A Framework for Robot-Generated Mixed-Reality Deixis
Tom Williams | MIRRORLab | Colorado School of Mines

Introduction
Human-robot interaction conducted in pure reality has well-explored the use of deictic gestures to complement robots' natural language utterances. But these physical gestures are just one category of deictic gesture available in mixed reality. In this work we present a framework for categorizing mixed-reality deictic gestures and explore differences between these categories along several dimensions.

Egocentric Gesture
- Egocentric gestures are issued from the generator's perspective.
- Example: pointing to an object.

Allocentric Gesture
- Allocentric gestures are issued from the perspective of the generator's intended interlocutor.
- Example: circling an object within a teammate's field of view.

Multi-Perspective Gesture
- Multi-perspective gestures require the connection of multiple perspectives.
- Example: drawing an arrow from the robot to an object within a teammate's field of view.

<table>
<thead>
<tr>
<th>Category</th>
<th>Required Perspective</th>
<th>Embodiment</th>
<th>Capability</th>
<th>Dynamic Legibility</th>
<th>Static Legibility</th>
<th>Generation Cost</th>
<th>Maintenance Cost</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E)</td>
<td>Robot</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>(A)</td>
<td>Human</td>
<td>No</td>
<td>No</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>(M)</td>
<td>Human</td>
<td>Yes</td>
<td>No</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>(E,A)</td>
<td>Both</td>
<td>Yes</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>(E,M)</td>
<td>Both</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>(A,M)</td>
<td>Human</td>
<td>Yes</td>
<td>No</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>(E,A,M)</td>
<td>Both</td>
<td>Yes</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>

We analyze these three categories and the combinations thereof, across eight sample framework dimensions: Whose perspective is required to generate the gesture? Does the generator need to be embodied? Does the generator need physical gestural capabilities? Is the gesture legible while it’s being executed? Is the gesture legible once complete? How expensive is the gesture to generate? How expensive is the gesture to maintain? And is the gesture visible to bystanders?

References

Contact
Tom Williams (twilliams@mines.edu)
Web: inside.mines.edu/~twilliams
Lab: mirrorlab.mines.edu

MIRRORLab
Mines Interactive Robotics Research