Social Therapy Robot

Arun Devarajulu
Manohar Anumolu
Pablo Sanhueza

Why a social therapy robot?

- Average cost to educate a child with Autism is approx. **$20,000/year**
- Incidence of Autism: **1 in 68** children
- Children with ASD are characterized by difficulties in **interaction, communication, and attention**: Therapy sessions with engaging tools are used to maximize long-lasting efficiency.
- When you are interacting with a person, there are a lot of social cues such as facial expressions, tonality of the voice, and movement of the body which are overwhelming and distracting for children with autism.
- A Robot offers a simple, predictive and stable interaction that is favourable to children with ASD.
- Goal: For the **Social Robot to become an assistive agent for children with learning disabilities**, leading them to adapt to social situations and to their integration into society.

User Requirements

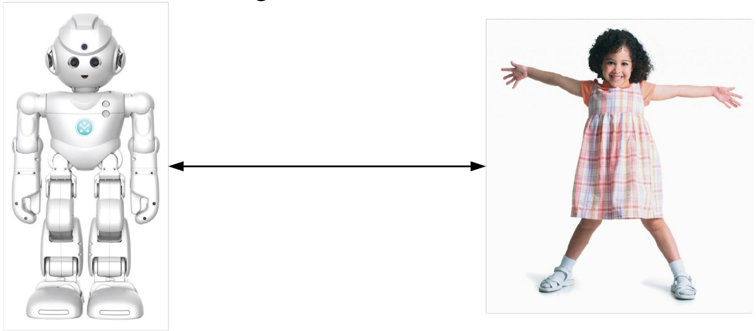
- ★ Behavior Analysis - Compare the speed, accuracy and frequency of user responses to **measure the scope of learning disabilities**.
- ★ **Exercises** for **children to improve their Social, Sensory, and Cognitive skills**
    - Practice/teach through games, instructional videos or other means to:
        - **Improve on eye-contact, friendliness, body-language**
        - Learn and Model Good Social Behavior
- ★ Gather information/data regarding eye contact, facial orientation, emotion, voice stress. This can help in the initial diagnosis and act as performance metrics to track improvements in the child
- ★ **Playful design** - Needs to grab child's attention and trust
- ★ Ability to remember different patients/users (**User Identification**)
- ★ Adapt and engage in **natural conversations**
- ★ Robot usability for ages of **3-14 years** old
- ★ Engage in a **safe and controlled environment** with the user

Robot Requirements

- ★ **Anthropomorphic** design (upper body) for interaction with children
- ★ **Wheeled Lower body** for cost-effective mobility
- ★ Power Source to run the robot for multiple sessions - Battery
- ★ Camera to analyze and gather information (duration of eye contact, facial orientation detection, simple gesture analysis)
- ★ Ultrasonic sensor, accelerometer for motion detection, control
- ★ Microphones, Speakers for Natural verbal interaction, voice analysis using Deep learning
- ★ Model emotions during tasks to elicit good social behavior from user using LED display via deep learning
- ★ Interactive, Detachable Tablet + Stylus module for task completion (various task modules to help develop good social behavior) by user
- ★ Network communication for data transmission
- ★ Cyber-resilient software design for security purposes
- ★ Event Logging to troubleshoot, debug and investigate system issues

High Level Functionality Design

Interaction between robot and child with audio & video feedback to model good social behavior.



Improved Social Interaction

Mechanical Design

The Components Include:

- ABS Polymer Chassis: Flexible, Cheap, Lightweight
- Exterior: Plastic Injection Molding
- Stepper Motor for Locomotion
- 5(R)+5(L) Arm Actuators, 3 on the shoulder, 2 at the elbow
- 2 Actuators at the Head and the Hip
- Battery for Power Source for easier movement during learning tasks

Segway Advanced Personal Robot

Daxter

Motion Control for Gestures

- Control of the arms head and hip - Use simple predefined configurations for general expressions of emotions, feedback from user is achieved via pressure sensors, and Kinect sensors for movements, gestures, and behavioral responses.
- Feedback from the kinect sensors also helps in ensuring user safety in case of any collisions.

Perception

Camera Video Feed is used as Input for the following tasks:

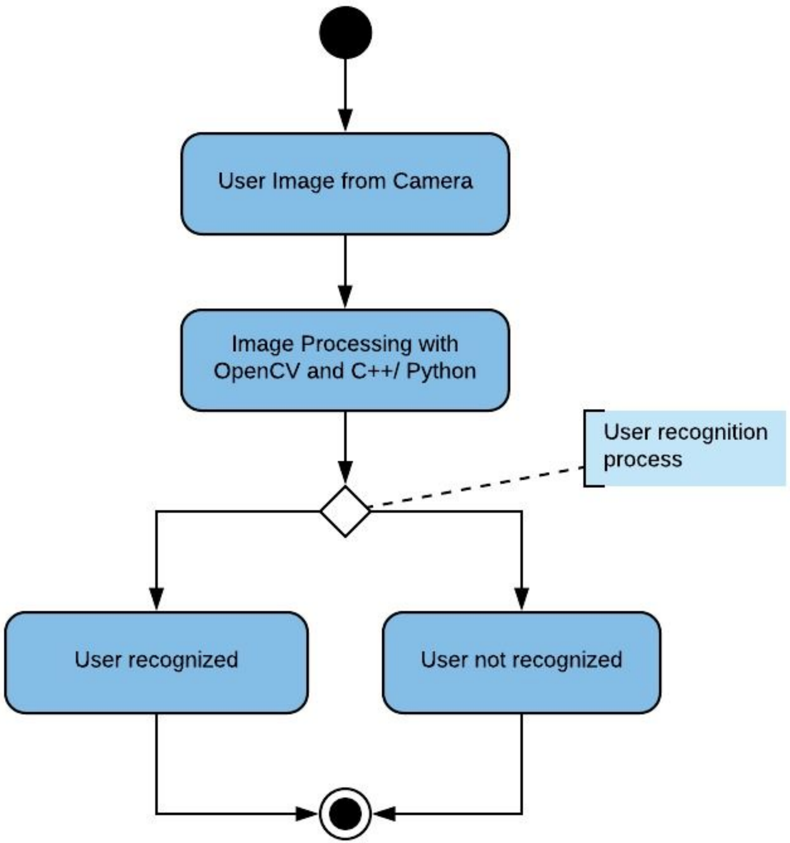
1. User Identification
2. Eye Tracking
3. Facial Orientation
4. Emotion Analysis

Voice Input to be used for Conversation (using Deep Learning) to interact with the user while measuring the following metrics:

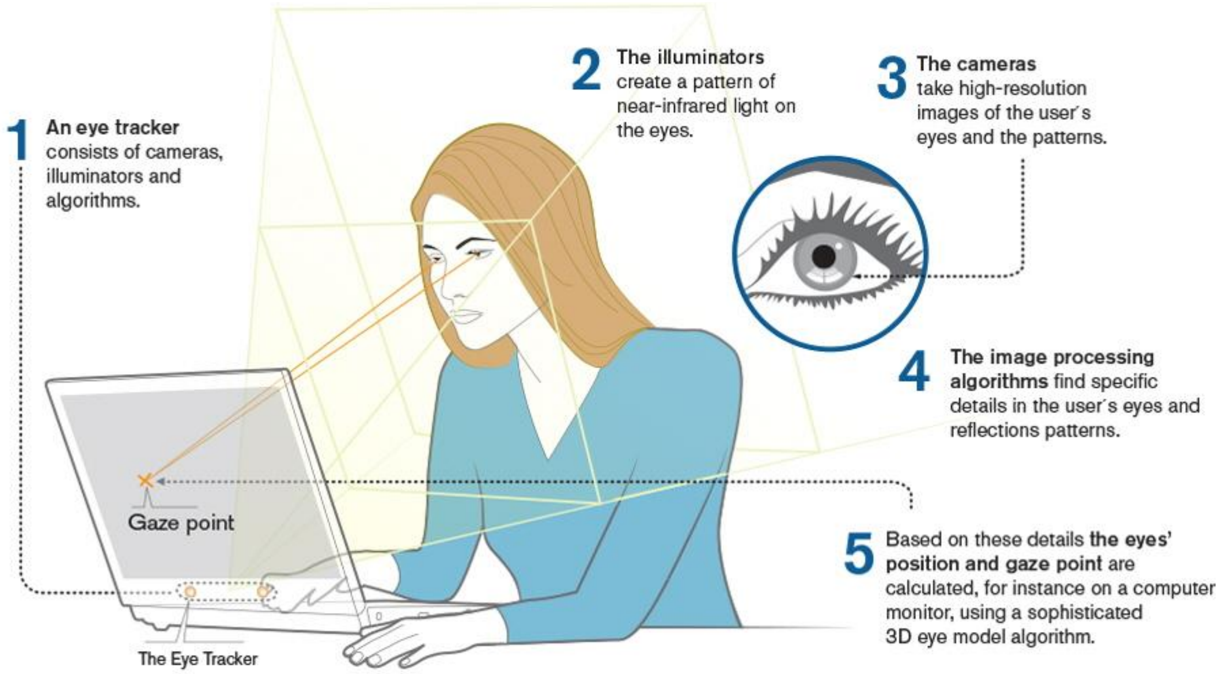
1. Speed, Accuracy and Frequency of User response
2. Voice Stress Analysis

User Identification

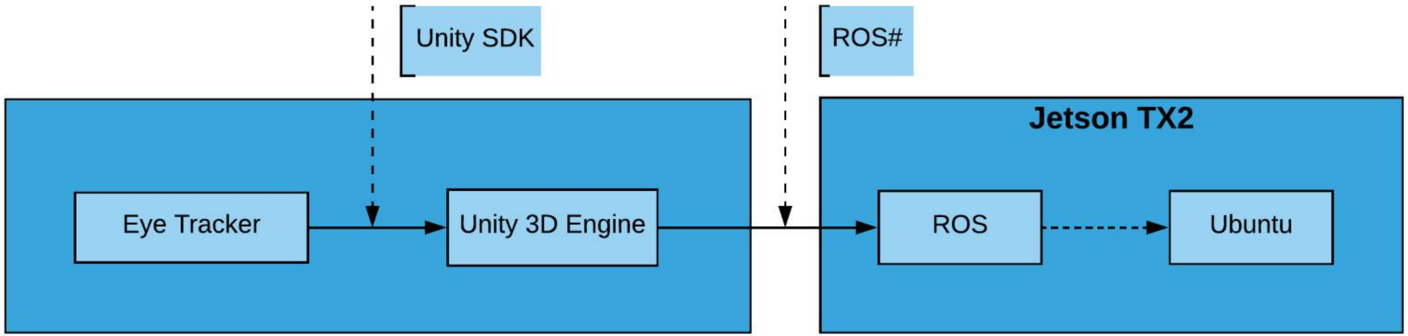
- ❑ All user pictures are entered in the database by the therapist
- ❑ SVM based on user pictures will be trained before start of therapy session using OpenCV, C++, and QT framework
- ❑ When user faces the robot main camera, the image is classified as known user or unknown user.
- ❑ If the user is known, further interaction is enabled.



Eye Tracking

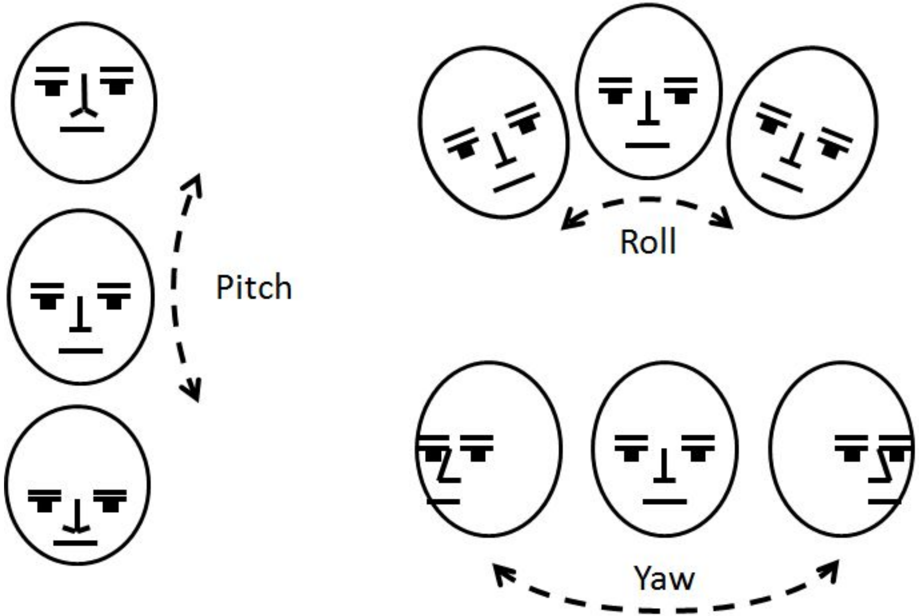


Eye Tracking - Design



- ❖ Interactive Eye Tracker gathers gaze data from user and implements eye-tracking features such as GazeAwareness, GazeDuration, GazePoint Data (gives location of gaze) etc. These eye-tracking features give a constant stream of data in real-time.
- ❖ Using ROS#, a set of open source software libraries and tools in C# for communicating with ROS from .NET applications, in particular Unity3D, we can share and store Gaze-Related Data in Jetson TX2 for Eye-Tracking Analysis. The Jetson TX2 Board runs Ubuntu which has ROS installed.

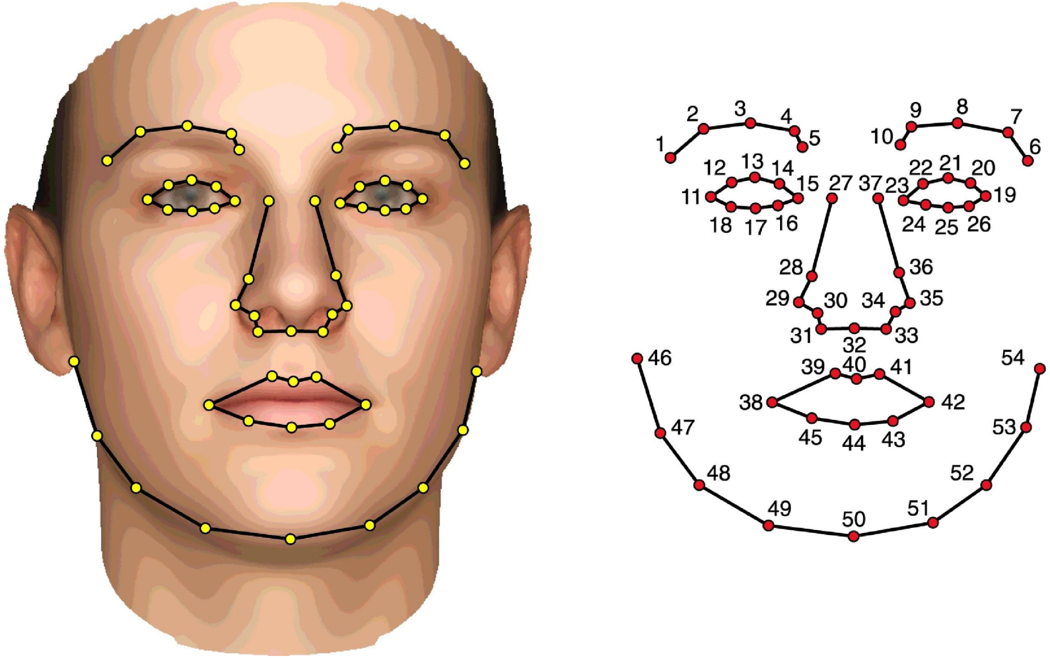
Facial Orientation



- Facial Orientation is a performance metric to analyse head movements (Roll, Pitch, Yaw).
- Compared with children without ASD, children with ASD exhibited greater yaw displacement, indicating greater head turning, and greater velocity of yaw and roll, indicating faster head turning and inclination.
- This helps us to track improvements throughout sessions, by comparing them to a baseline performance metric (the baseline can be that of an average normal childs, or the actual initial diagnosis of the particular user with ASD)

Facial Orientation

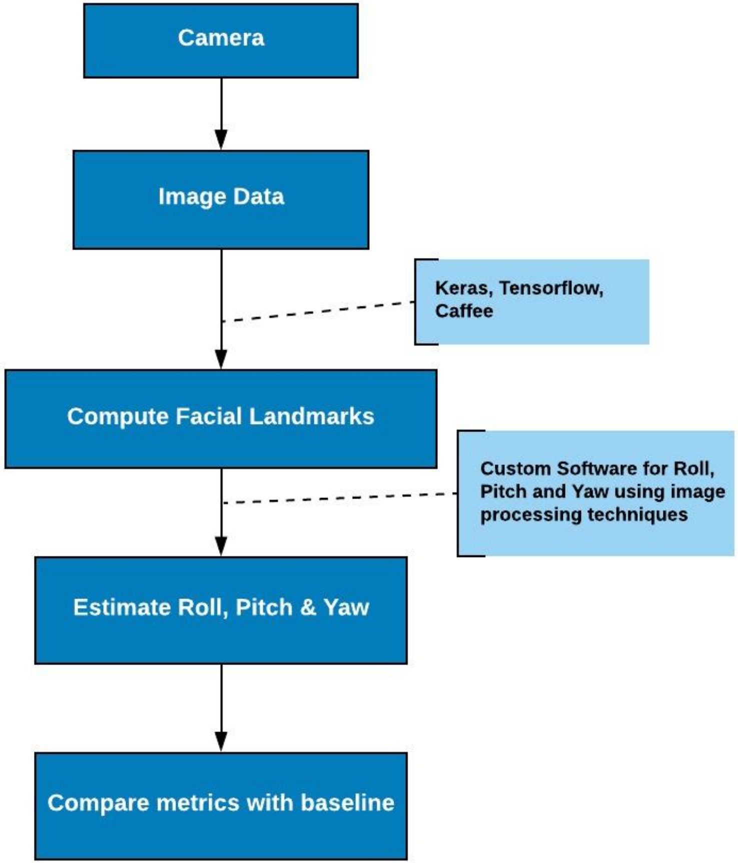
- ❖ Facial Landmarks calculated from Video Feed using DLib, OpenCV
- ❖ Compute pairwise euclidean distances between facial landmarks to generate a feature vector for each pose.



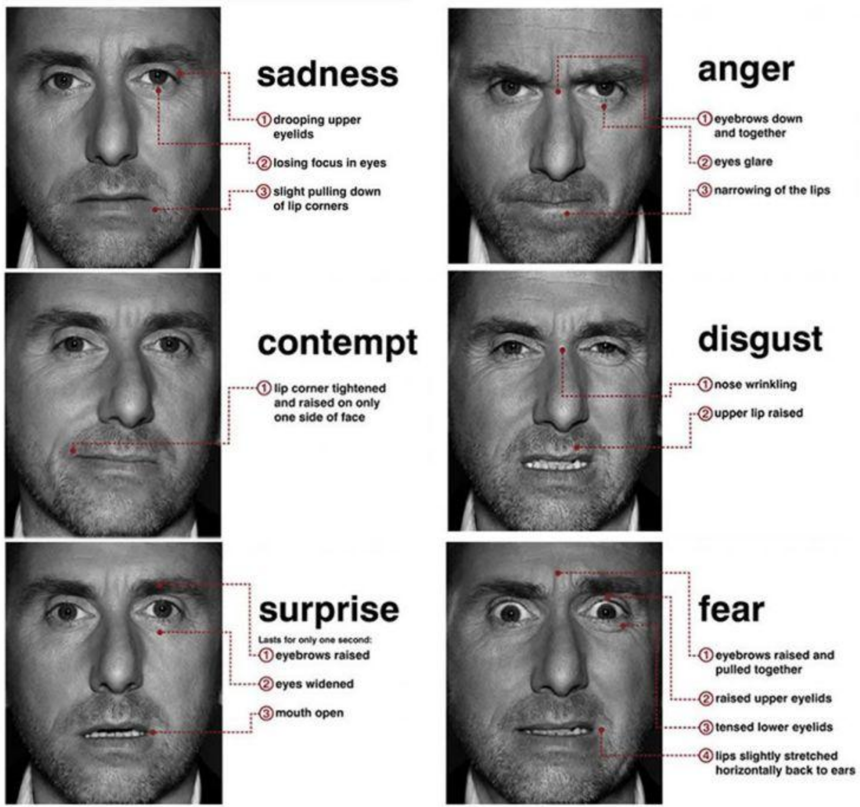
Facial Landmarks

Facial Orientation -Design

- ❖ The Roll, Pitch and Yaw is labeled for each pose using Amazons Face-Detection API.
- ❖ The Data can be compared to a baseline metric to show increase or decrease in head movements over multiple sessions, and can give a good indication of the treatments efficacy.

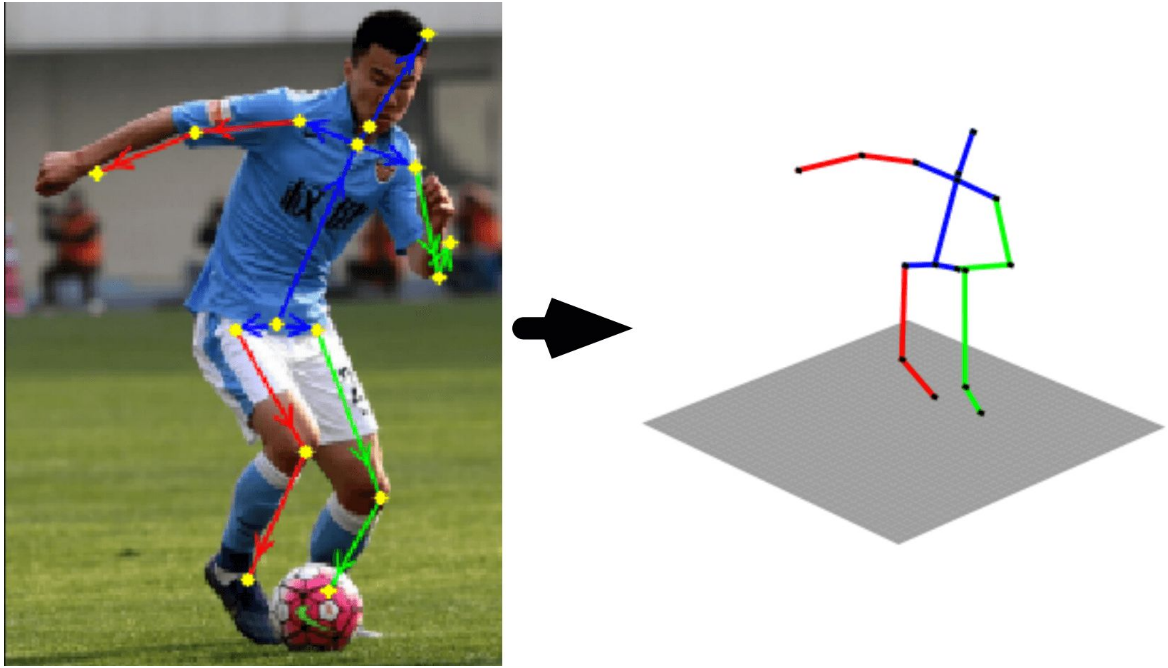


Emotion Recognition



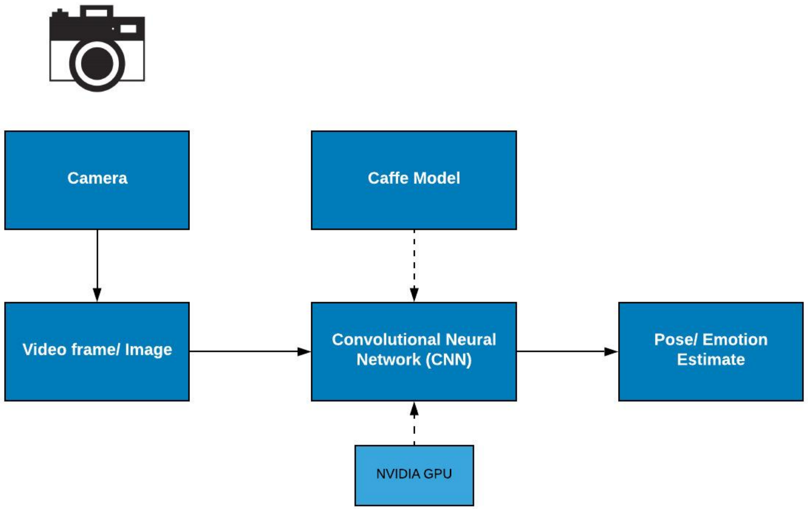
- ❖ Goal: Emotion Recognition via Facial, Gesture Analysis. The objective is to observe a user's face and body language and try to derive possible emotional information from the subject from a computational perspective.
- ❖ Facial Landmarks can be passed through a Deep Convolutional Neural Network to predict emotions.

Emotion Recognition



- Estimate Keypoints (detect and localize major parts of the body to generate the human pose and compare gesture with known labeled gesture-emotion data.
- These Points of interest together provide a robust emotion recognition subsystem.

Keypoint/Landmark Analysis- Emotion Recognition



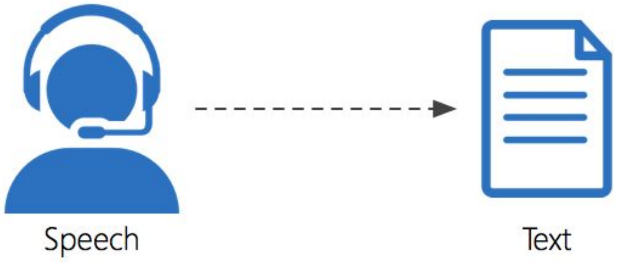
- ❖ Image Data is taken from the Video Feed and fed to a pre-trained Caffe model (COCO Keypoints Challenge 2016, COCO is a large-scale object detection, segmentation, and captioning dataset) to generate pose estimates. For the face data, use a deep convolutional neural network with keras and tensorflow backend to generate an emotion estimate.
- ❖ CUDA Dependency is required for GPU Computation
- ❖ To speed up the caffe model, it is accelerated by drop-in integration of NVIDIA cuDNN GPU Accelerated Library
- ❖ Emotion estimate derived from pose estimate using a known gesture-emotion dataset.

Task Planning/ Task Fulfillment

- The Task Modules are designed for the robot to specifically engage with the user to understand emotions and expressions and to learn and practice social behavior and responses.
- Task Modules are graded, and more difficult social interactions are introduced when the user achieves a desired level of social adaptability.
- Tasks are programmed to keep the user experience fresh ,data is tracked and logged and progress can be checked in monthly intervals.
- Introductory Module: Greetings, Introductions etc
- Physical Response Module: Waving your Arms, Handshake, Hi-Five, Following the robot etc
- Stress-relief Module: Deep Breaths, Count to 10, Stress Ball (usually when angry)
- Conversational Modules: Conversation Dynamics, Taking Turns to Speak, Difficult Conversations
- Emotional Understanding: Identifying and Imitating the major emotions, demonstrating and target emotions when required
- Situational Modules: (Birthdays, Meet and Greets)
- Tasks are also influenced by user emotions, for example, the stress-relief module would be triggered if the user is sad/angry/irritable, otherwise task randomization is used.

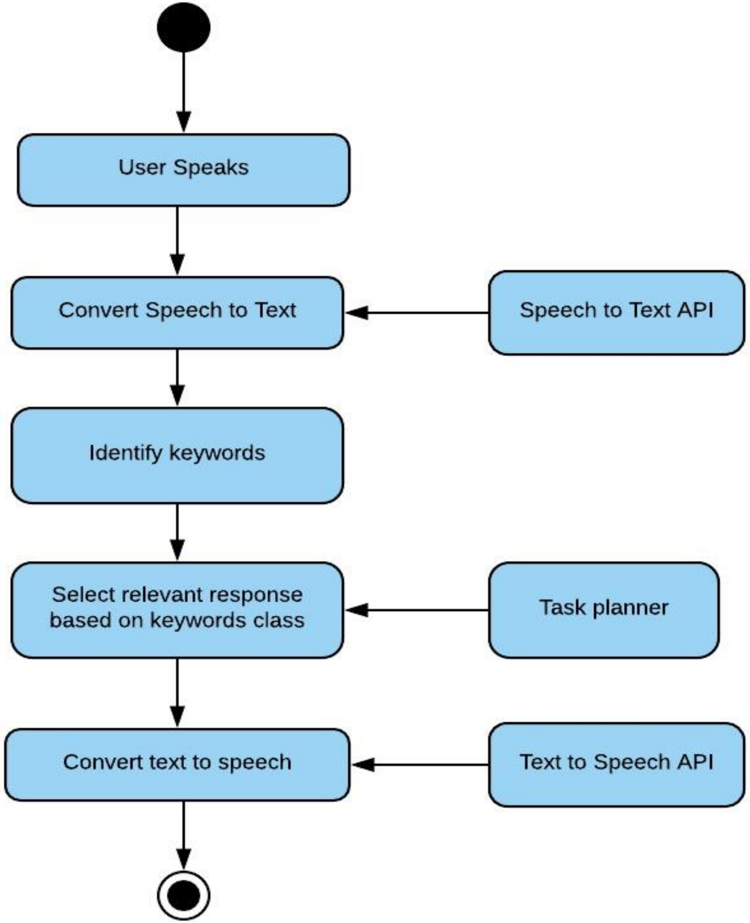
Natural Verbal Interaction/Voice Analysis

- Recurrent Neural Networks (RNNs) are the backbone of speech recognition systems.
- Speech recognition is difficult to program because different people can have different dialects and pronunciation speeds and therefore the same word can be heard in many different ways.
- Baidu Deep Speech 2 is supposed to be the best speech to text conversion tool available in the market. But unfortunately it is not available for public release.
- Our best alternative is Microsoft's Azure API. In particular, the Speech SDK provided by Microsoft's Speech Services can be easily integrated into any system running Linux, Windows, Mac, Android, iOS, etc.
- Subscription works on a hourly basis and rates are starting at $1 per hour for speech to text translation.

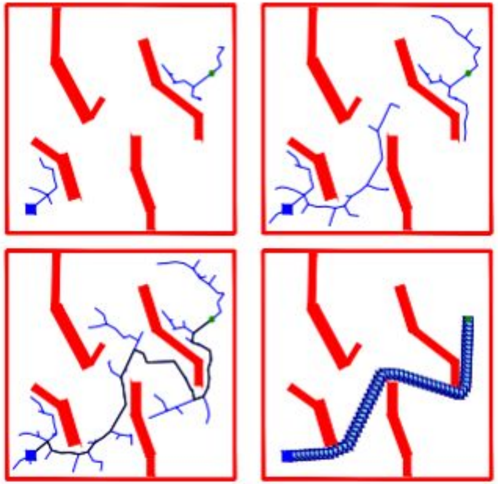


Voice Interaction

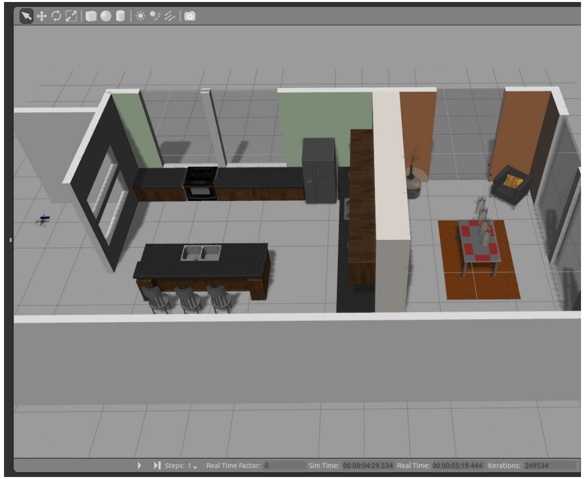
- Voice Interaction is necessary component for modeling good social behavior.
- As the user speaks, the speech is converted to test using an RNN, and the keywords are identified.
- A relevant response is selected from the task planner which takes in voice and emotion input data to generate appropriate response selections.
- For example, a stressed voice and extreme emotions (sad/angry/panic) might generate a calm-down module as a response, which might include taking deep breaths, counting to ten, etc.
- After the response is generated, a Text to Speech API (from Azure) is used to generate speech responses



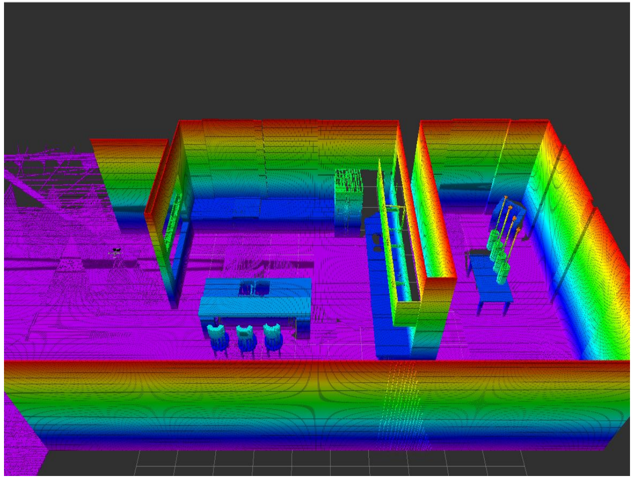
Path Planning



Bi-directional RRT



Local Environment



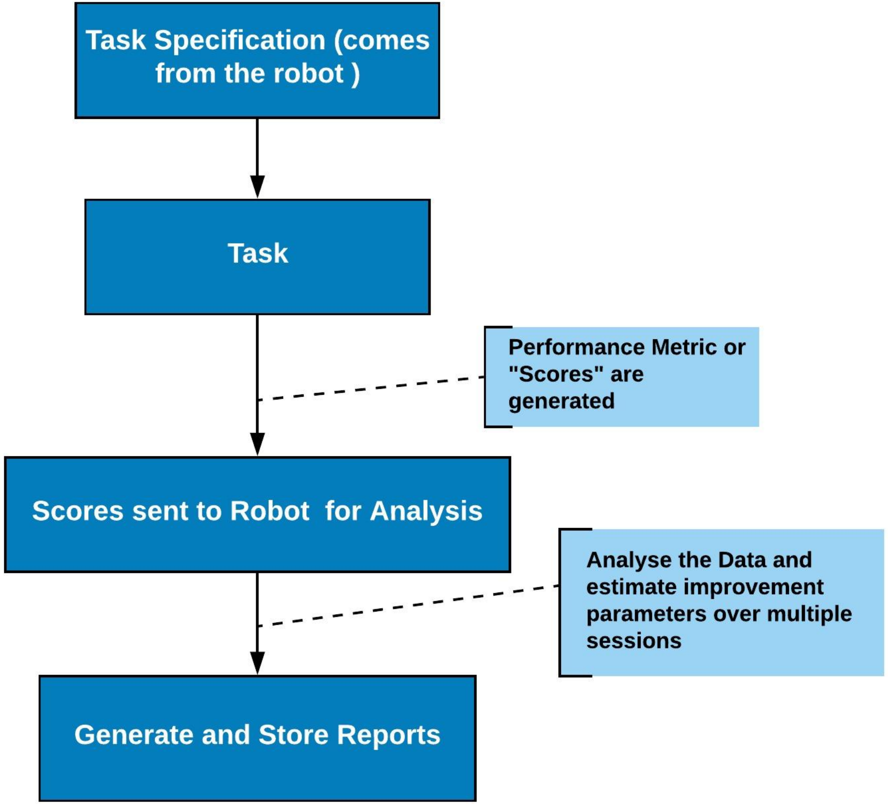
3D Occupancy Model

- The Kinect Sensor on top of a Pan-tilt unit can be used to map out the room and generate a model of the room
- OcTree and Octomap can be generated based on Kinect sensor data and the low probability regions (obstacle free regions) can be used to set navigation waypoints
- OMPL Planner can be coupled with ROS and C++ to create sampling based solutions like RRT, RRT* and RRTConnect based on Octomap probability values
- Collision avoidance is handled by move-base local planner available in moveit motion planning library under ROS.

Tablet UI

The tablet is used to get the performance metrics for various tasks such as:

- ❖ Reading
- ❖ Writing
- ❖ Puzzles
- ❖ Arithmetic
- ❖ Spatial Reasoning Games: Tetris etc.
- ❖ Motor Control Games: Pressing Buttons of different colors upon request, reconstructing 3D puzzles etc.
- ❖ Watching Instructional Videos



Security

- Hardening: Hardening is usually the process of securing a system by reducing its surface of vulnerability, which is larger when a system performs more functions; in principle a single-function system is more secure than a multipurpose one.
- Reducing available ways of attack typically includes changing default passwords, the removal of unnecessary software and services etc
- There are various methods of hardening Unix and Linux systems. This may involve:
  - Applying a patch to the kernel such as Exec Shield or PaX
  - Closing open network ports
  - Setting up intrusion-detection systems, firewalls and intrusion-prevention systems.
- There are also hardening scripts and tools like Lynis, Bastille Linux, for example, which deactivate unneeded features in configuration files or perform various other protective measures.

System Integration w/ Cyber-resiliency

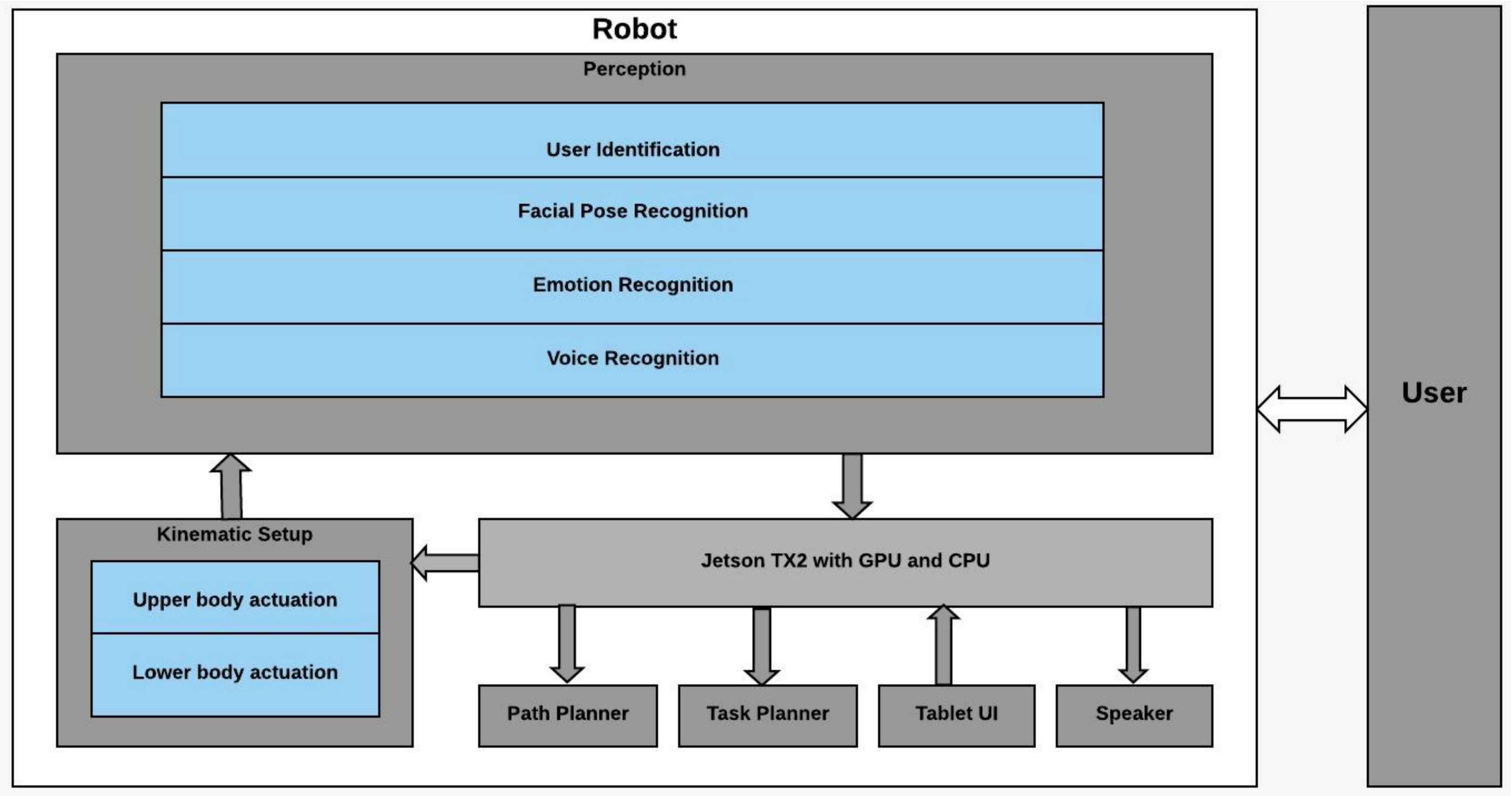
The robot has a centralized processor (Jetson TX2) which performs all the computations taking in the perception/task inputs gathered from various peripherals and outputs actions. The resulting actions by the user are analyzed in real time and the data gathered regarding the task performance metrics or “scores” are logged into the storage device (in the robot) and can be securely transmitted to a nearby computer for further data analysis if needed.

The Jetson TX2 is running a Linux Operating System (Ubuntu etc.) which has ROS Installed. The communication between the peripherals and the processor (Jetson TX2) is via ROS nodes. This makes the system highly modular and stable and the failure of one node does not influence the performance of the other nodes.

ROS provides you with the architecture to achieve concurrency, inter-communication, and extensibility for robotics processes.

All the modules are containerized in Docker and they run as individual ROS nodes. Containers are isolated from each other and bundle their own application, tools, libraries and configuration files; they can communicate with each other through well-defined channels.

System Integration



Mechanical Hardware w/ Cost

Hardware Type	Quantity	Total Cost for each Hardware
Microsoft Kinect	1	$30.00
Jetson TX2 (GPU: NVIDIA Pascal, 256 CUDA Cores ; CPU)	1	$569.99
2pc 1500mAh	4	$107.96
Tablet - Asus ZenPad 3S (9.7”)	1	$299.95
Stylus - Asus Z	1	$35.95
ABS Skeleton Material 1.75mm (1kg)	4kg	$63.16
Speakers - 20W 4 Ohm	2	$19.98
LED Display 7”	1	$10.00
Actuators for Arms - DC Motor 12V, 5600 RPM	10	$75.00
Wheel Motors - High Torque 6000 RPM 80W DC Motor	2	$116.00
Tobii Eye Tracker	1	$170.00
Samsung 850 PRO - 2TB - 2.5-Inch SATA III Internal SSD	1	$260

Software Cost

Software	**Cost ($)**
Unity Plus	$35.00/month
Microsoft Azure (speech to text)	$1.00/per hr

Total Prototype Cost

Approx. $6,500.00 (Tentative with 3rd Party API’s and not taking into account software development cost and additional labor)

Additional Work

- Material Analysis - Robustness in skeletal structure, Improved exteriors for interaction with users.
- Cost Effective Functionality - Optimizing Hardware/Software with future products
- Open-Source Software: Open-source Libraries for Voice Recognition will lead to a more cost-effective robot as the software cost for Azure is on a hourly basis, blowing up the cost price.
- Cultural Assimilation: To deploy it in Countries with various cultural differences - Hand-Gestures, Greetings, Language, Form of Communication (Implicit or Explicit) are vastly different.
- Ensure Reliability, Performance, Trialability, Security and Privacy Concerns are addressed appropriately.
- Introduce Possible Variants in the Product Offerings to match user requirements/budget

Product Offering

Features:

- ❖ Voice/Face Recognition
- ❖ Emotion Detection
- ❖ Optimized for Human Interaction
- ❖ Danger-Free Use
- ❖ User- Friendly
- ❖ Assistive Social Care

Social Therapy Robot

Price (Tentative) - 15,999 USD

References

- https://www.iidc.indiana.edu/pages/the-use-of-technology-in-treatment-of-autism-spectrum-disorders
- https://www.nhs.uk/conditions/autism/symptoms/
- Unity SDK for Eye Tracking: https://developer.tobii.com/consumer-eye-trackers/unity-sdk/
- ROS# - https://github.com/siemens/ros-sharp
- Emotion Estimates via Human Pose - https://www.vision.ee.ethz.ch/publications/papers/articles/eth_biwi_00545.pdf
- Human Pose Estimation using OpenCV - https://www.learnopencv.com/deep-learning-based-human-pose-estimation-using-opencv-cpp-python/
- Unity on linux - https://blogs.unity3d.com/2015/08/26/unity-comes-to-linux-experimental-build-now-available/
- Facial Pose Estimate - https://medium.com/analytics-vidhya/face-pose-estimation-with-deep-learning-eebd0e62dbaf
- Caffe - http://caffe.berkeleyvision.org/installation.html
- DLib- http://dlib.net/
- Amazons Face Detection API - https://docs.aws.amazon.com/rekognition/latest/dg/faces-detect-images.html
- Head Movements in children with ASD - https://molecularautism.biomedcentral.com/articles/10.1186/s13229-018-0198-4
- Robots4Autism Curriculum - https://robots4autism.com/wp-content/uploads/R4A_ScopeSequence_013018.pdf
- Microsoft Speech Services - https://azure.microsoft.com/en-us/pricing/details/cognitive-services/speech-services/
- Microsoft Speech SDK - https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/speech-sdk
- Therapy Robot Bandit - https://www.popsci.com/science/article/2010-05/humanoid-robots-are-new-therapists#page-3