Human Computer Interaction

Multimodal Interaction

Dr Pradipta Biswas, PhD (Cantab)
Assistant Professor
Indian Institute of Science
http://cpdm.iisc.ernet.in/PBiswas.htm

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

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.

Basic Probabilistic Model

- Random Variables
  - Boolean
  - Discrete
  - Continuous
- Conditional Probability
- Expected everybody is familiar

Think: You are tossing a fair coin. The first three outcomes are head (incidentally). According to conditional probability of the prior events, what will be the outcome of all the next tosses?

Conditional Probability

Product Rule

$$P(a \wedge b) = P(a \mid b)\, P(b)$$
$$P(a \wedge b) = P(b \mid a)\, P(a)$$

$$P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$$

$$P(a \mid b) = \frac{P(b \mid a)\, P(a)}{P(b)}$$

Bayes' rule

- Relates prior knowledge (prior to present experiment) for taking present inference

- **Intuition:** Before we observe the data, the parameters are described by a *prior* which is typically very broad. Once we observed the data, we can make use of Bayes' formula to find *posterior*. Since now we can utilize some more facts in the form of evidences, the *posterior* is narrower than *prior*.

Bayes' Rule (2)

Posterior Probability

$$P(Y|X)=P(X|Y)*P(Y) / P(X)$$

Prior Probability

Y=Cause

X=Evidence or Symptoms

- Bayes' rule with more than one evidence

$$P(Y|X,e)= P(X|Y,e)P(Y|e) / P(X|e)$$

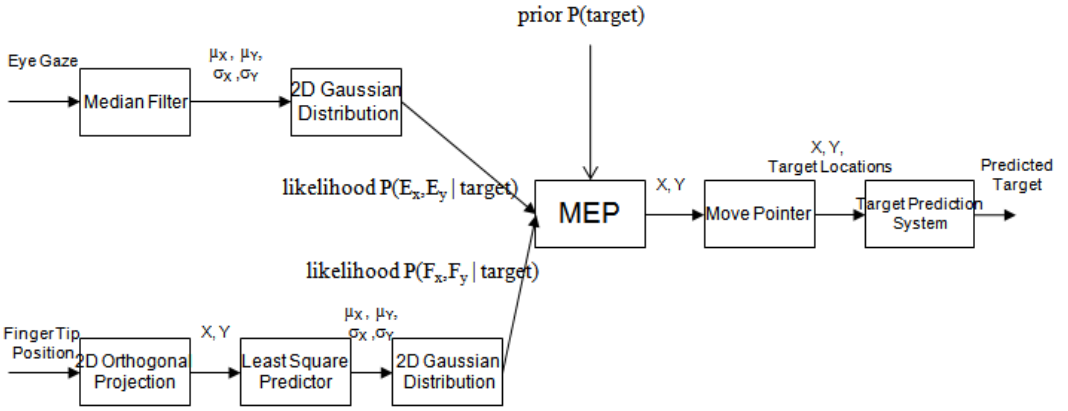
Example of Bayes' Rule

- We perceive an object
- We stretch arm to touch / gasp it
- If we have sensors to track eye gaze and finger movement can we predict movement ?
- Bayes' Rule can help in sensor fusion

Assumptions for a 2-D screen

- Sensors are inaccurate
- They generate x, y values inaccurately
  - We assume they return a random 2-D Gaussian distribution
    - With mean x, y
    - Std dev depends on accuracy of the sensor
- Prior probability tells us which area of screen (or which object) is more likely to be selected

Sensor Fusion using Bayes' Rule



13

Case Study 1

Multimodal Gaze Control

Objective

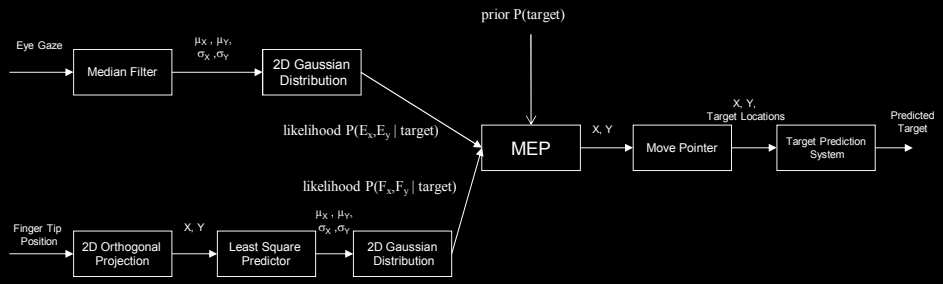
- Exploring use of non-invasive gaze-controlled interface and finger tracking technology in automotive environment
- Comparing different multimodal fusion strategies

15

Secondary Task- Point & Select

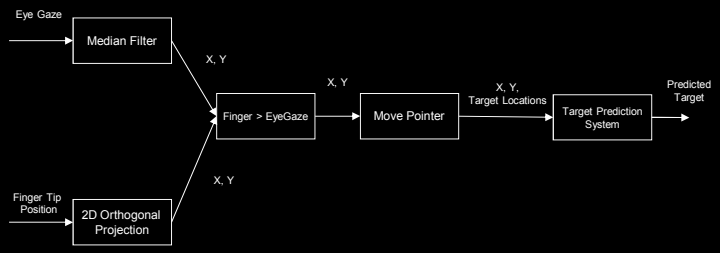
16

MultiModal Bayesian (MMB)



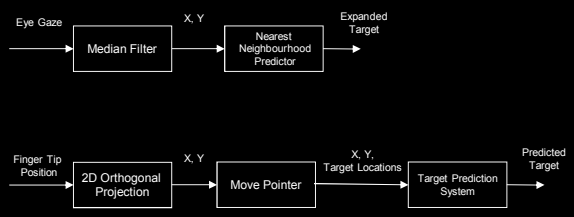
17

MultiModal



18

Finger Tracking



19

Design

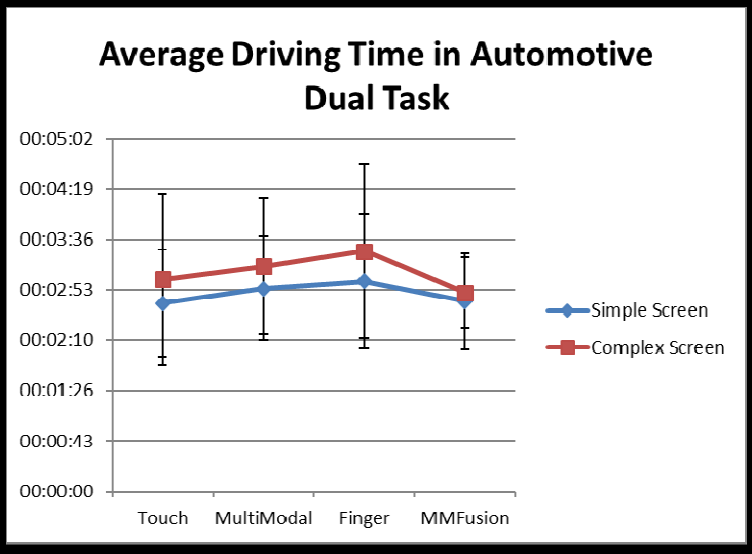
- Dual Task Study
  - Primary task was driving
  - Secondary task was point and selection
- Repeated Measure 2 × 4 Design
  - Two screen layouts
  - Four different interaction strategies
    - Touchscreen and 3 different finger / eye-gaze tracking

20

Primary Task

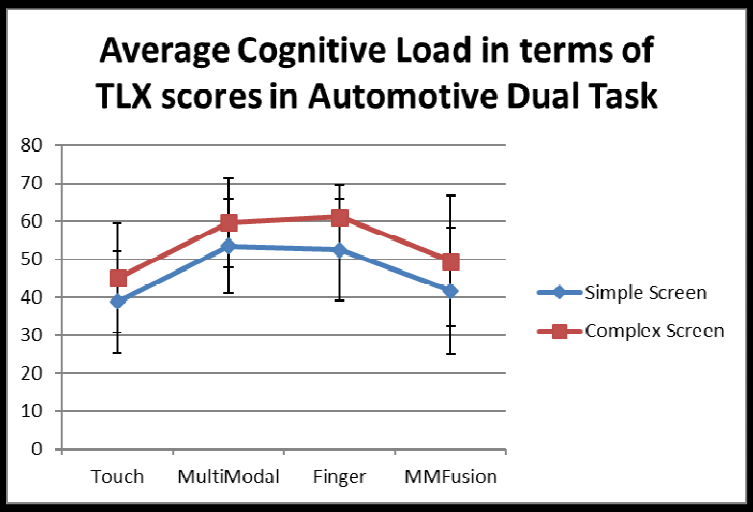
21

Track Completion Time



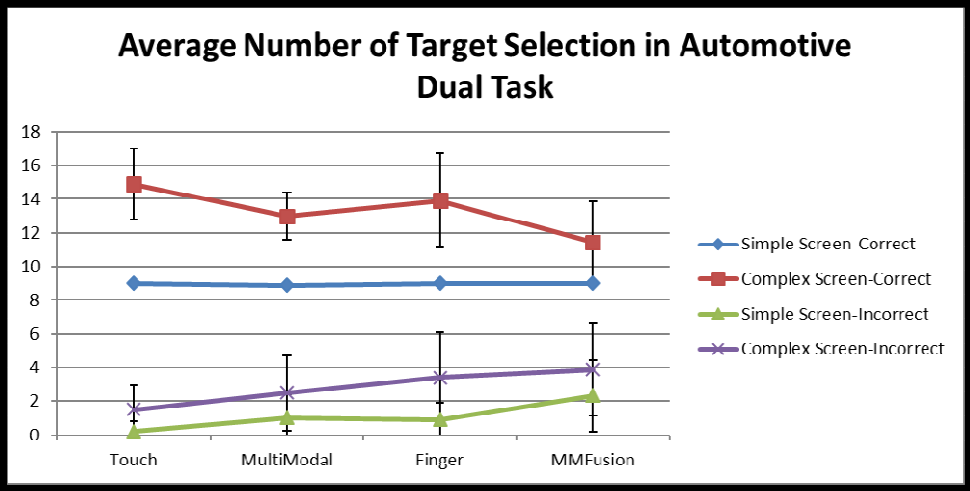
22

Cognitive Load



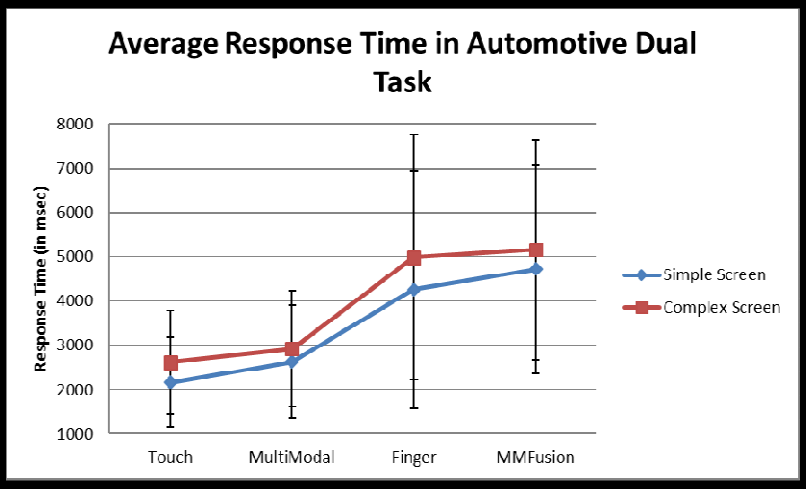
23

Number of Selections



24

Response Times



25

Summary of Results

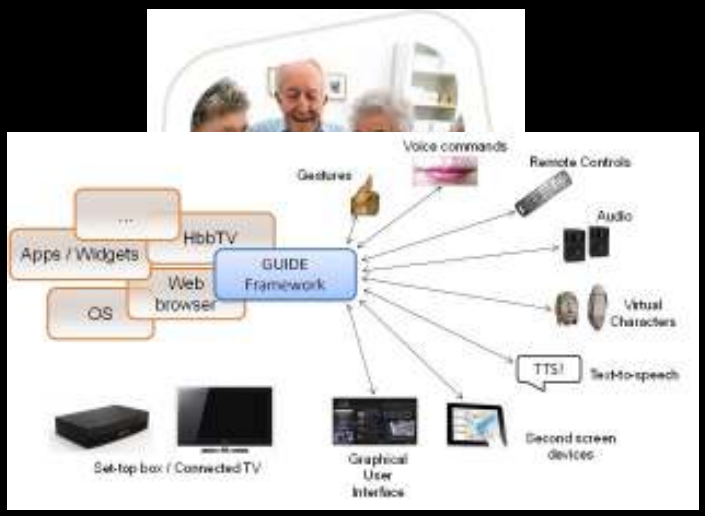
- Using eye gaze for pointing
  - Either reduced driving time
  - Or reduced response time than finger tracking based pointing
- We did not yet have the best fusion strategy
  - The Bayesian Fusion improved driving time and reduced cognitive load but the naïve multimodal system reduced pointing and selection 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

26

Case Study 2

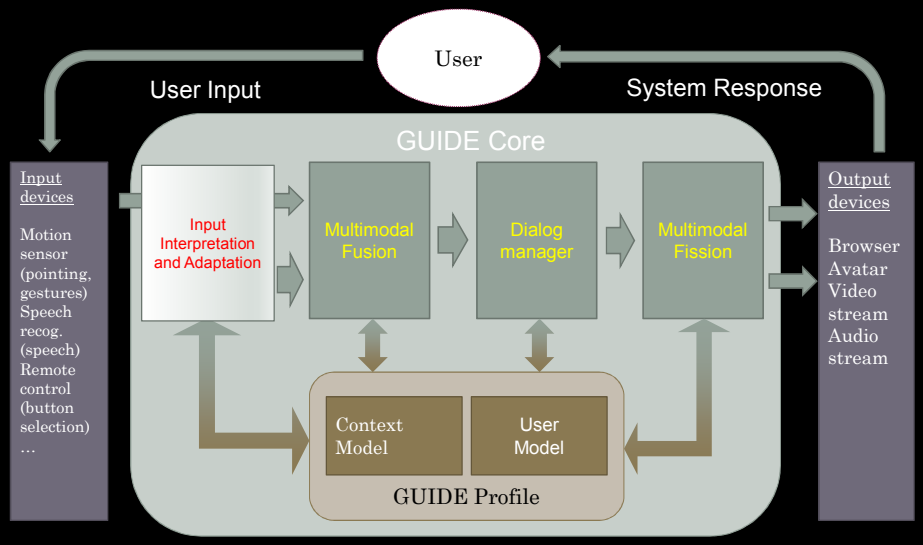
Digital TV Framework

EU GUIDE Project



28

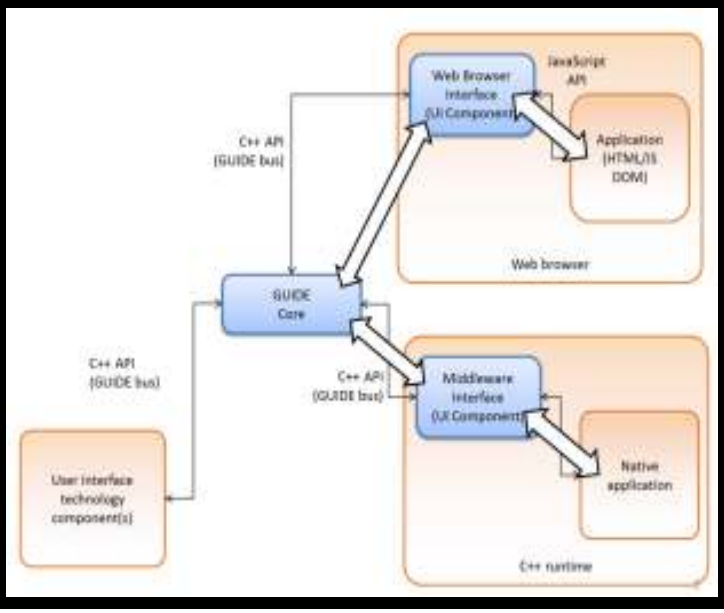
Conceptual Framework



29

29

Framework Application Interface



30

30

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