

Vision PSU

Capstone Project

Team Members:

Mohammed Hoque
Troy Tancraitor
Jonathan Lobaugh
Lee Stein
Joseph Mallozi

Pennsylvania State University

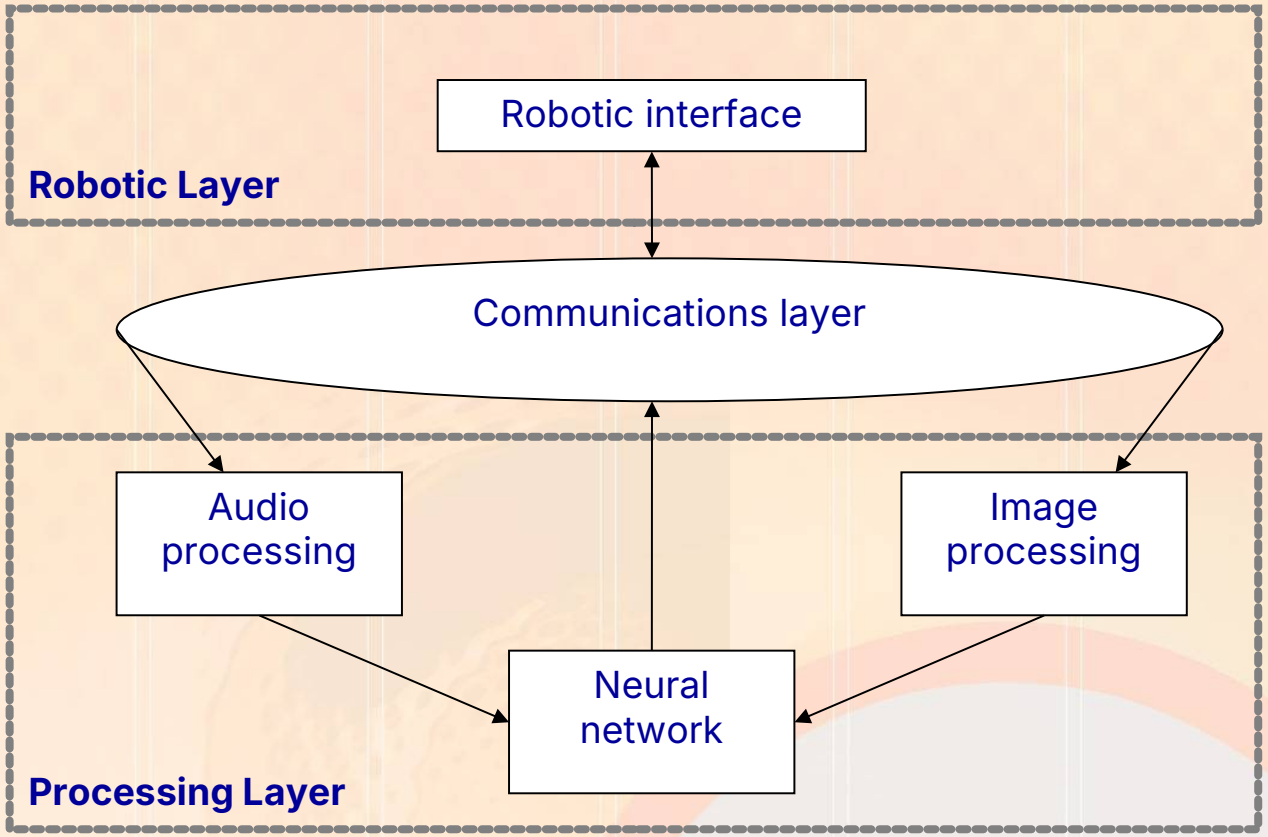
Presentation Overview

- Problem Statement
- Architectural design
  - Robotic layer
  - Processing layers
  - Communications layer
- Testing and Results
- Realization of Requirements

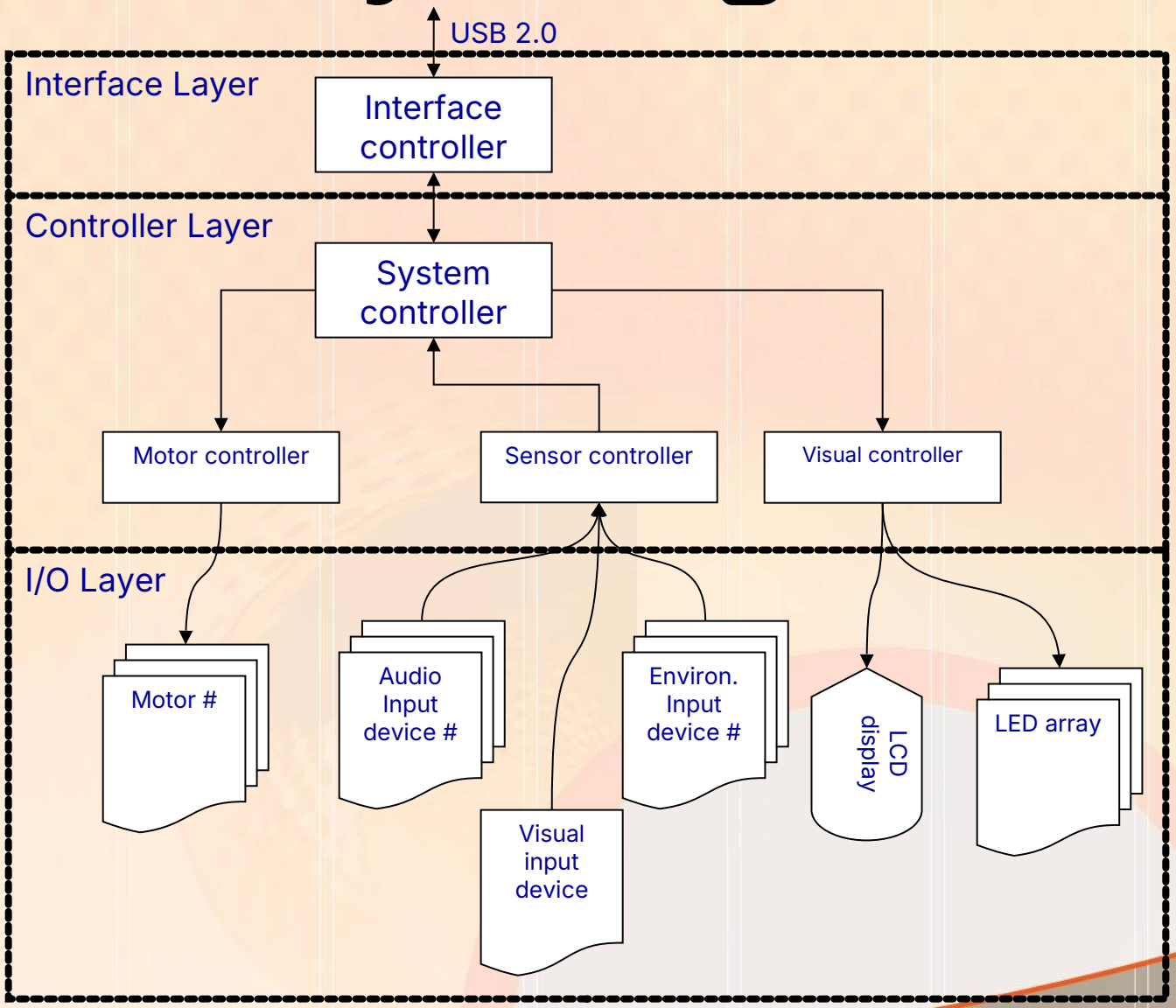
The Big Picture

- We built a robot with vision and hearing capabilities.
  - **Objectives:**
    - The robot must be able to detect a specific user from a known set of users, based on audio and video information.
    - The system should facilitate an improved platform for human computer interaction.
  - **Limitations:**
    - The recognition is solely limited to the five group members and project advisors.

High-level Architectural Design



Robotic Layer Diagram



Processing Layer Diagram

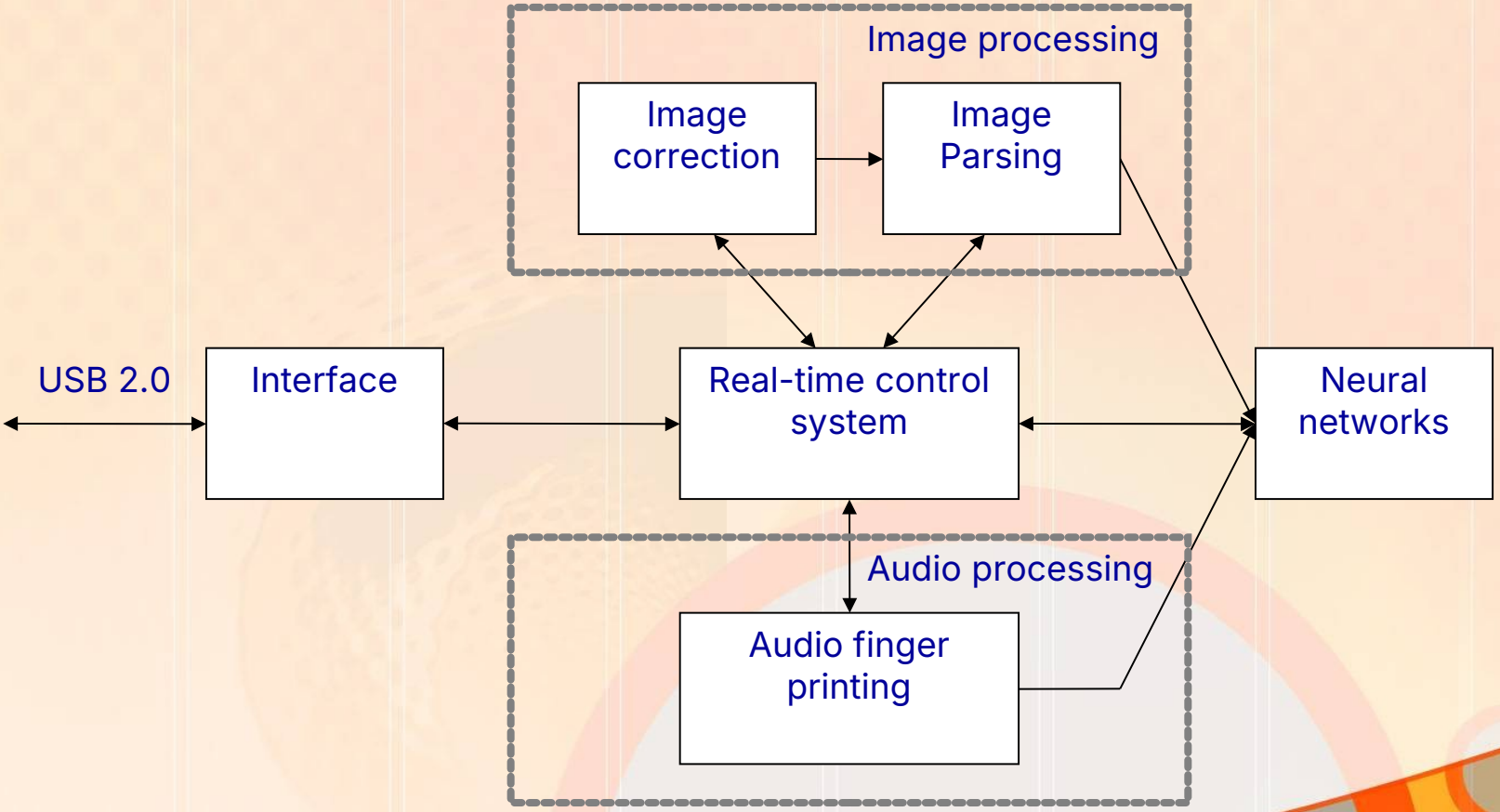
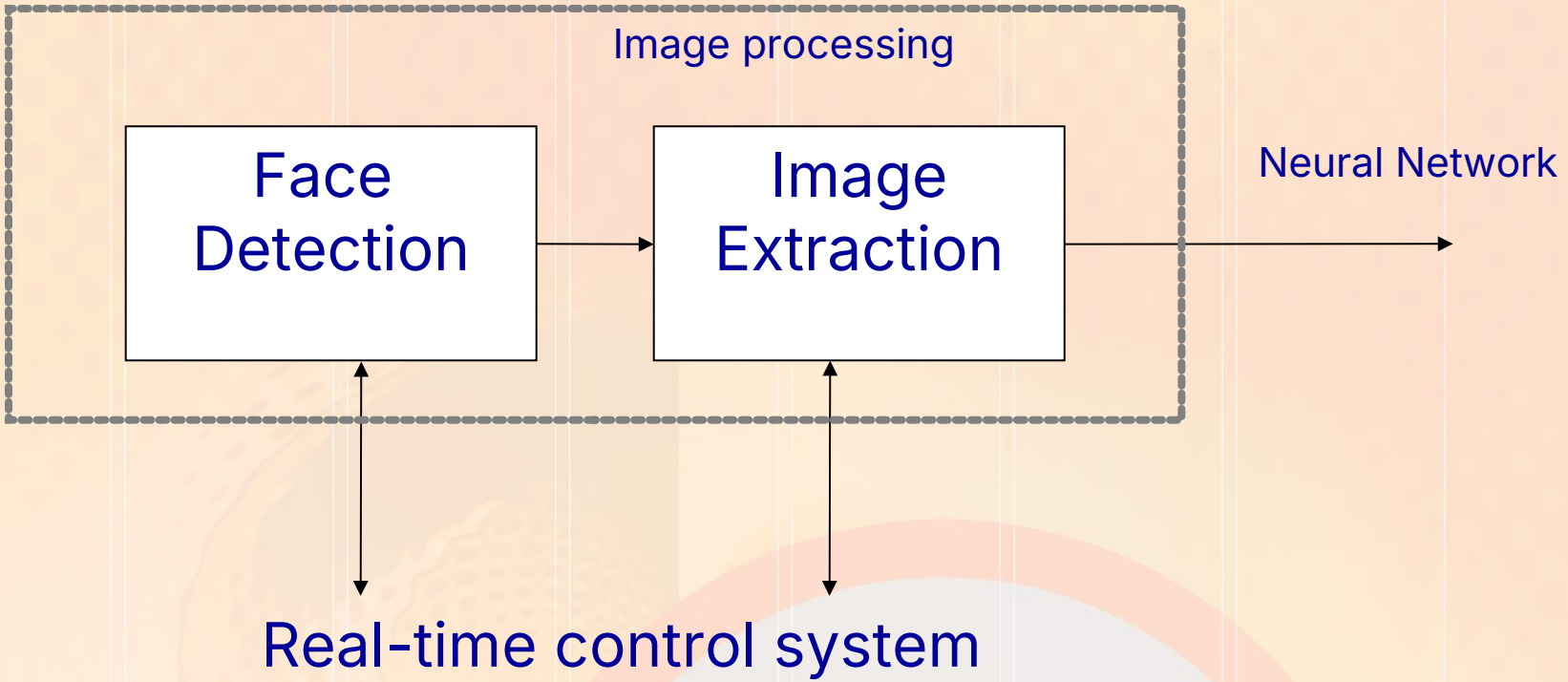
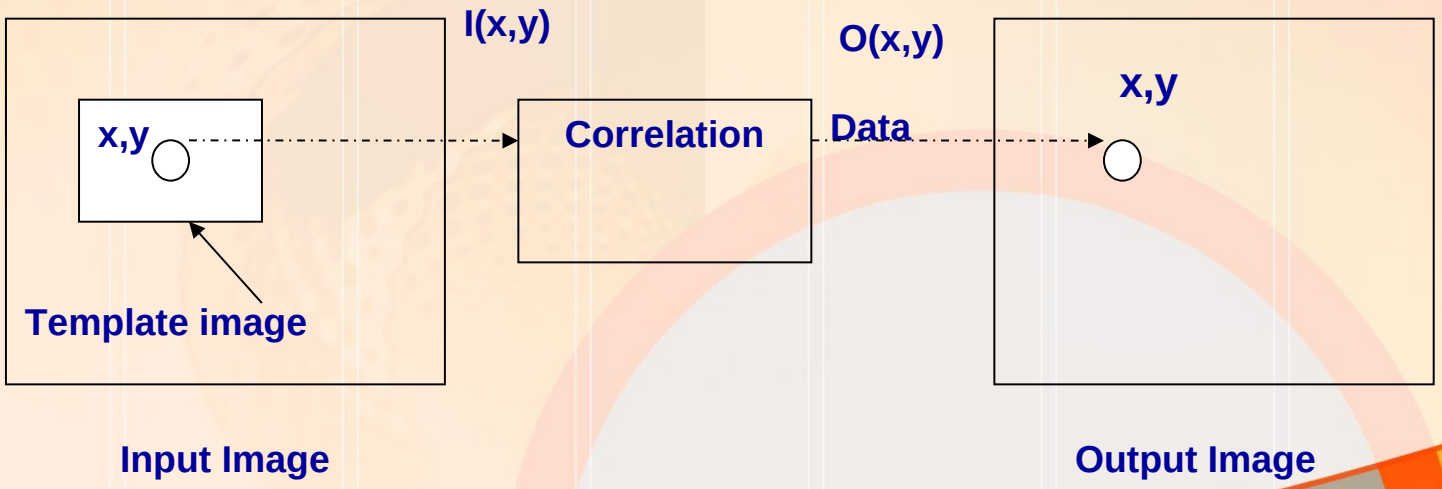


Image Layer Diagram



Facial Vector Extraction Process

- The location of the distinctive features of the face are identified using template matching, and the resultant position vectors are fed to the neural network.



Input Image



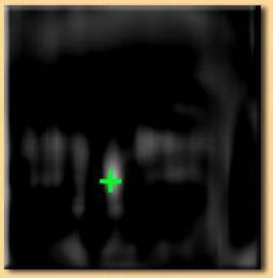
Input Template Patterns







Detected Center Point Positions

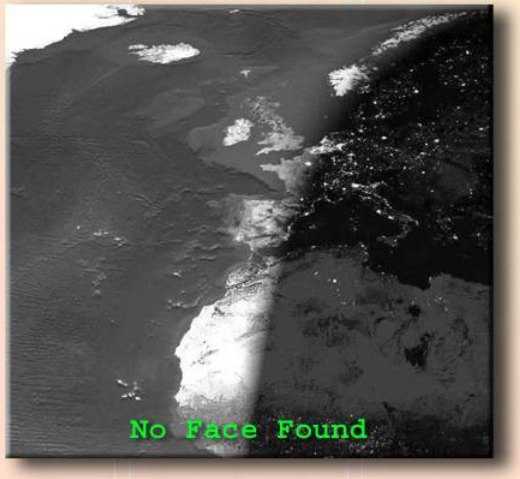






Template Matching Feature Extraction Example









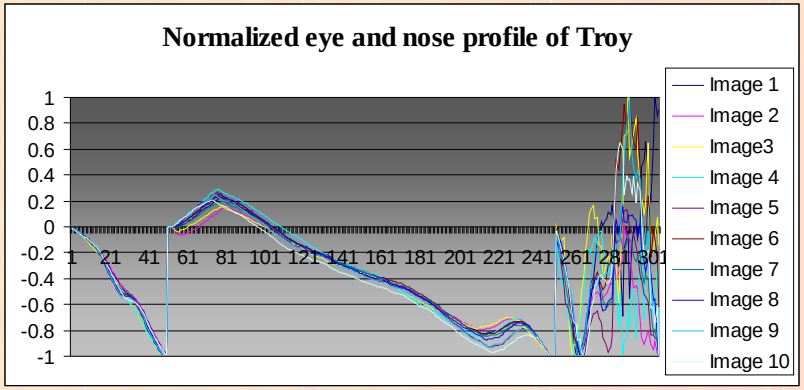


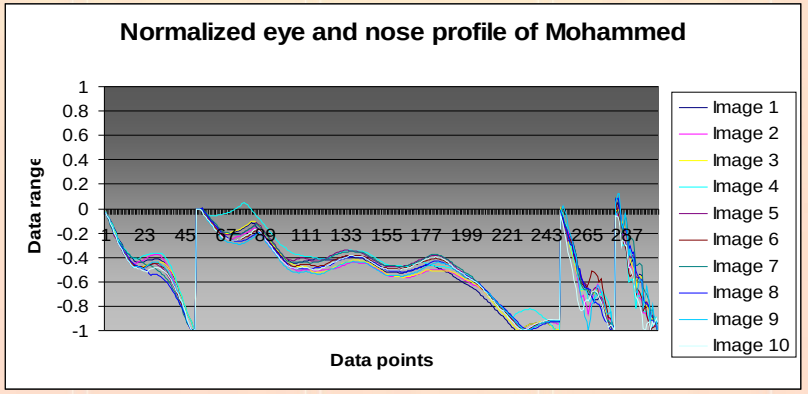


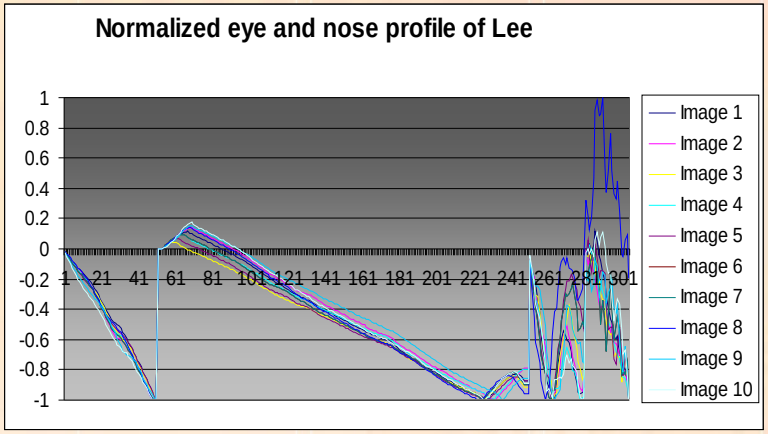


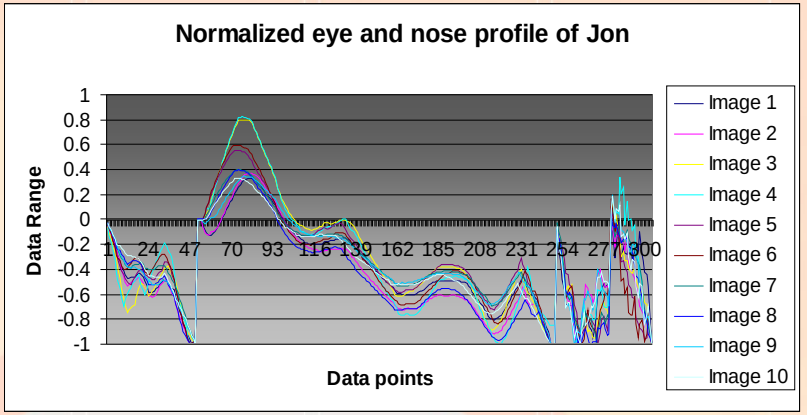
Face Processing Results

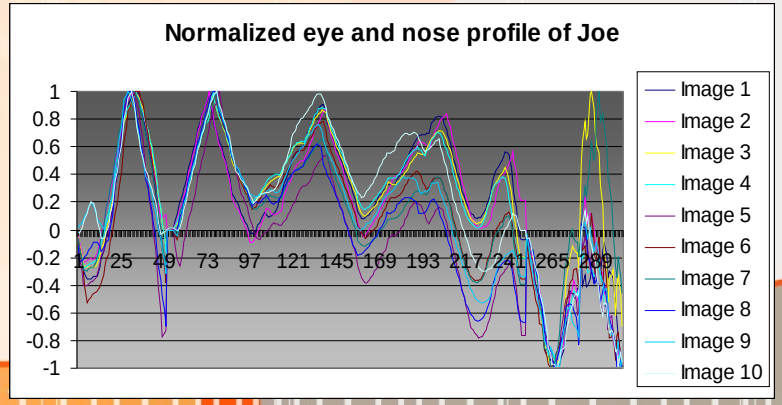
Eye and nose profiles





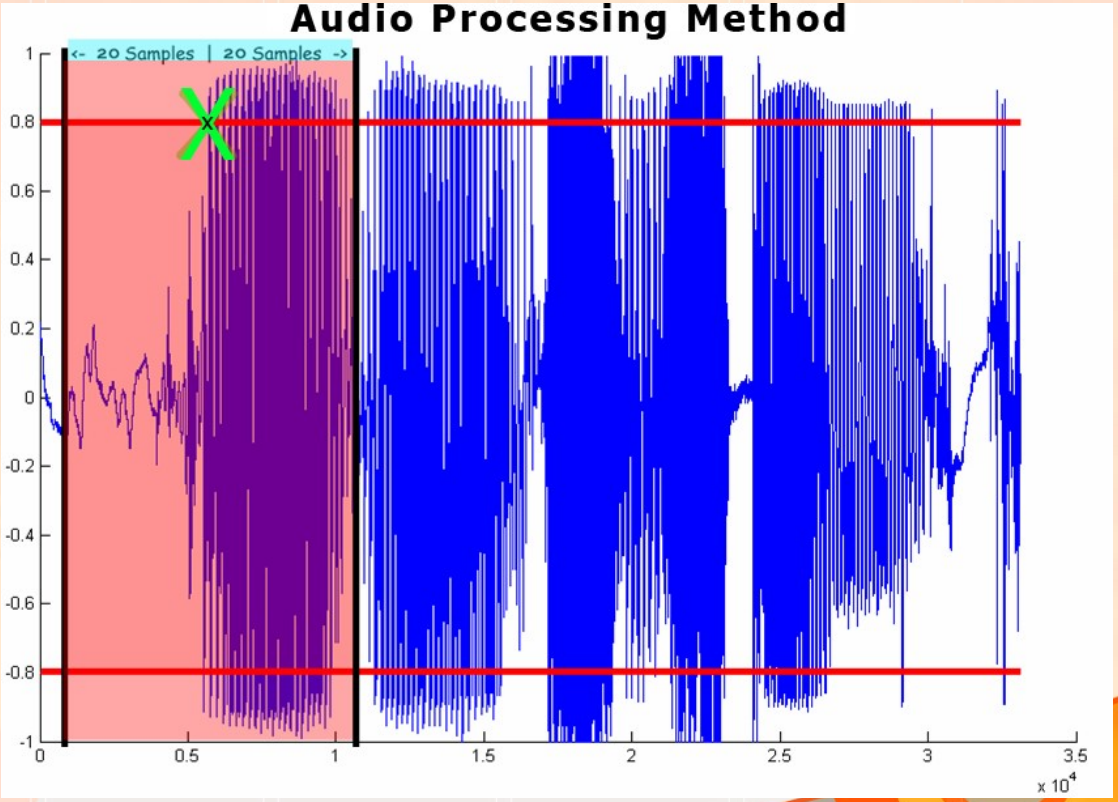






Audio layer

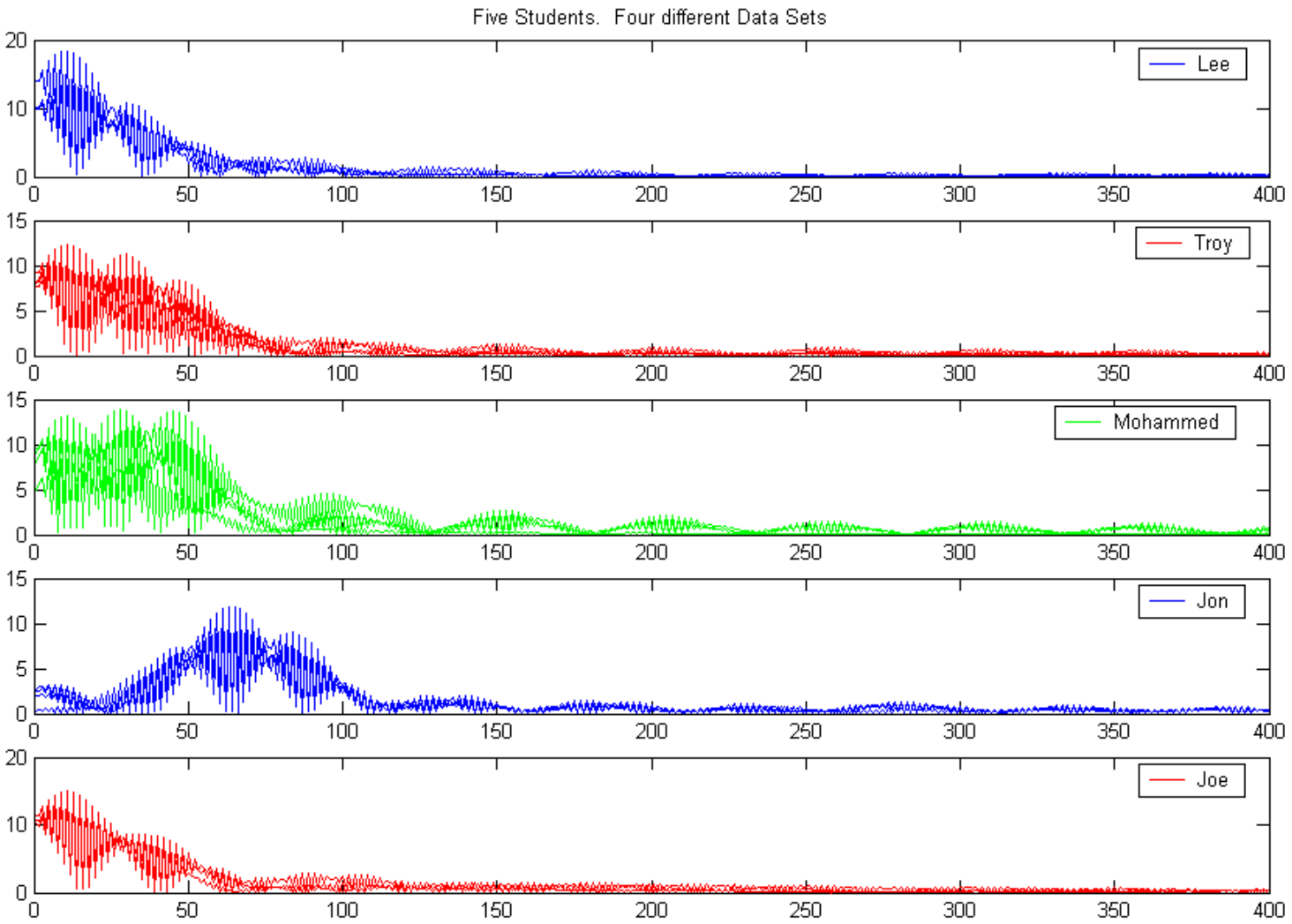
- Finds the normalized highest amplitude of the waveforms (should be less than or equal to 1).
- Looks for amplitudes greater than the 80% of the highest amplitudes.
- Creates an array of amplitudes of 40 points after finding the first one crossing the 80% boundary.



Audio Layer (cont)

- Pads secondary array with ending zeros (984) to get a 1024-point array.
- Performs a Fast Fourier Transform on 1024 points.
- Sends the first 400 absolute points to neural-network for processing.

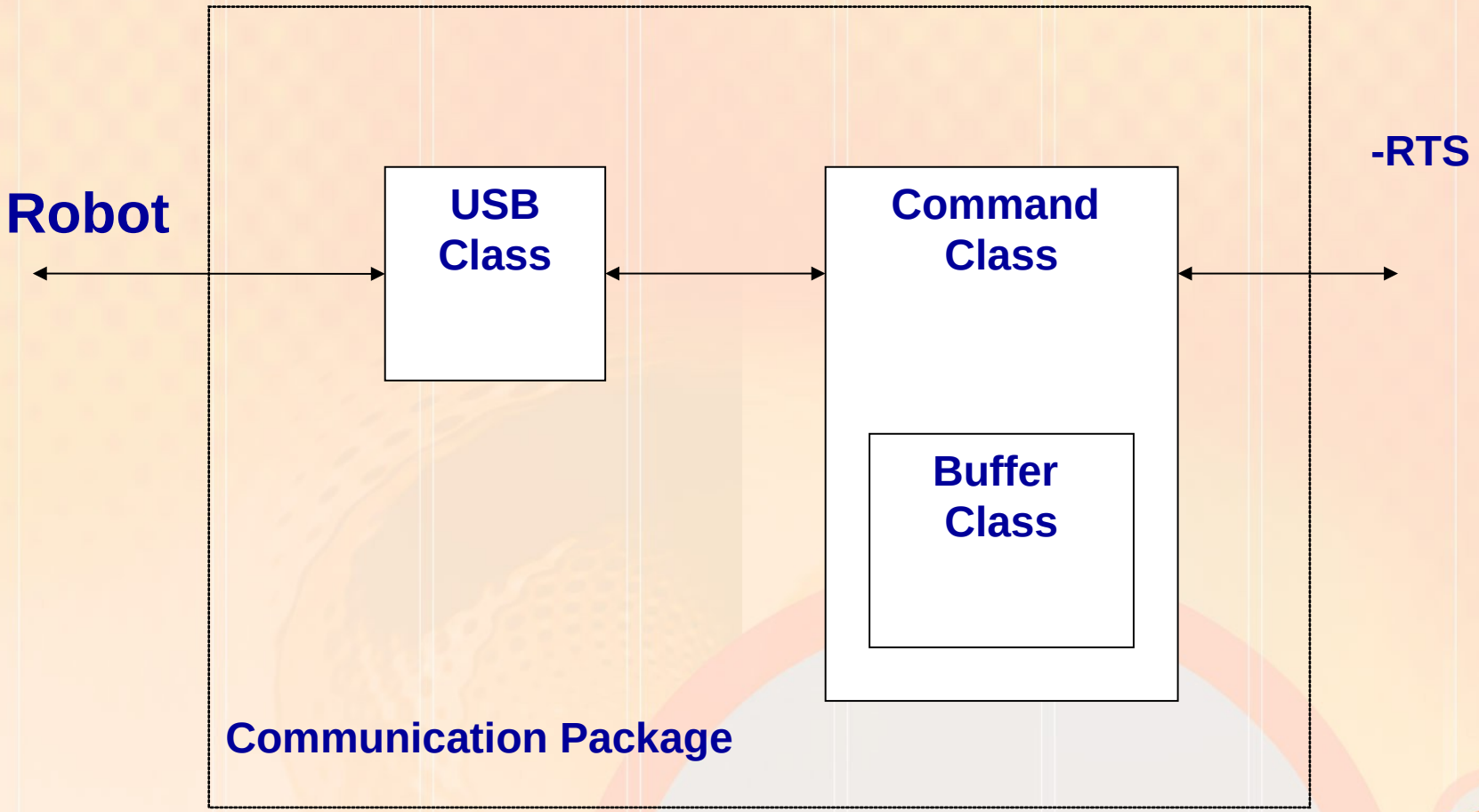
Audio results



Neural network layer

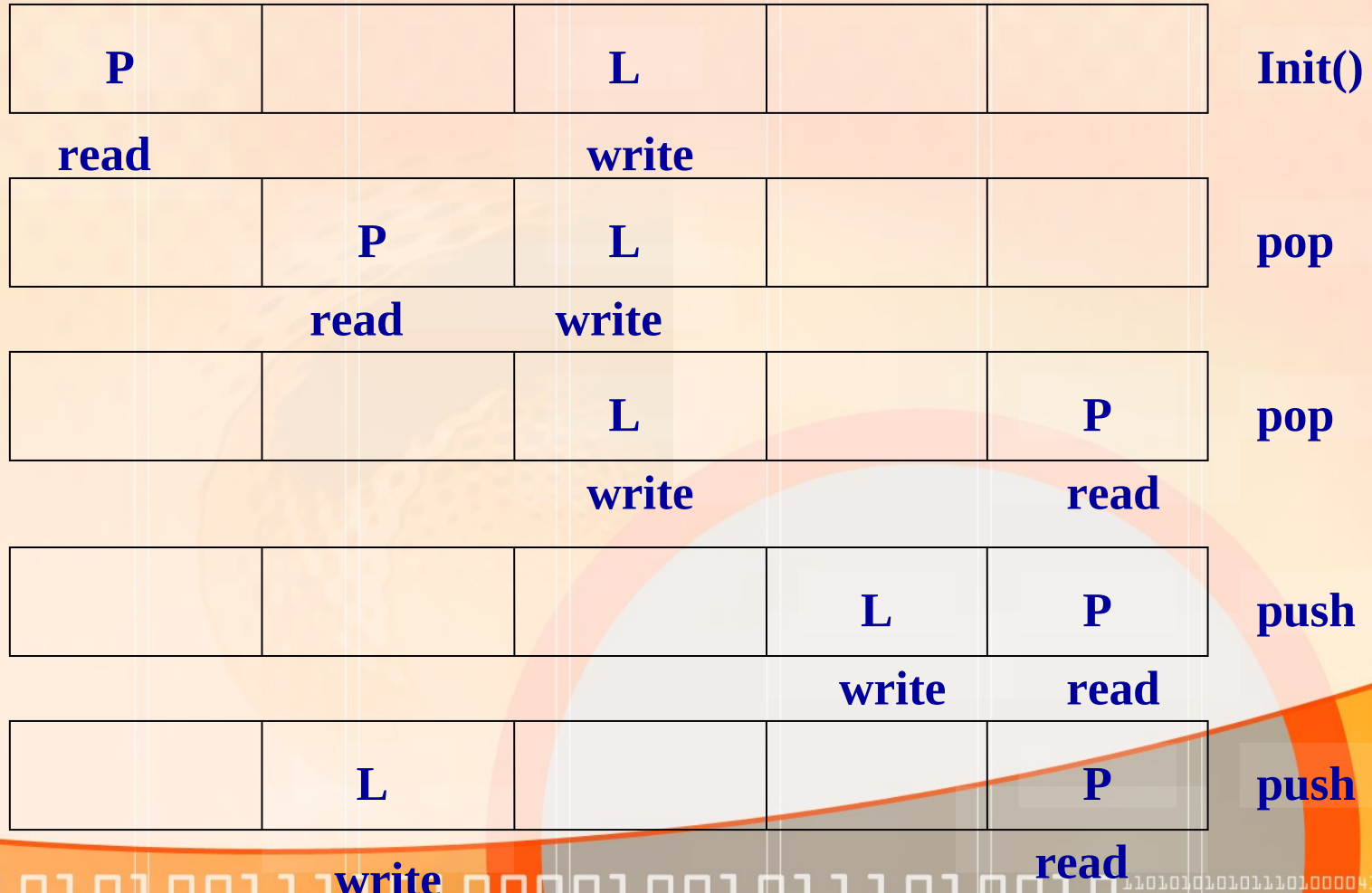
- Feed-forward back propagation 3 layer network
  - Neural networks provide a way of allowing the system to generalize the input data and determine the output.
  - Two separate networks; one trained for audio data, the other trained for image data.
- Inputs
  - Vectors containing facial feature position information (Eyes and nose)
  - Audio vector of 400 known FFT points
- Output
  - Percent similarity of known users based on audio/ imaging
- Utilized Matlab's neural network toolbox to train the system weights.

Communications Layer



Lock and Key Algorithm

- Setup in the image and audio arrays.
- Developed to prevent data collision.



Communications Layer (cont.)

- Controls the flow of data between the robotic layer and the processing layer.
- Communicates with the devices.
- Command class interfaces all of the software packages
- Interprets and sends commands to the robotic layer.

Real-time control system

- Manages input and output of data to the interface layer
- Determines robot's actions during idle processing time
- Controls robot to follow a face once detected.
- Controls programmed responses to robotic interface; audio, optical, and mechanical

Realization of Requirements

- Ease of Use
  - Centralized control
  - Minimize connections needed (connections <=2)
  - Standardized adaptors
- Cost
  - School funded expenses = $102.00
  - Team funded expenses = $420.00
  - Previously owned = $200.00
  - Total = $722.00
- Efficiency

  Performs audio and image processing in less than .5 seconds.

Realization of Constraints (cont)

- Economical

  Total expenses exceeded the initial budget of $250.

- Environmental
  - Audio System is sensitive to background noise.
  - Imaging system works the best in ideal lighting condition.
  - It also works the best when the user directly looks at the camera from a reasonable distance (3-4 feet).

- Accuracy
  - The overall accuracy rate of our system is close to 80% which is what we initially had set up.

Conclusions / Questions

The next evolutionary step would be to expand the interaction between humans and machines to a social interface. This system is created to help bridge the gap between humans and computers.

Someday this distinction may change the meaning of computers to our everyday lives.