Nonverbal Behavior Generation for Social Bite Timing

Abrar Anwar[1], Tapomayukh Bhattacharjee[2]
[1]University of Southern California, [2]Cornell University

Motivation

- Millions of people in the world need help with feeding
- Robots have the potential to help these people with feeding
- Stakeholders find eating with friends/family the most memorable

- The process of feeding consists of three steps:
  - Bite acquisition consists of learning how to pick up food
  - Bite timing focuses on when to feed a participant
  - Bite transfer deals with the physical transfer of food into someone's mouth
- Previous work on these steps focused on dyadic scenarios
  - Bite timing in social dining is challenging because it is a cadence of various rich multimodal implicit and explicit cues [1]

- Gestures and nonverbal communication are powerful cues for implicit communication.
  - Nonverbal behaviors reveal the focus of someone's attention

Methodology

- Since human trials can become costly to run, we generate simulated data from speech in YouTube videos of people eating
  - Using Assistive Gym simulator, modified for social dining

- In the simulation, there are three friends eating together, one of whom needs to be fed with an assistive feeding robot

- We collect and label internet videos of groups of people eating together for participant's roles and bite timings
  - For each video, we assign roles to the speaker and the listeners:
    - **Active listeners** pay full attention to an ongoing conversation
    - **Passive listeners** do not pay full attention, and tend to not engage in conversation or focus on other things (like eating)
    - **Direct conversation** is an action initiated by a speaker towards a specific listener as direct comment or inquisition
    - **Broadcasting** is an action initiated by a speaker who speaks without targeted listeners
  - These labels are used to determine where the agent is placing their attention.

Infrastructure

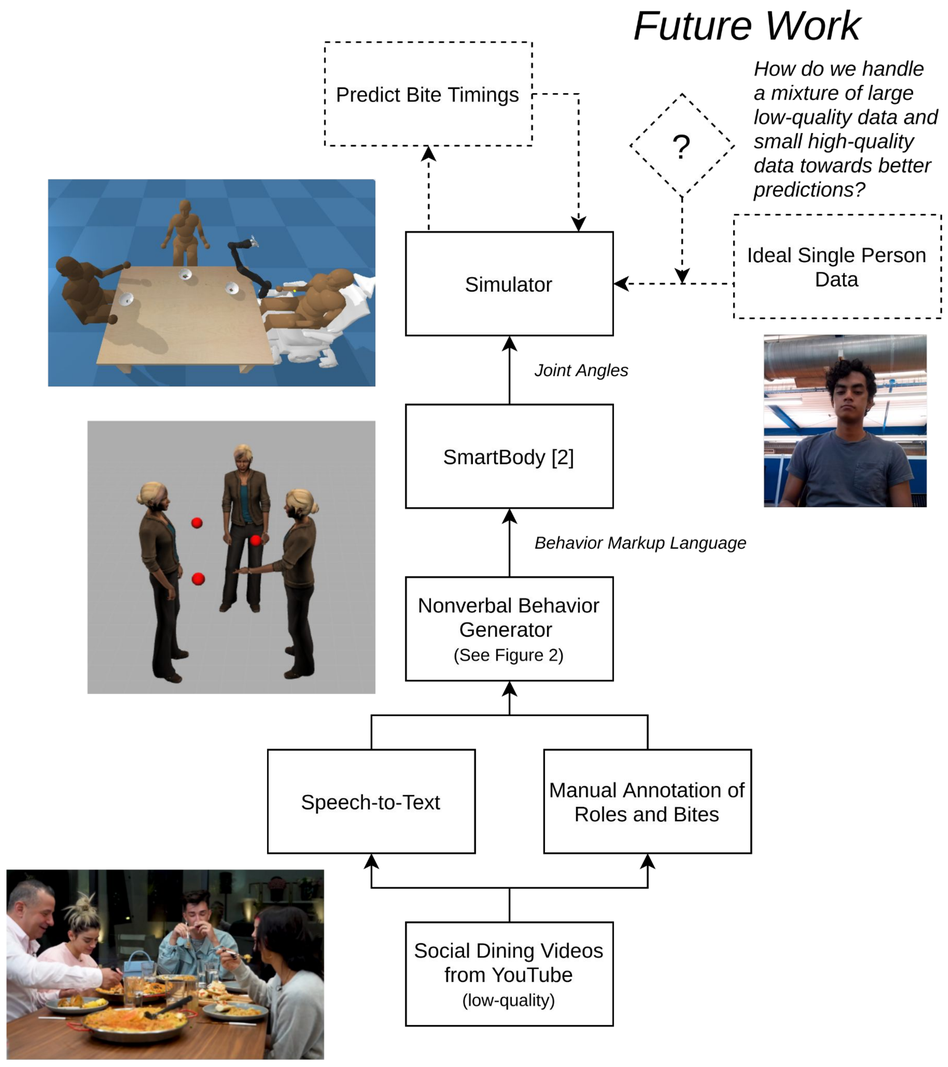

Figure 1. Infrastructure for a simulator based on videos

- We use social dining videos:
  - Video is used to annotate the bite timings and participant's roles
  - Nonverbal behavior is generated according to Figure 2
  - The joint angles and positions are mapped to the Assistive Gym environment

- Video of realized behaviors in SmartBody: [Link]
- Video of behaviors mapped to simulator: [Link]

Nonverbal Behavior Generation

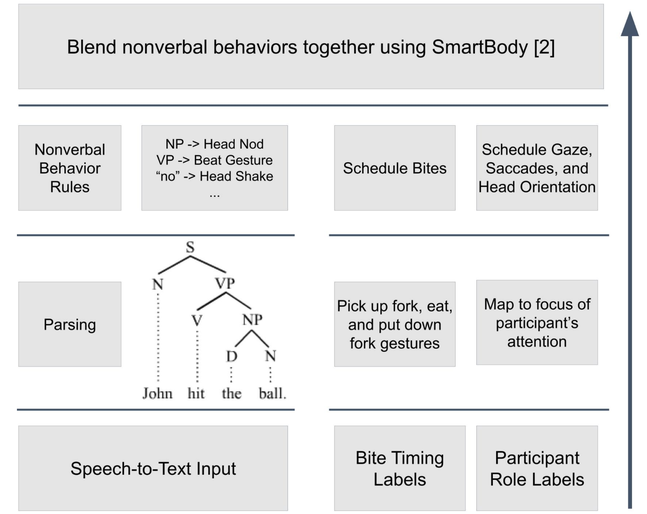

Figure 2. There are three main components to generating nonverbal behavior and actions for bite timing (left to right):

1. Speech-to-text is parsed, and rules are applied based on NonVerbal Behavior Generator (NVBG) to generate gestures
2. The bite timing labels are used to generate action behaviors
3. Participant roles decide what they are focusing on

These behaviors are scheduled to occur at the right time, blended together and realized using SmartBody [2].

Future Work

- We plan to build a prediction algorithm to predict bite timing
- We hope to answer which set of social cues matter based on who is being looked at

Acknowledgement

I would like to thank Google Research exploreCSR and the University of Texas at Rio Grande Valley for organizing and hosting these research experiences for students. Special thanks to Frank Bu for his work on the Assistive Gym social dining environment and his past work on this project.

References

[1] Laura V Herlant. 2016. *Algorithms, Implementation, and Studies on Eating with a Shared Control Robot Arm.* Ph.D. Dissertation.
[2] Marcus Thiebaux, Andrew Marshall, Stacy C. Marsella, Marcelo Kallmann. SmartBody: Behavior Realization for Embodies Conversational Agents. 2008. In International Conference on Autonomous Agents and Multiagent Systems (AAMAS)

Contact

Abrar Anwar abrar.anwar@usc.edu