

Motivation

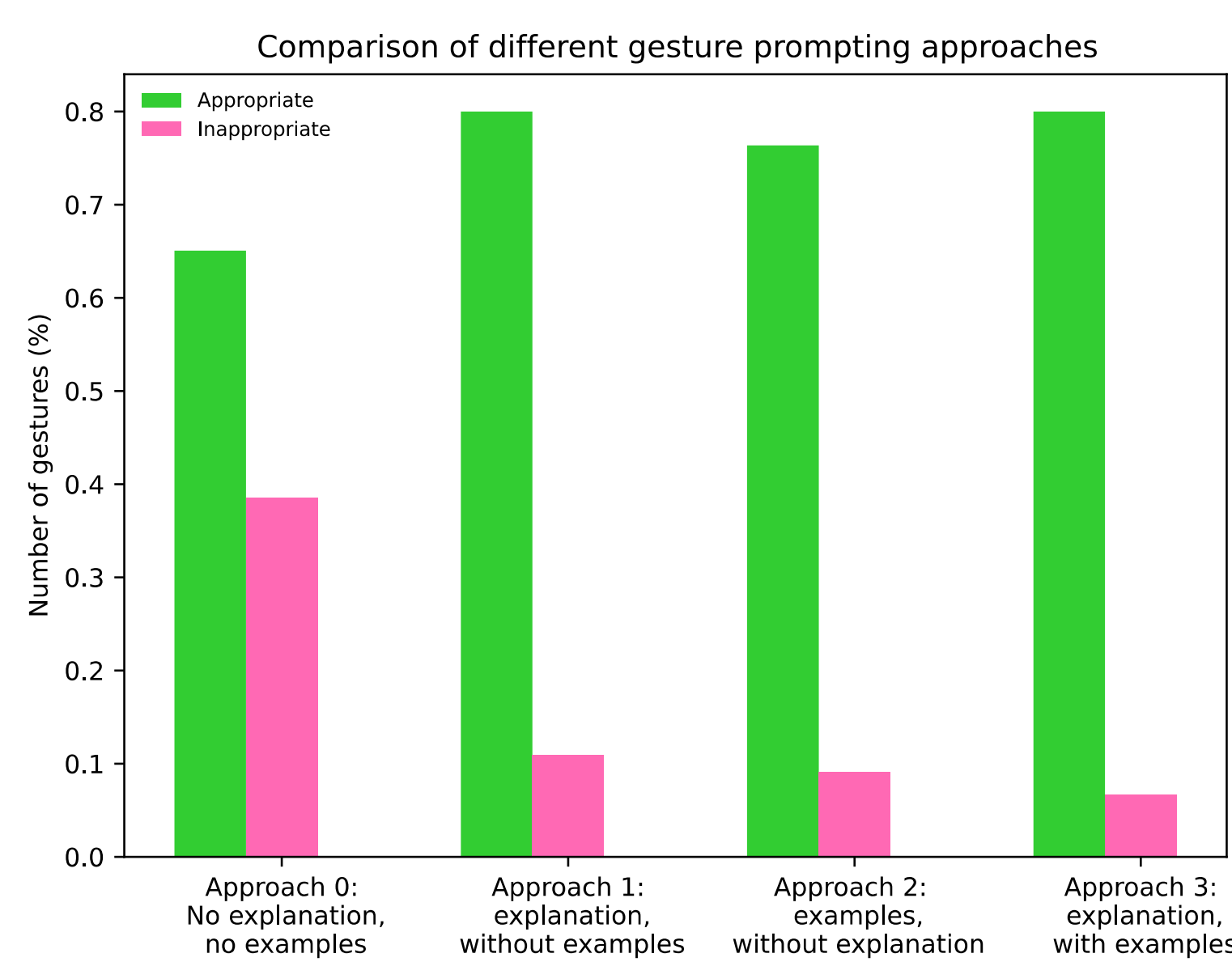
- Face-to-face interaction is multimodal.
- Gestures
 - communicate information: **deictic** gestures establish reference, **beats** emphasize words, and **metaphoric** and **iconic** gestures illustrate concepts.
 - convey social information, such as personality, and emotions.
 - Impact affective states of observers, including liking and trust
- Generation of gestures is a key agenda in virtual agent research.

SIMA: LLM-based Gesture Selection System

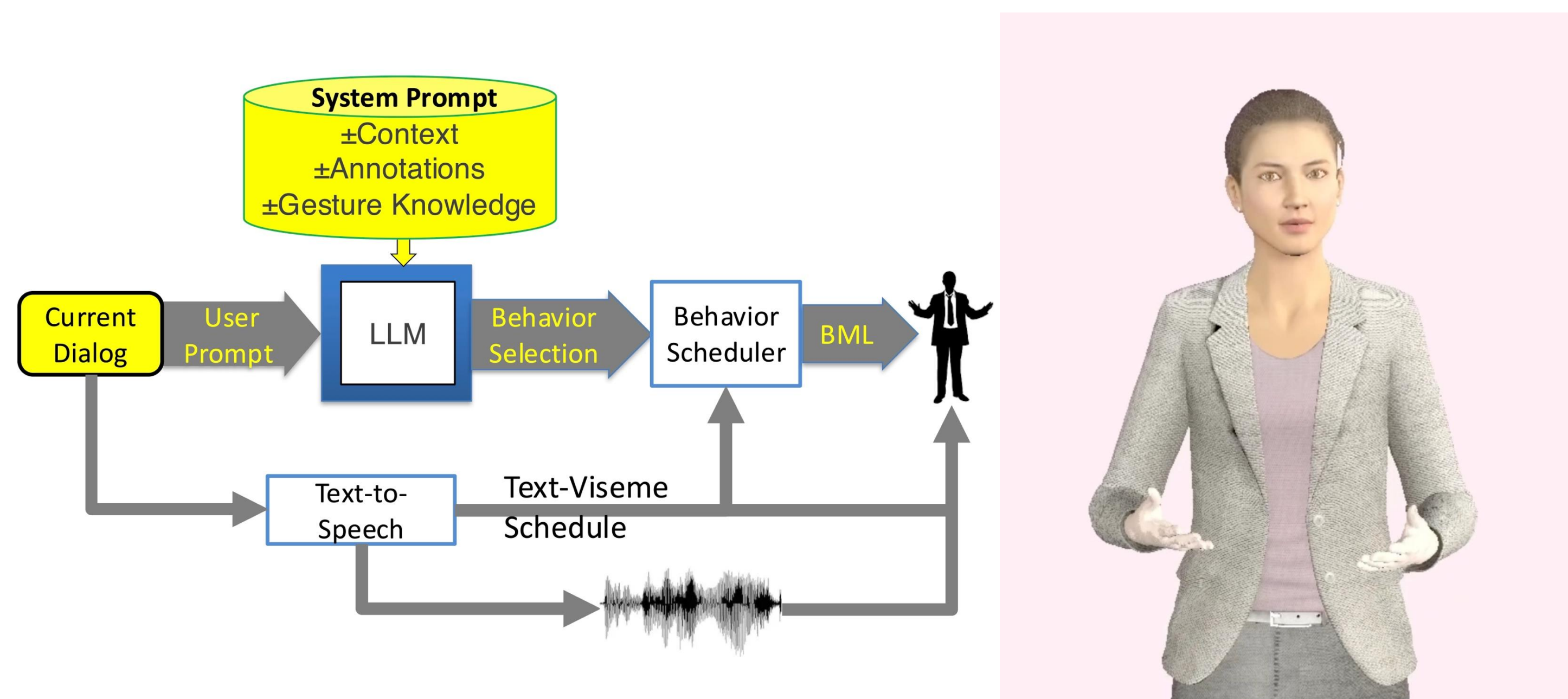
- Example Metaphoric Gesture:
 - Collect Gesture conveys bringing together into one entity



- Different prompting approaches were evaluated and their selected gestures were assessed as meaningful and appropriate to the role and context.



- End-to-end Implementation within Virtual Human platform



Current Work

Extend this LLM-based gesture selection system to select appropriate **Deictic** gestures.

- Example: Football referee pointing where foul occurred
- Example: SIMA virtual human pointing at addressee



Can LLMs infer deictic gestures?

Preliminary study of this question

We modeled the interaction between a virtual tutor and a student in an assembly task.

- The LLM was prompted with
 - description of a structural workspace including the workspace table, tools and materials, and location of each of them, and
 - the step-wise procedure of the task.
- The LLM is prompted to return , for each step of the task, the multimodal behaviors of the tutor including
 - the verbal instructions, and
 - deictic/or abstract deictic gesture pointing to a referent (object or region in the workspace) when appropriate

Preliminary Results:

LLM-selected deictics as appropriate & task-relevant.

Future Work:

- Extend these work to all types of nonverbal behaviors (facial expressions, posture, gaze, head movements)
- Conduct human-subject studies to study to user-perception of these LLM-selected behaviors.



- Calbris, Geneviève. "Elements of meaning in gesture." (2011): 1-398.
- McNeill, David. "Hand and mind: What gestures reveal about thought." (1992).
- Torshizi, Parisa Ghanad, et al. "Large Language Models for Virtual Human Gesture Selection." *arXiv preprint arXiv:2503.14408* (2025)
- Hensel, Laura B., Stephanie Cheng, and Stacy Marsella. "A richly annotated dataset of co-speech hand gestures across diverse speaker contexts." *Scientific Data* 12.1 (2025): 1748.