Large Language Models for Virtual Human Gesture Selection

Parisa Ghanad Torshizi, Laura Hensel, Ari Shapiro, Stacy Marsella

Presenter: Parisa Ghanad Torshizi

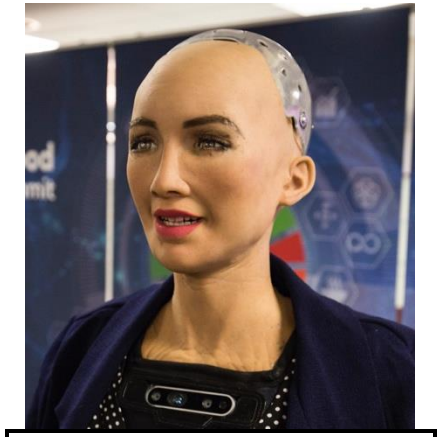
This research is based upon work supported by the NSF Grant #2128743.

Gestures

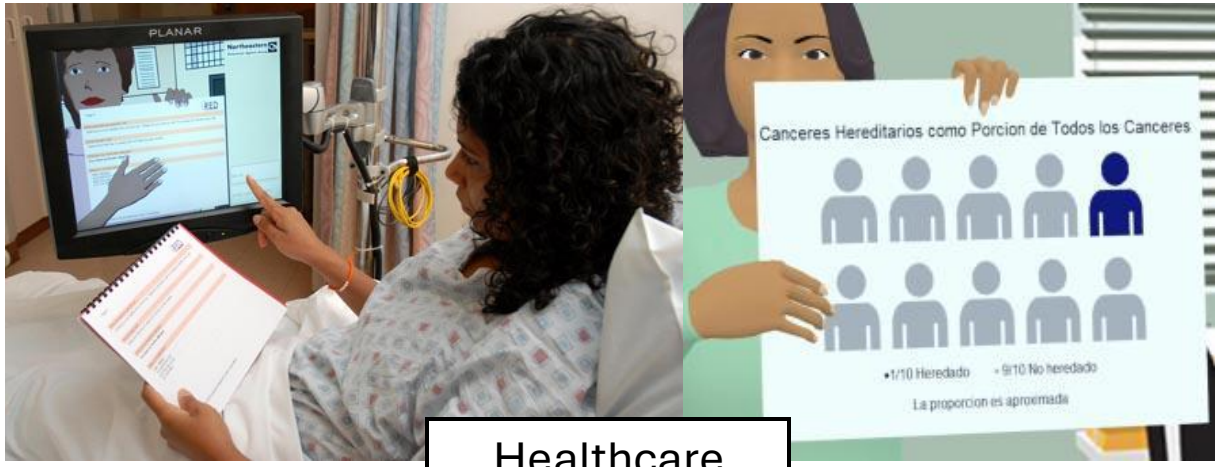
- Convey a wide variety of meanings
- Integral to face-to-face interactions
- Impact the persuasiveness of the speaker
- Impact trust, engagement, and recall in the addressee

Gestures and Virtual Humans/Robots: Context and Speaker Role

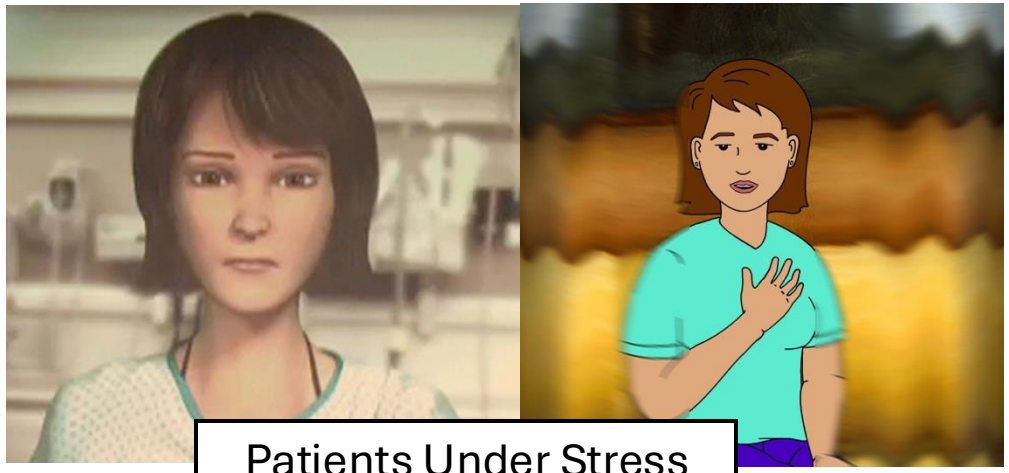
Mental Health Clinicians



Humanoid Robots



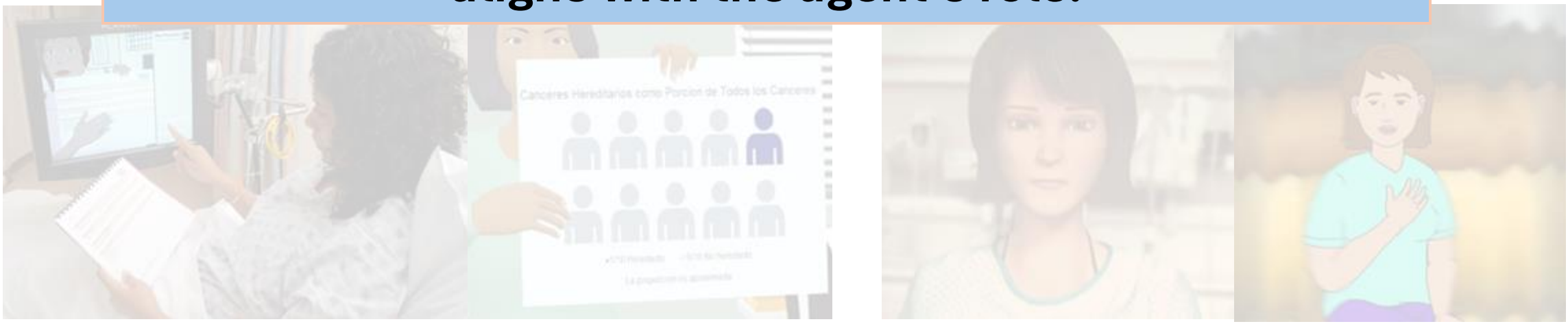
Healthcare



Patients Under Stress

Gestures and Virtual Humans/Robots: Context and Speaker Role

We propose an approach for generating virtual human gestures that both conveys intended meaning and aligns with the agent's role.



Gesture types (McNeill, 1992)

Gesture generation in virtual humans has explored different kinds of gestures:

Deictic (pointing) Gestures: Point to objects in the space.

Beat Gestures: Rhythmic gestures to emphasize a point in speech.

Iconic Gestures: Depict the physical properties of concrete objects

However, the least studied are:

Metaphoric Gestures: Grounded in image schemas that allow abstract concepts to be conveyed as if they had physical properties.

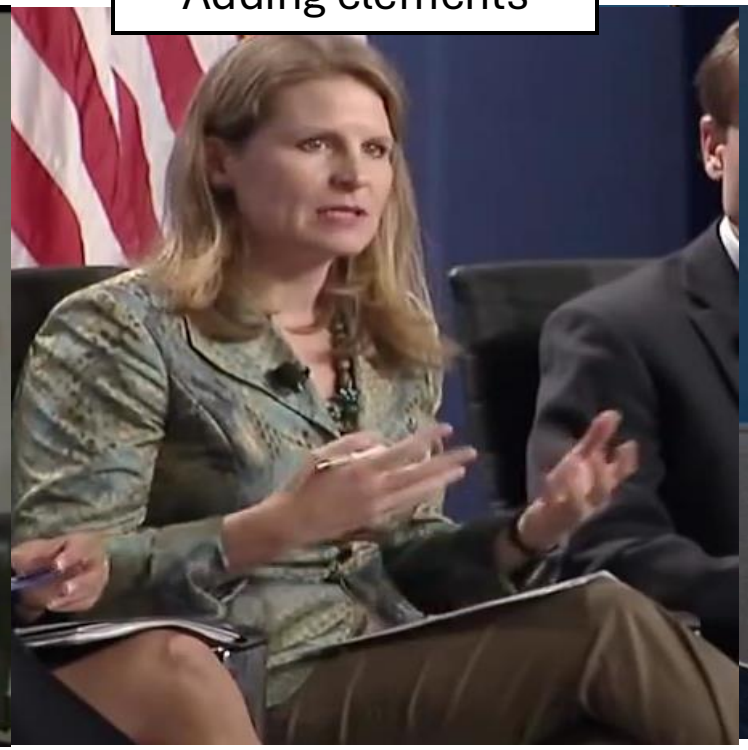
Image schema & Metaphoric gesture

The **Container Image Schema** represents an abstract set, with metaphoric gestures revealing cardinality, union, as well as removing and adding elements.

Cardinality

Removing elements

Adding elements



Union

What Gesture to use??

Data-driven and ML approaches that map utterances to gesture motion

- ❑ Tend to produce a statistical average over gestures
  - → Ignore individual and role differences
    - Politicians, Educators, Clinicians, and Comedians gesture differently
- ❑ Difficulty conveying meaning (REF)
- ❑ Lack of designer control

LLMs as a Solution

The paper demonstrates LLMs encode rich knowledge of human gestures

From abstract to specific

- Abstract/Theoretical knowledge, such as
- Image schema
- Types of gestures
- How different roles use different gestures
- How well-known individuals gesture

Key: Prompting allows Individualization and role specification

Well-known Individuals Gesturing

Style: Controlled, deliberate, and measured
Precision grip: thoughtful emphasis
Open palms: convey honesty, openness
One-handed chops: structured and logical delivery

Style: Bold, repetitive, and performative
OK sign: confidence, correctness
Pinching gesture: dramatization of size or precision
Double-handed chops: assert dominance, create rhythm

Data Annotations

We annotated a six-minute video speech of a labor activist

Classes of metaphoric gestures

- Progress
- Regress
- Cycle
- Collect
- Container
- Oscillation
- Temporal

Experimental Design

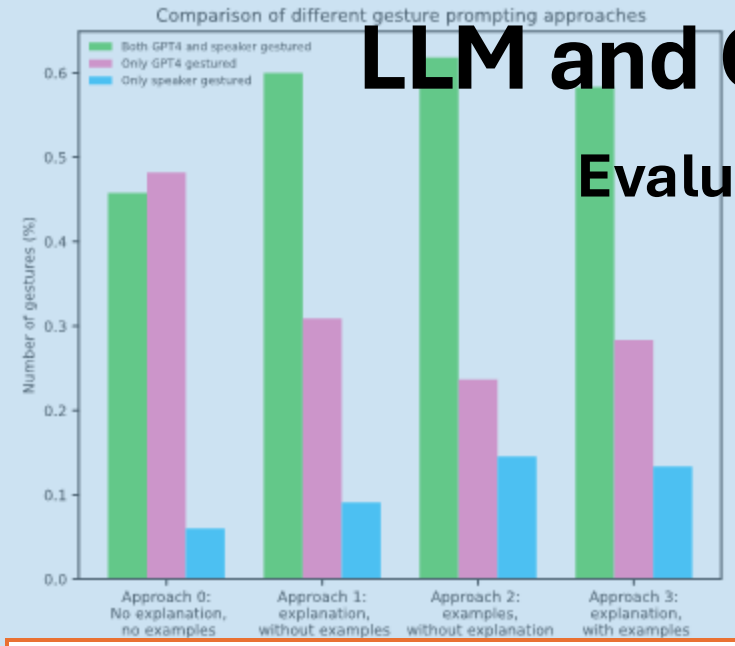
What are the impacts of alternative prompting approaches on the
1. Appropriateness of gesture selection,
2. Speed of inference?

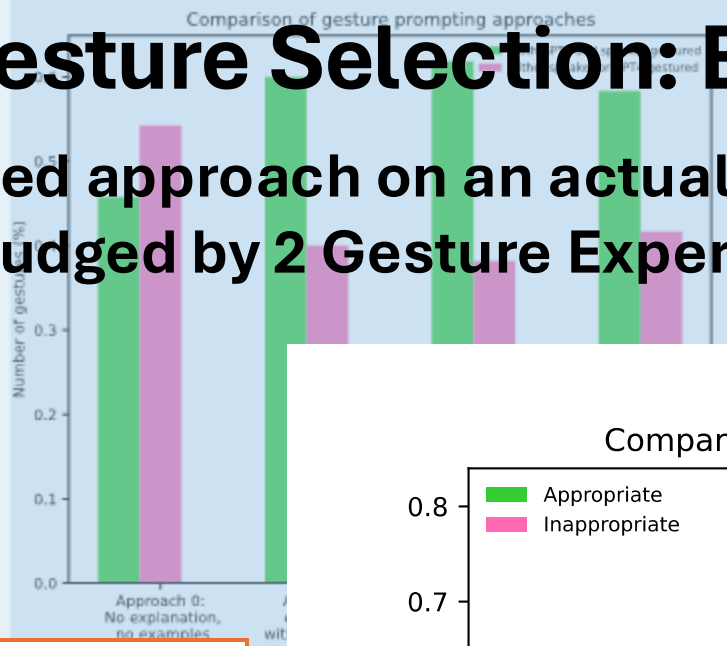
2X2 Factorial design of alternative prompting approaches

		Explanation of gesture classes provided	
		No	Yes
Annotated examples provided	No	Approach 0	Approach 1
	Yes	Approach 2	Approach 3

LLM and Gesture Selection: Evaluation

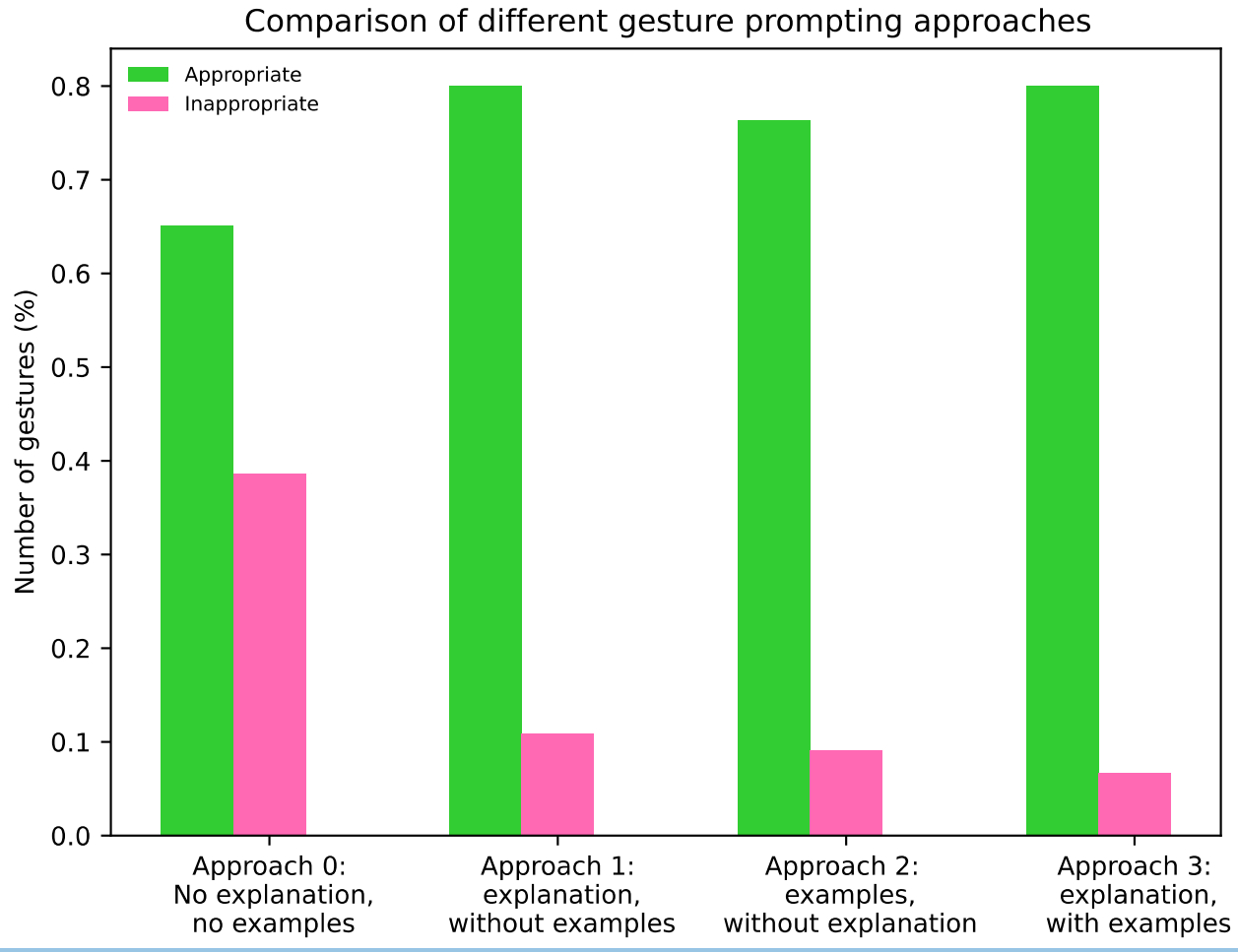
Evaluated approach on an actual speaker
Judged by 2 Gesture Experts





Explanation of gesture classes provided

Annotated examples provided	No	Yes
No	Approach 0	Approach 1
Yes	Approach 2	Approach 3



What Gesture to use??

GPT-4-Driven Discourse Analysis: Theme and **Rheme** → **New, emphasized information**

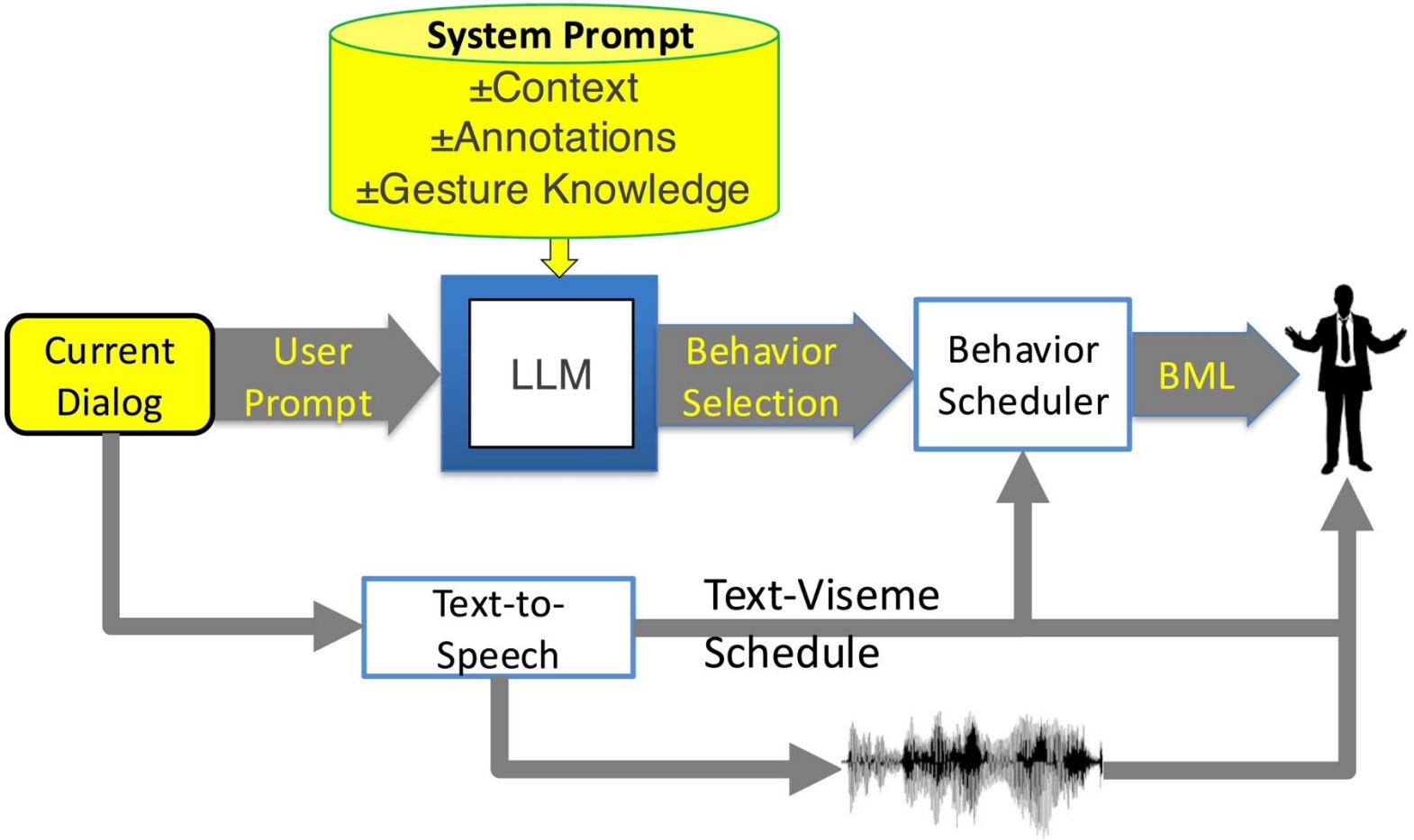
Gestures occur on the rheme parts of an utterance

Example:
And so I think always keeping those values in mind is really what propels people,
you know, to keep their eye on the prize and move these policies forward. → Progress gesture

Results:
43 out of 49 gestures occurred within the rheme identified by GPT-4

Implementation:
SIMA(Socially Intelligent Multimodal Agent)

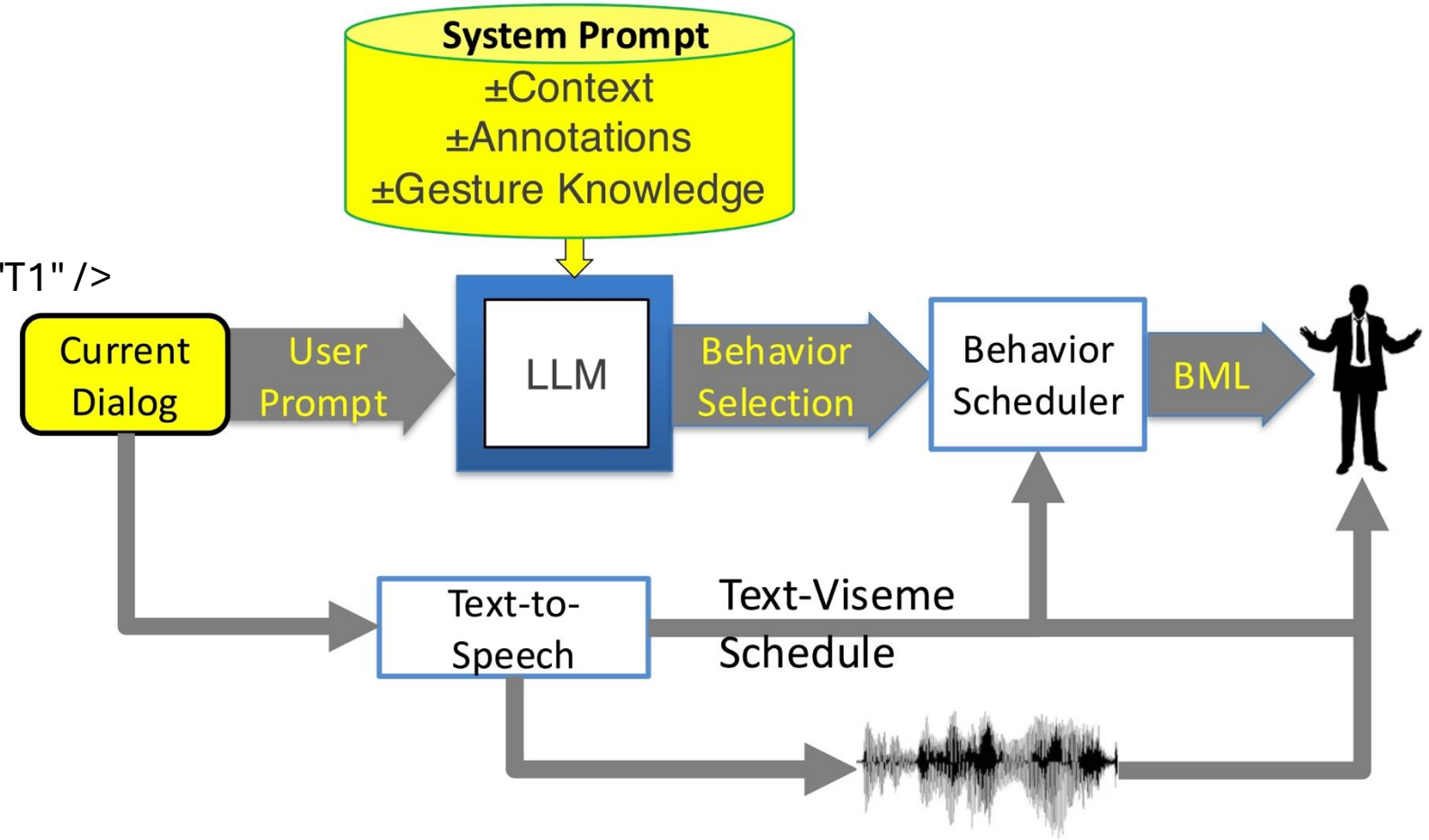
- Gesture Generation
  - Input processing
  - Gesture Prompting
- Animation realization
  - Behavior scheduler
  - Motion Controller



Implementation:
SIMA(Socially Intelligent Multimodal Agent)

- Gesture Generation
  - Input processing

<mark name="T0" />word<mark name="T1" />



Implementation:
SIMA(Socially Intelligent Multimodal Agent)

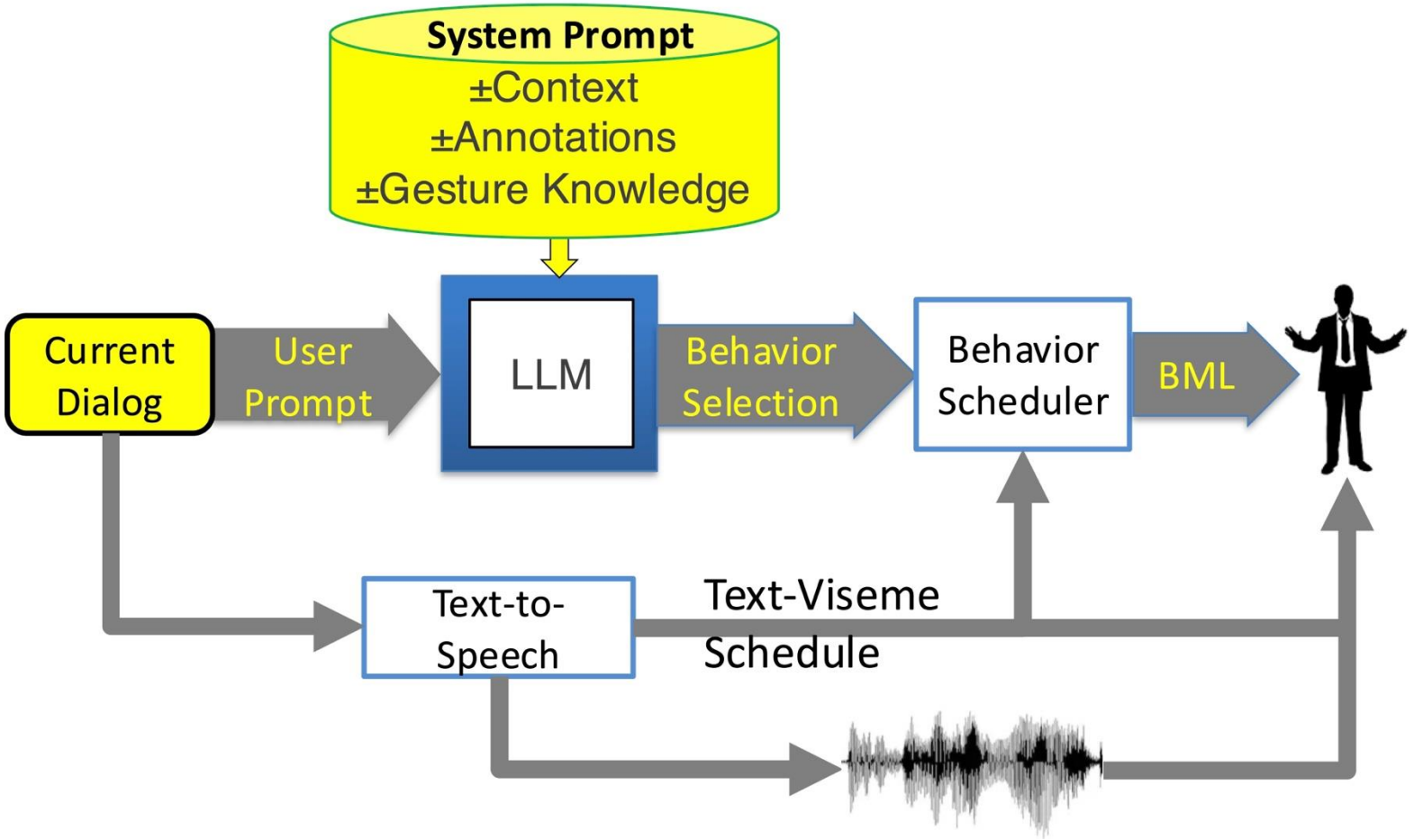
- Gesture Generation
  - Input processing
  - Gesture Prompting

Prompt:
Here is a list of Gesture types, their description, and some examples:

- Cycle
- Container
- Oscillation
- Negation
- Collect

Report the following in JSON format

- Gesture type
- The phrase where the gesture happens



Implementation:
SIMA(Socially Intelligent Multimodal Agent)

- Gesture Generation
  - Input processing
  - Gesture Prompting

<gesture stroke_start="sp1:T4" lexeme="SURROUND" type="METAPHORIC" emotion="neutral" />

<head id="action2" type="shake" amount="-0.2" repeats="0.5" start="sp1:T4" relax="sp1:T4+0.8" group="shake_nod" />

<head type="nod" velocity="1" amount="0.15" repeats="1" start="action2:start" relax="action2:relax+0.5" group="shake_nod" />'

Concluding Remarks

- **Overall, the semantic appropriateness of gestures was impressive.**
- **Alternative prompting approaches are suitable for different uses.**
  - **Prompting approach 0: No prior information on gestures**
    - Output: Flexible and creative gesture variations
    - Use case: Suggesting gestures to a designer
  - **Prompting approach 1,2,3: Prior information on gestures**
    - Output: Gestures constrained to the provided list of gestures/annotated examples
    - Use case: Automating gesture selection, personality/role consistency in gesturing
- **Rheme and Theme analysis can be used to narrow down the selection of gestures to gestures of the actual speaker.**