Emotional Speech Recognition toward Modulating the Behavior of a Social Robot

Chris Lytridis, Eleni Vrochidou, Vassilis Kaburlasos

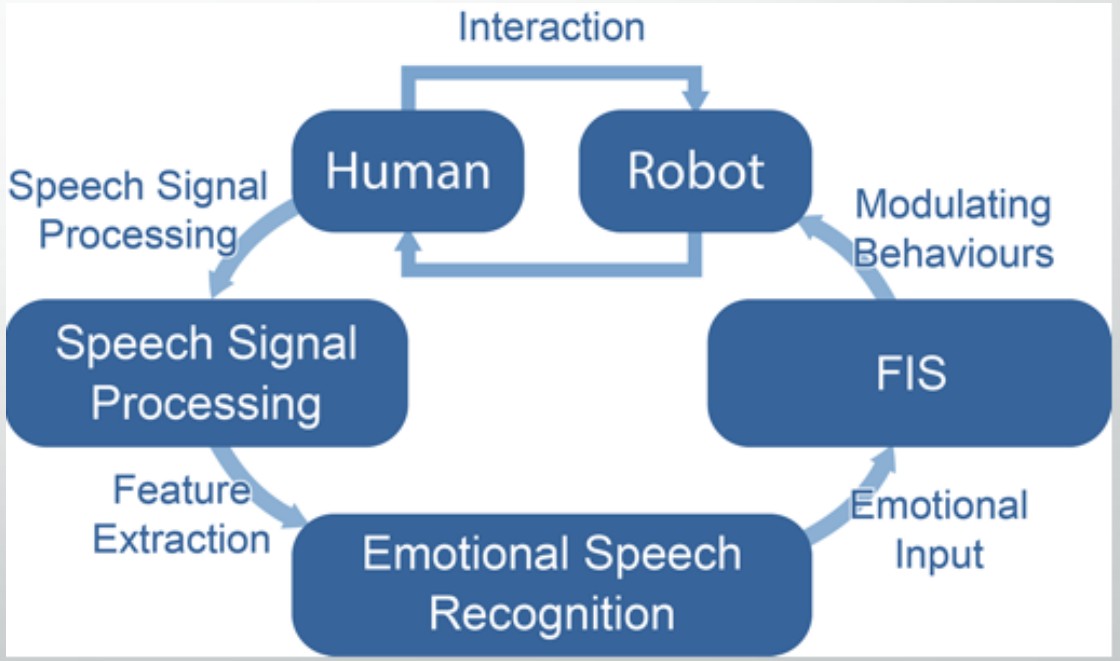
HUman-MAchine INteraction (HUMAIN) Lab

Social robots

- The incorporation of social robots in education, special treatment and therapy, requires that robots can interact socially with individuals.
- The effectiveness of human-robot interaction depends on the ability of the social robot to adapt its behaviour according to human emotional states.
- This work proposes a novel behaviour modulation system for social robots based on emotional speech recognition.

The Behaviour Modulation System (BeMoSys)

- The proposed methodology involves three stages:
  1. Feature extraction from the voice signal.
  2. Signal classification to various emotional states.
  3. Behaviour modulation.



Emotion Recognition

- The OpenVokaturi library was used for emotion recognition from the human voice.
- Validated with existing emotion databases.
- Predicts five emotions; happiness, sadness, anger, fear, and neutrality.
- Returns probability of each emotion being present in a given signal.
- The highest value among all predictions, determines the dominant emotion.
- Result: 66.5% accuracy.

Robot actions

- Three categories of robot actions:
  - Facial expressions (eye LEDs).
  - Sounds (audio tracks and music).
  - Animation (gestures and postures, using the robot's head, arms and legs).

Action category	Individual Behaviours
Facial expressions	• Eye LEDs (colours, blinking frequency)
Sounds	• Happy sounds • Relaxing sounds • Music
Gestures/Postures	• Head movements • Arm movements • Leg movements

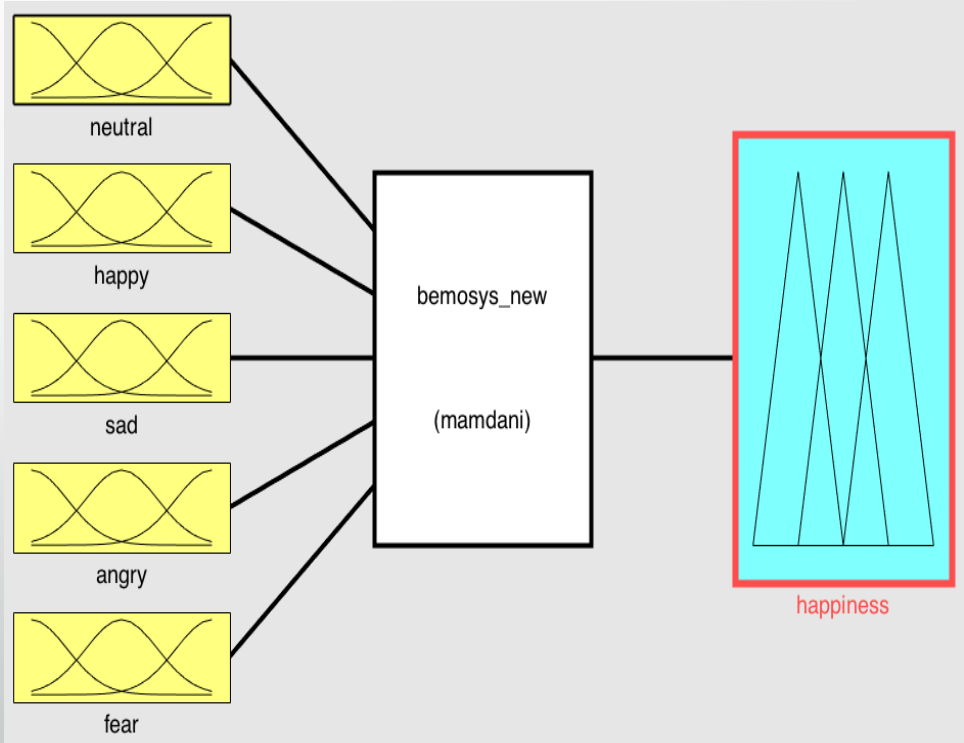
Robot behaviours

- Robot actions are combined and form the five pre-set distinct robot behaviours, A to E.
- Behaviour A is the robot's response to lowest levels of estimated happiness.
- Behaviour E is the robot's response to the highest levels of estimated happiness.

Behaviour	Description
A	Eye LEDs (colour set to white); Relaxing sound; Relaxing animation
B	Eye LEDs (blinking); Relaxing sound; Animation to attract attention
C	Eye LEDs (rapidly blinking); Cheerful sound; Zestful animation
D	Eye LEDs (rapidly alternating colours); Enthusiastic sound; Cheerful animation
E	Eye LEDs (rapidly alternating colours); Upbeat music; Enthusiastic animation

The Fuzzy Inference System (FIS)

- There are five inputs to the FIS which correspond to the five emotions extracted from the speech signal.
- The knowledge base of the inference system consists of 47 rules, which map the inputs to a single output value referring to the level of estimated happiness.
- Based on the value of estimated happiness, one of the five pre-defined robotic behaviours is selected.



Experiment (1)

- 399 speech recordings retrieved from the Berlin Database of Emotional Speech were used:
  - 74 recordings of neutrality.
  - 71 recordings of happiness.
  - 126 recordings of anger.
  - 67 recordings of fear.
  - 61 recordings of sadness.

Experiment (2)

- For each recording, the resulting emotional context of the speech signal was calculated and supplied to the FIS.
- The FIS calculated the estimated level of happiness.
- A robot behaviour was selected according to the estimated level of happiness.

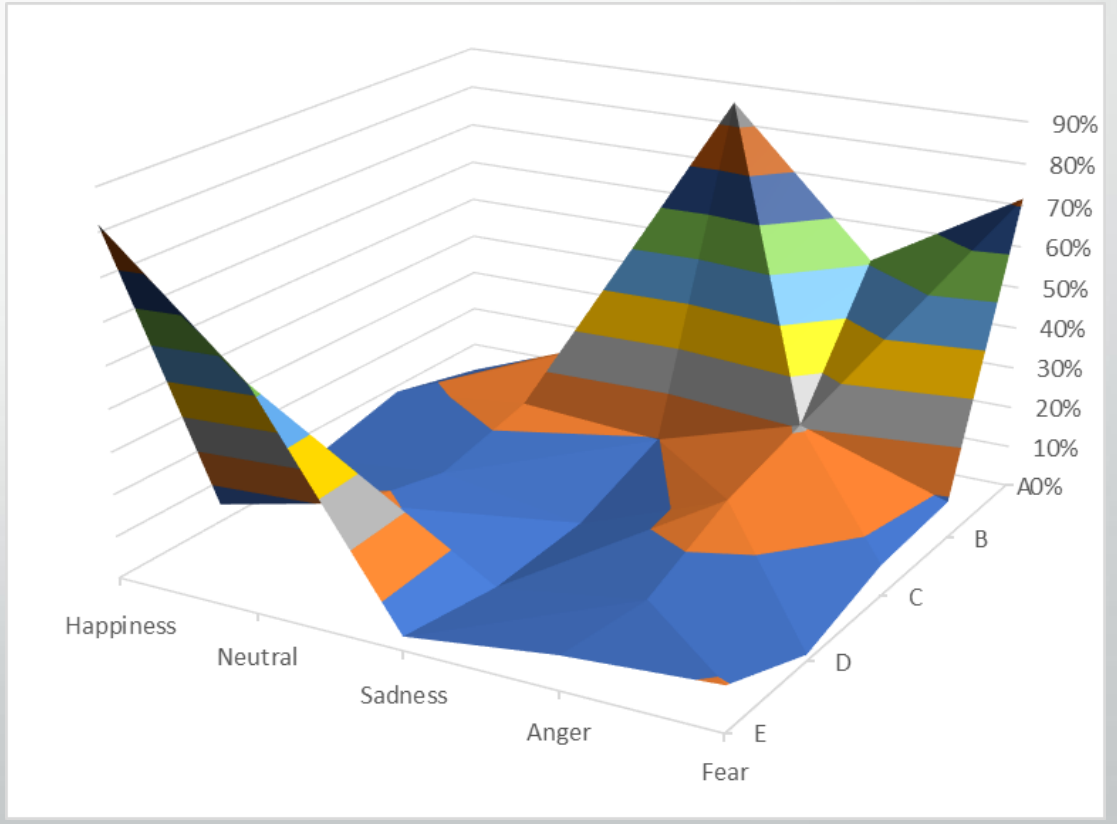
Results

- The BeMoSys was implemented on a NAO robot.
- Selection of an appropriate behaviour given an emotional speech signal is successful most of the times.
- Discrepancies are observed due to the limited accuracy of the emotion recognition library.

Behaviour / Emotions	A	B	C	D	E
Happiness	3%	8%	3%	4%	82%
Sadness	85%	10%	2%	0%	3%
Anger	51%	21%	15%	6%	7%
Fear	72%	9%	7%	2%	10%
Neutrality	16%	12%	7%	12%	53%

Selected behaviours based on calculated happiness level

- BeMoSys appears to be very consistent.
- When the detected emotion is anger, sadness or fear, then the system is driven to lower values of happiness level, and triggers the appropriate robot behaviours.
- Clearly happy emotional input, leads to the activation of reinforcing robot behaviours.



Conclusions

- A behaviour modulation system was implemented on the social robot NAO.
- The robot records the human voice and extracts the emotional content with the help of a well-known software application.
- Extracted emotions act as inputs to a FIS in order to map emotions to robotic actions.
- Future work will focus on applying the proposed system in the areas of special treatment and education.

Thank you

This work has been supported, in part, by the European Commission Horizon 2020 MSCA-RISE Project no. 777720 "Cyber-Physical Systems for PEdagogical Rehabilitation in Special EDucation (CybSPEED)".