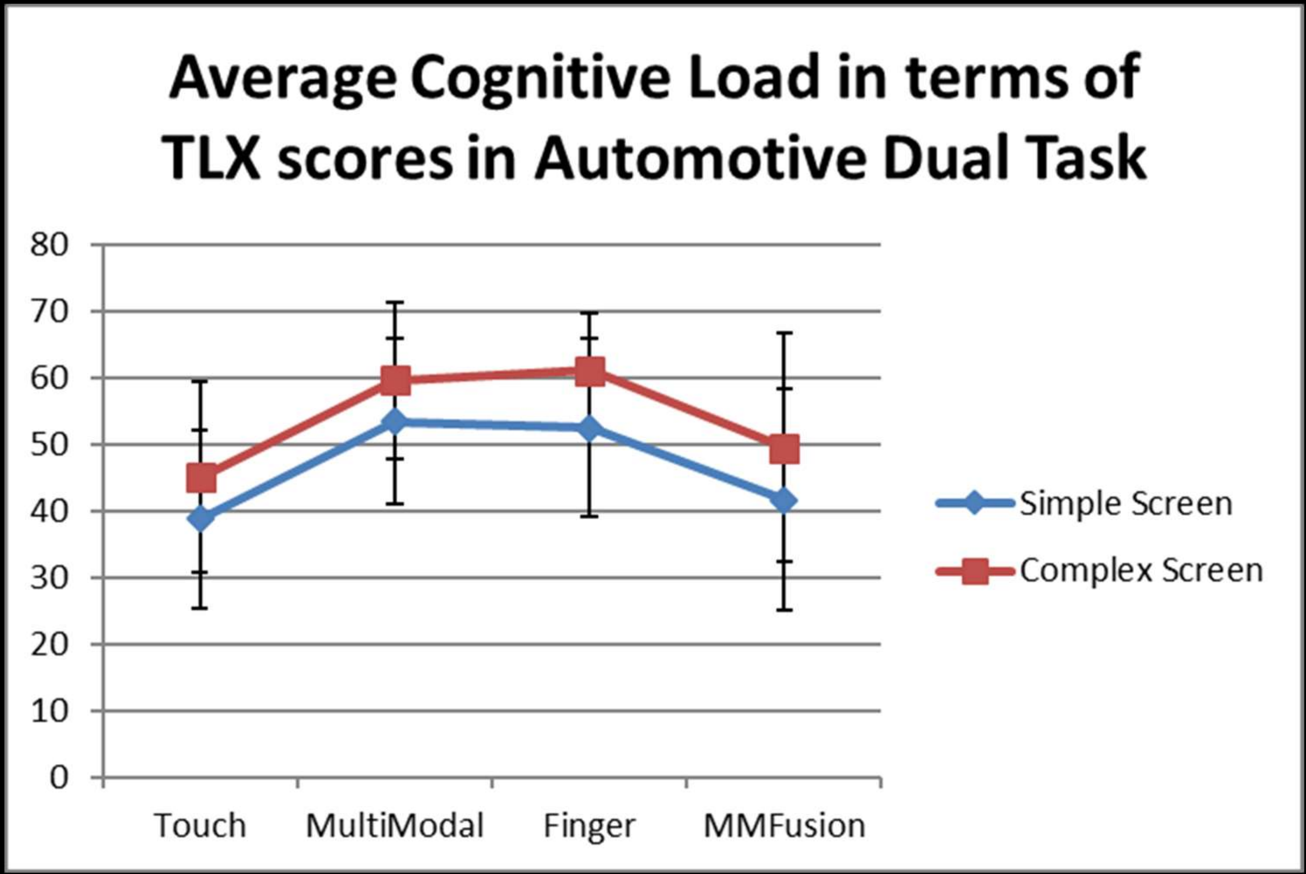
Primary Task



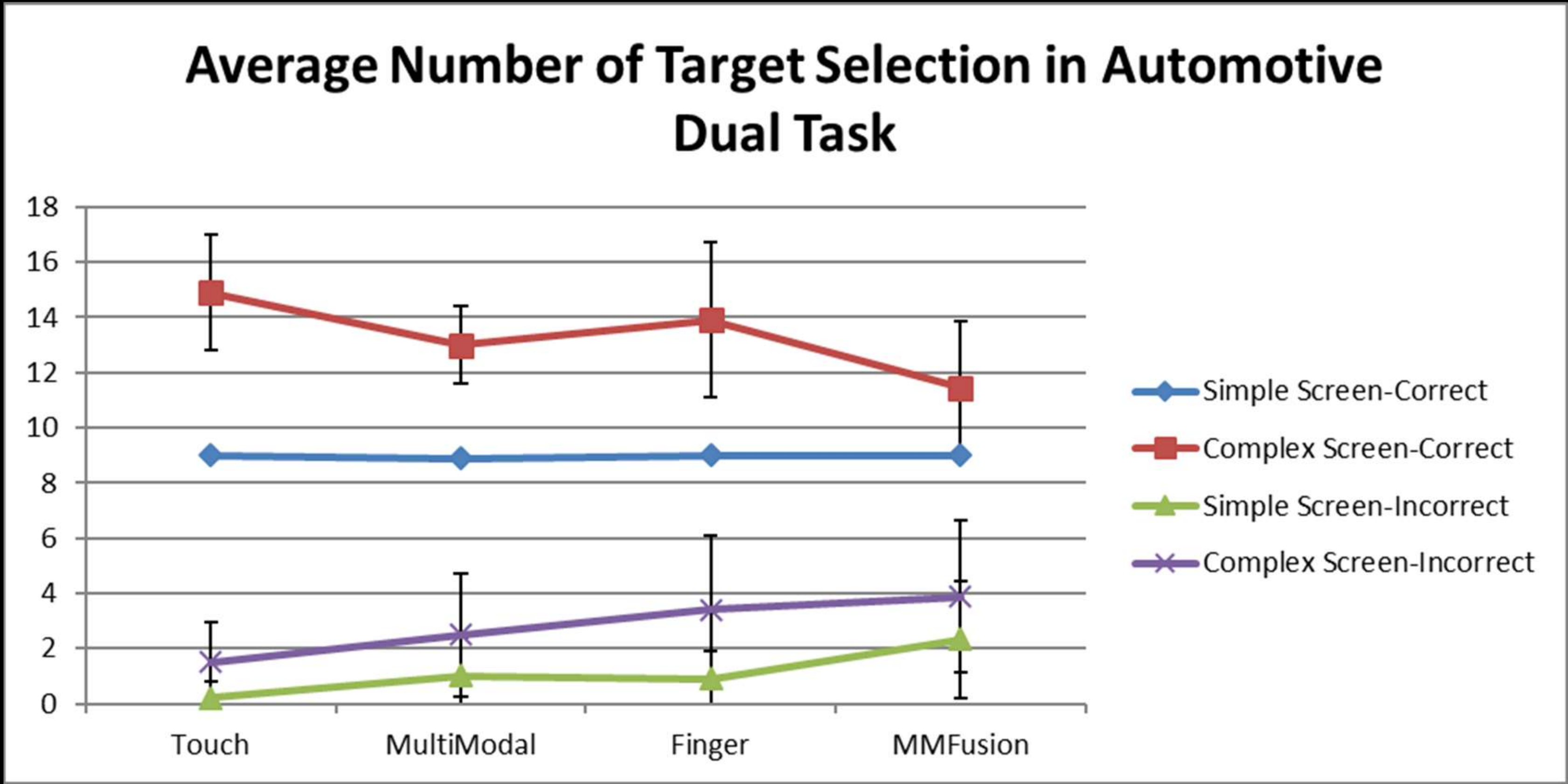
Track Completion Time



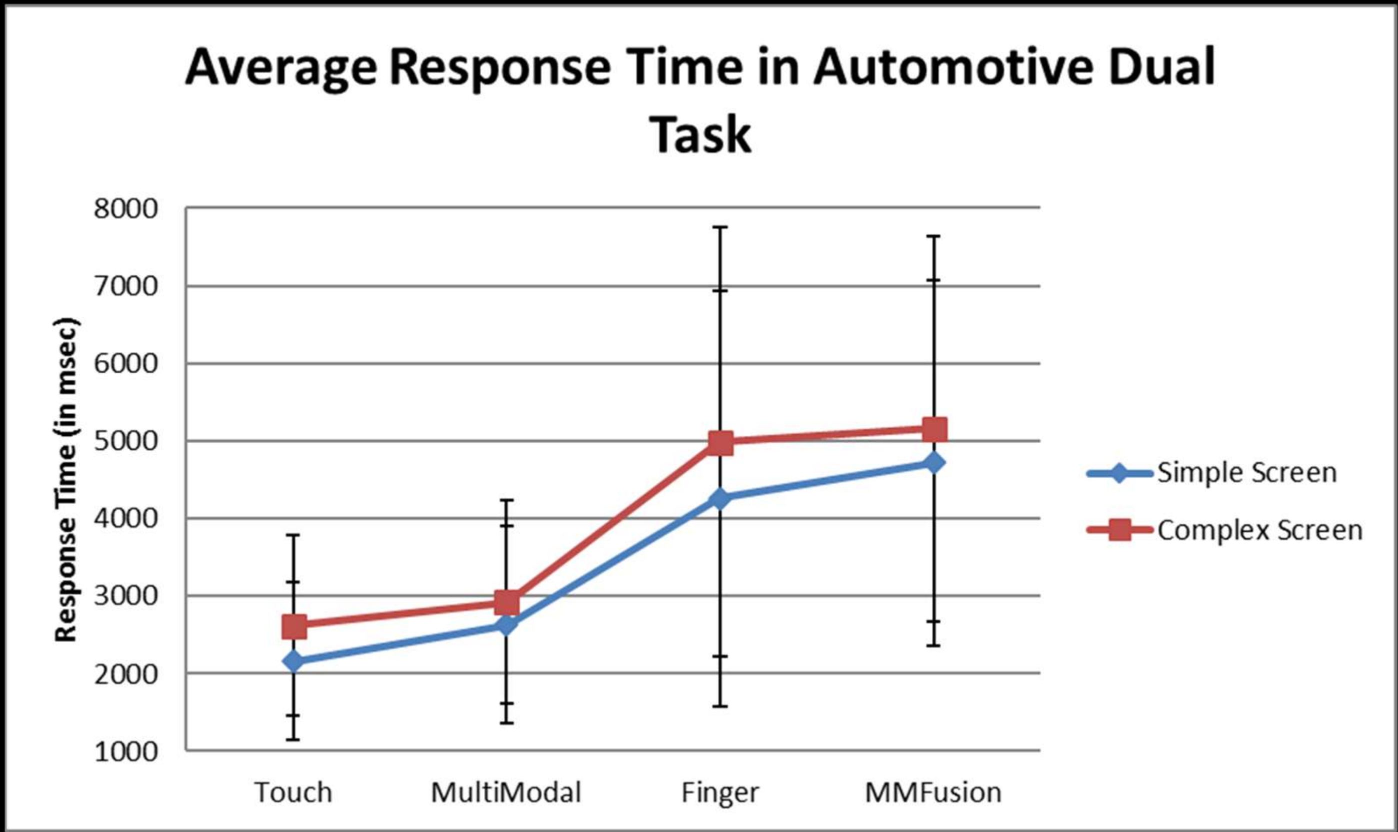
Cognitive Load



Number of Selections



Response Times



Summary of Results

- Using eye gaze for pointing
  - Either reduced driving time
  - Or reduced response time than finger tracking based pointing
- We did not yet have the best fusion strategy
  - The Bayesian Fusion improved driving time and reduced cognitive load but the naïve multimodal system reduced pointing and selection times
- The latency in the finger tracking system reduced its utility as a pointing modality
  - In the touchscreen system only in 10% pointing tasks LeapMotion could track hand movement

Gaze Controlled HUD / HMD

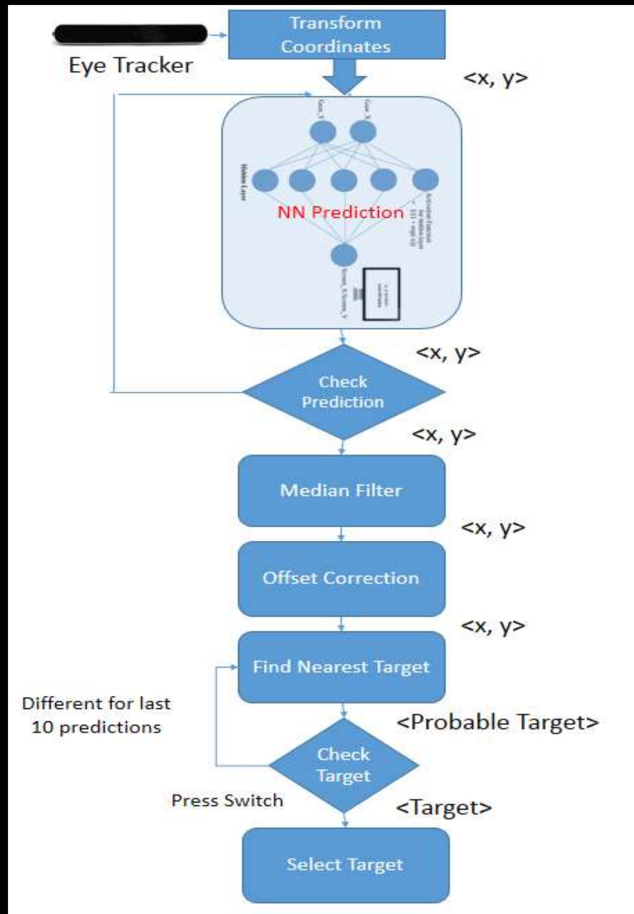
Challenge

- Existing eye trackers are developed for desktop computing environment where
  - Tracker is attached below display
  - Display is a flat screen
- We used eye tracker to track eyes on windshield
- Display was away from eye tracker
- Display surface was not flat like a computer screen

- G Prabhakar, A Ramakrishnan, LRD Murthy, VK Sharma, M Madan, S Deshmukh and P Biswas, Interactive Gaze & Finger controlled HUD for Cars, Journal on Multimodal User Interface, Springer, 2019
- P. Biswas, S. Deshmukh, G. Prabhakar, M. Modiksha and A. Mukhopadhyay, System and Method for Monitoring Cognitive Load of a Driver of a Vehicle, Indian Patent Application No.: 201941052358, PCT Application No.: PCT/IB2020/062016, US Patent Application No.: 17/437,003

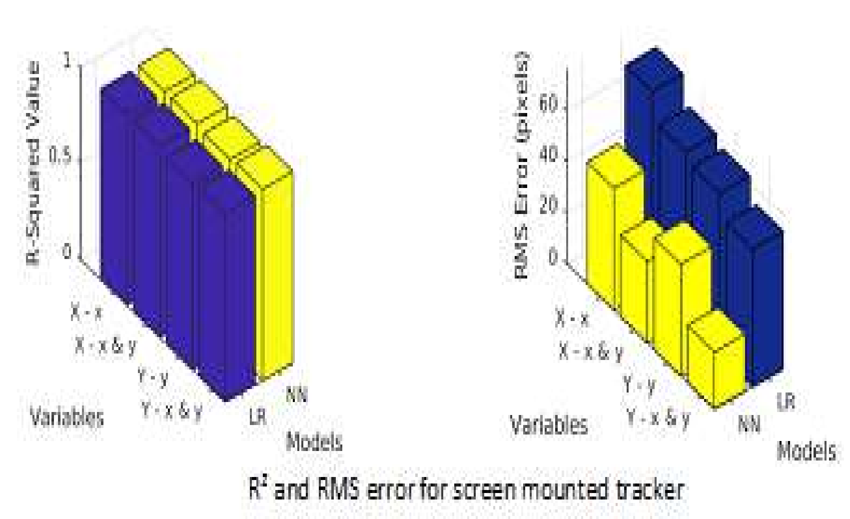
Implementation

- Transform raw gaze coordinates geometrically for inverted image
- Run calibration program to train neural net
- Filter predicted gaze coordinates
- Correct offset based on initial calibration
- Activate target nearest to predicted gaze location

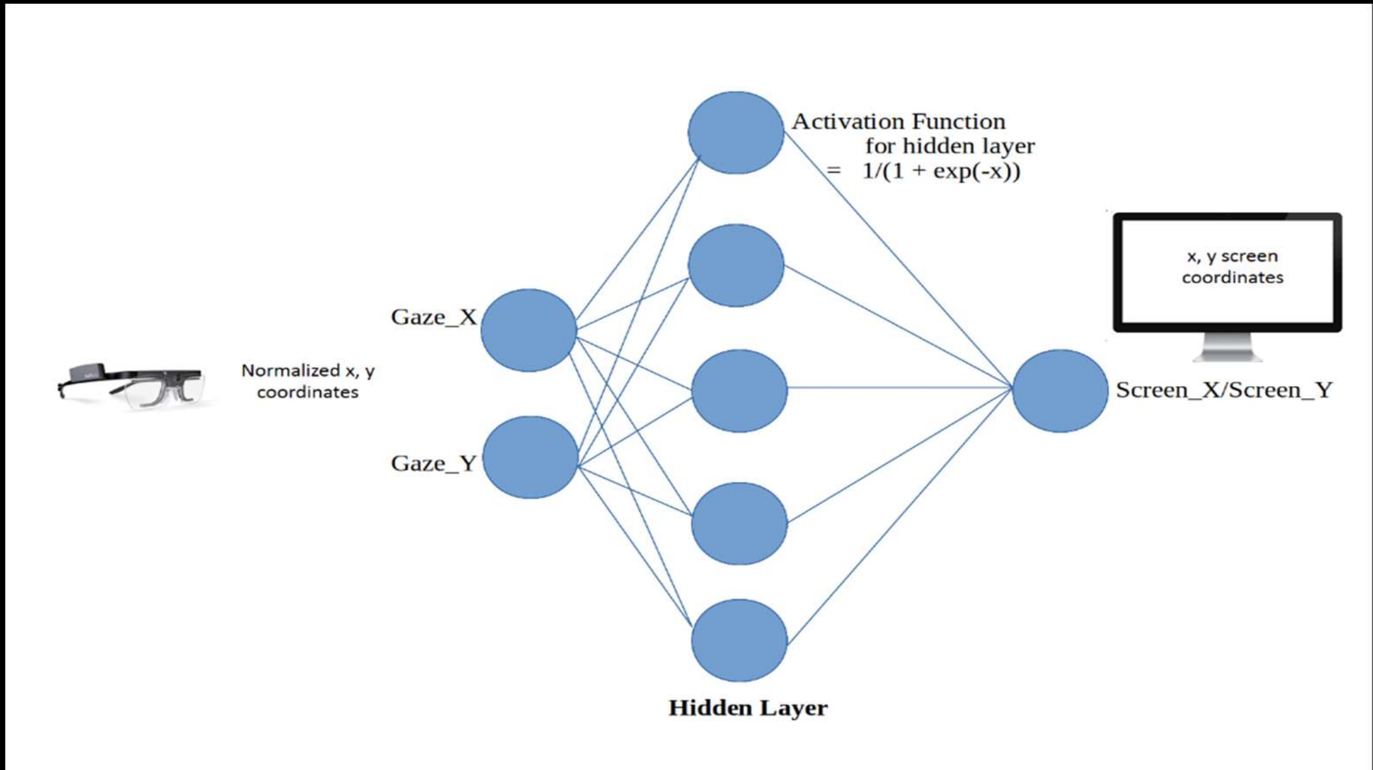


Exploration

- Compared ML systems to convert eye gaze coordinates to screen coordinates on windshield
- Set up Linear Regression and Backpropagation Neural Network Models for
  - Predicting x-coordinate in screen from x coordinate recorded by gaze tracker
  - Predicting x-coordinate in screen from x and y coordinates recorded by gaze tracker
  - Predicting y-coordinate in screen from y coordinate recorded by gaze tracker
  - Predicting y-coordinate in screen from x and y coordinates recorded by gaze tracker
- Compared R^2 and RMS error
- Neural Network model worked better than Linear Regression

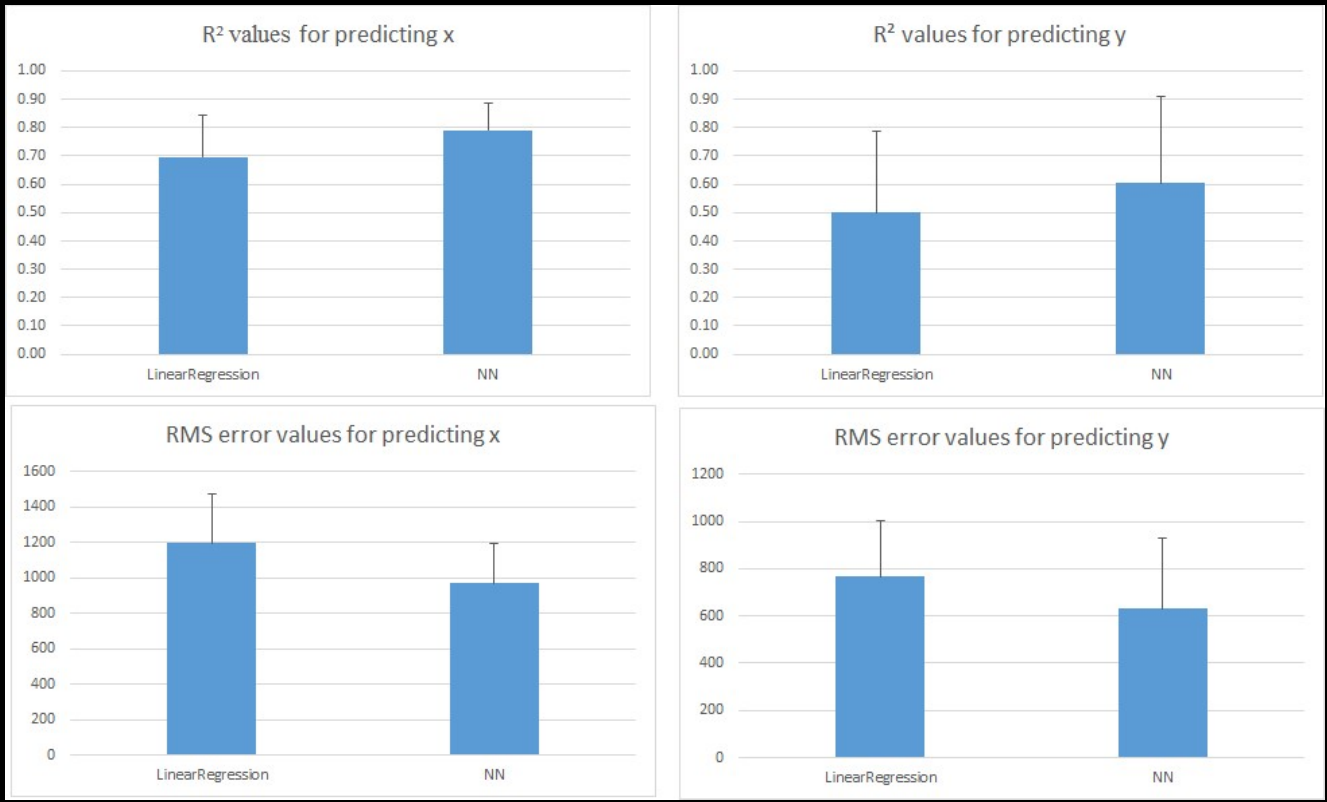


HMD



P. Biswas, Interactive Gaze Controlled Projected Display, Indian Patent Application No.: 201641037828

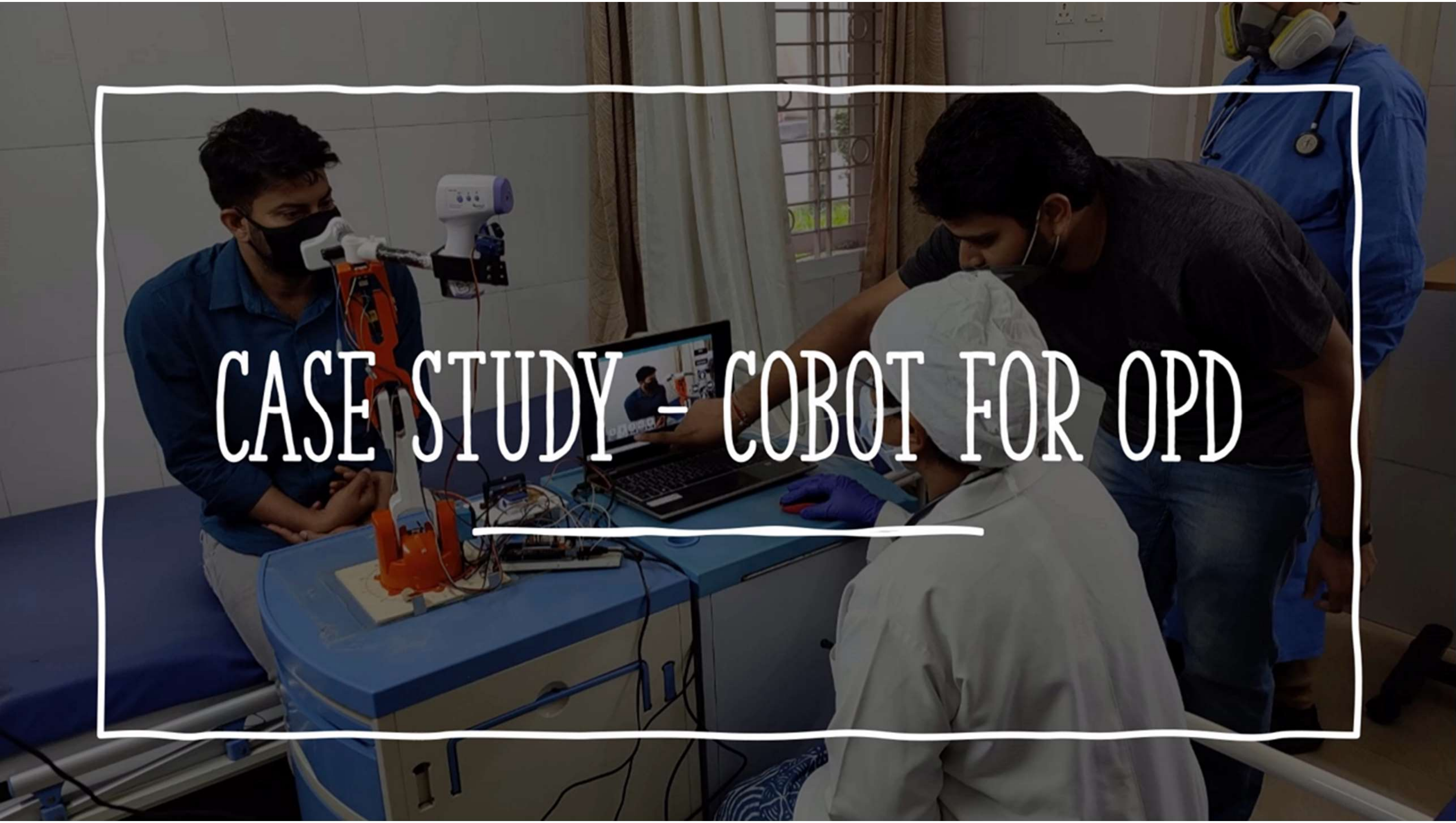
Comparison


R² values for predicting x
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
LinearRegression
NN
R² values for predicting y
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
LinearRegression
NN
RMS error values for predicting x
1600
1400
1200
1000
800
600
400
200
0
LinearRegression
NN
RMS error values for predicting y
1200
1000
800
600
400
200
0
LinearRegression
NN

Video Demonstration

2019
AERO INDIA
The Runway to a Billion Opportunities

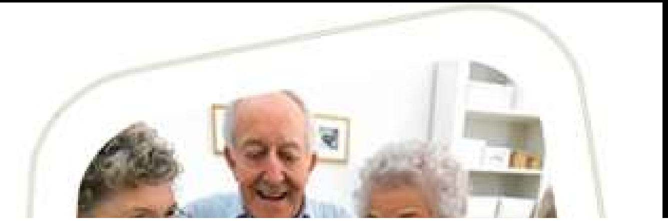
- LRD Murthy and P. Biswas, Deep Learning Based Eye Gaze Estimation Algorithms for Military Aviation, IEEE Aerospace 2022
- LRD Murthy, A. Mukhopadhyay, V Yelleti, S Arjun, P Thomas, MD Babu, KPS Saluja, JeevithaShree DV and P. Biswas, Evaluating Accuracy of Eye Gaze Controlled Interface in Military Aviation Environment, IEEE Aerospace 2020
- JeevithaShree DV, KPS Saluja, LRD Murthy and P. Biswas, Operating different displays in military fast jets using eye gaze tracker, Journal of Aviation Technology and Engineering 8(1), Purdue University Press, 2018

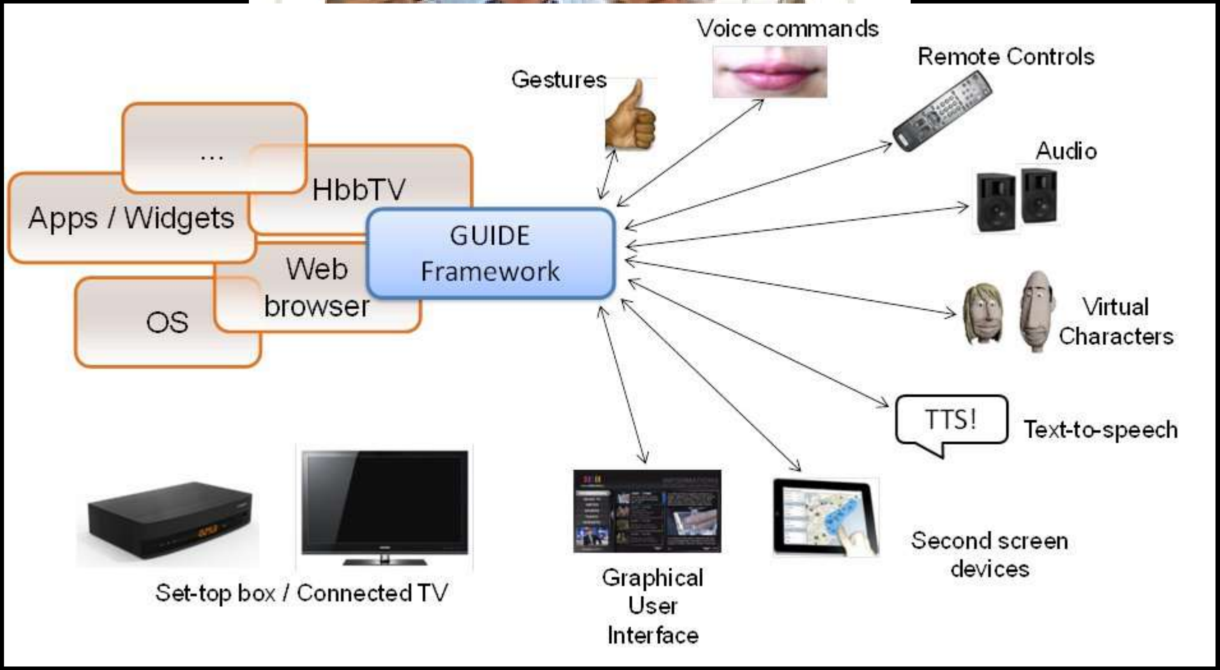

CASE STUDY – COBOT FOR OPD

Case Study 3

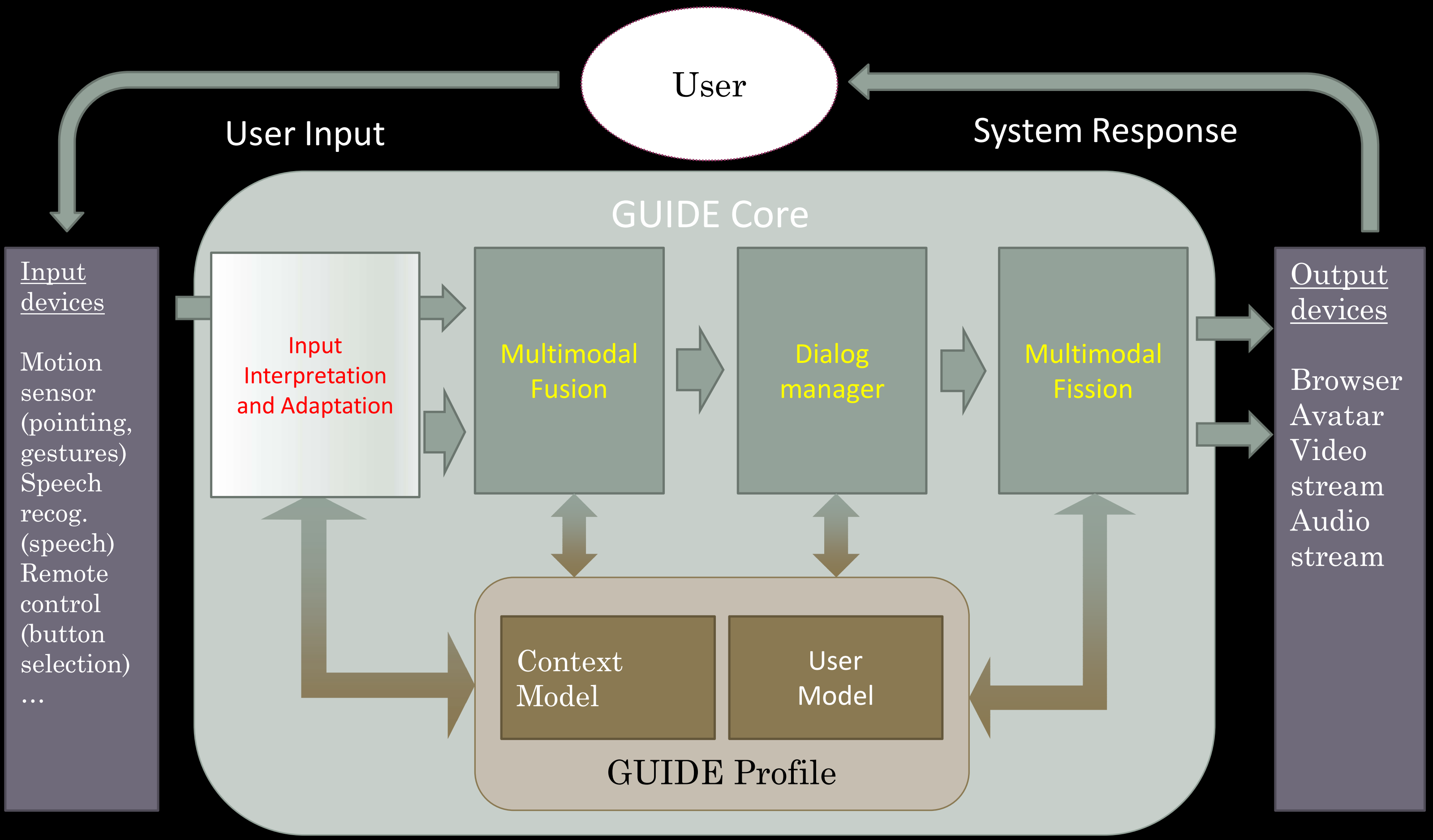
Digital TV Framework

EU GUIDE Project



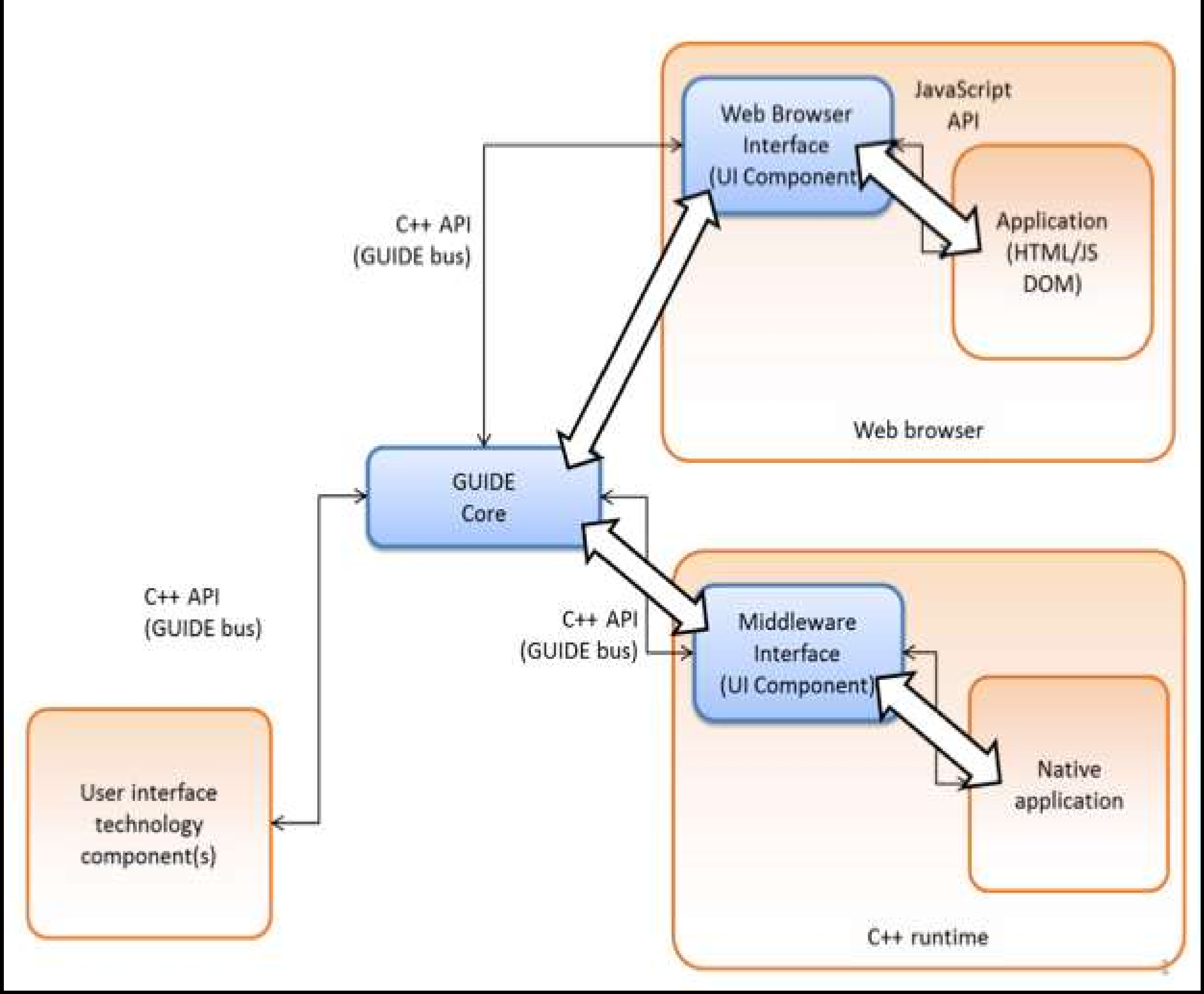


Conceptual Framework



31

Framework Application Interface



GUIDE
Adaptive Multi-Modal User Interfaces for Web Applications

GUIDE supports HTML5 web applications

More information on the GUIDE Open Source Software Framework: www.guide-project.eu

Take Away Points

- Introduction to multimodal systems
- Fusion strategies
- Case studies of fusing modalities including
  - Eye gaze
  - Finger movement
  - Speech input
  - Text-to-speech output
- Evaluation of multimodal systems