TITLE
SUBTITLE

CS 7075: Artificial Intelligence and Robotics
Dr. Majeed
Matthew Hensel, Ian Taylor

INTRODUCTION

Abstract—As artificial intelligence becomes more ubiquitous, interfacing with machines begins to resemble a sort of social psychology. This paper investigates the underlooked aspect of nonverbal communication in human-robot interaction, addressing the limitations of current interfaces. We propose a comprehensive approach that considers human-to-robot nonverbal communication through the development of expressive movement primitives, social cues, and affective vocal characteristics. Our methodology combines insights from psychology and human-computer interaction with artificial intelligence techniques including Large Language Models (LLMs) and Vision Language Models (VLMs). Through iterative design, prototyping, and user studies, we aim to develop design principles for creating more socially intelligent robots capable of conveying and interpreting intention, attention, emotion, and attitude. The outcomes of this research will contribute to more intuitive and "effective affective" human-robot collaboration. We expect to see heightened understanding and more seamless communication because of the added nonverbal awareness we equip the LLM with.

PROBLEM STATEMENT

The core problem of our research paper is: how can we design robots to better perceive human emotion? Our key points include perception and planning as we must design ways for our AI to comprehend gestures and emotion conveyed through one's facial structure which will then be interpreted and used in communication.

The typical paradigm of interaction between humans and artificial intelligence consists of natural language conversations [1]; these lack key body language and nonverbal communication clues, both of which make up a significant portion of trust between interlocutors [2]. By developing a set of "Gesture primitives"

PROPOSED SOLUTION

To solve this problem, we will consider the following approaches:

1. We will explore methods for robots to perceive and interpret human nonverbal cues. This may involve utilizing machine vision to analyze human body language, facial

expressions, and gaze. For vocal cues, we will investigate audio processing techniques to understand changes in pitch, tone, and speech patterns that convey emotion or intent. We will aim to develop mappings between these perceived human cues and the robot's internal understanding and subsequent actions. Recognizing affective communicative intent in robot-directed speech, such as praise, prohibition, attention, and comfort, is important.

2. We will explore the use of Large Language Models (LLMs) and Vision Language Models (VLMs) to generate contextually appropriate robot gestures and nonverbal behaviors based on user language input and environmental context. We will also investigate machine learning models to learn mappings between human nonverbal cues and robot responses, and to personalize robot nonverbal behavior based on user characteristics and interaction history. Frameworks like EMOTION utilize LLMs for generating expressive motion sequences.
3.

METHODOLOGY

A. Vision-Based Facial Emotion Recognition
    1. Facial feature extraction
    2. Emotion classification
    3. Contextual understanding

B. Speech Analysis System
    1. Acoustic feature extraction
    2. Paralinguistic analysis

EXPECTED OUTCOMES

SIGNIFICANCE

REFERENCES

[1]
[2]
A. Mehrabian, “Inference of attitudes from the posture, orientation, and distance of a communicator.,” *Journal of Consulting and Clinical Psychology*, vol. 32, no. 3, pp. 296–308, 1968, doi: 10.1037/h0025906.