A Computational Model for Embodied Visual Perspective Taking: From Physical Movements to Mental Simulation

Tobias Fischer and Yiannis Demiris
www.imperial.ac.uk/PersonalRobotics

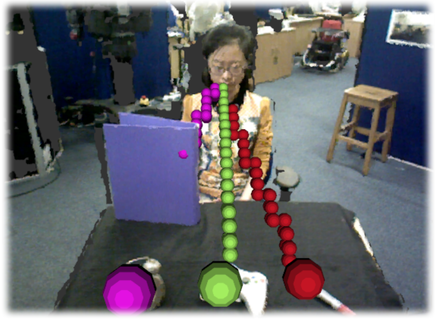
Imperial College London
PERSONAL ROBOTICS LABORATORY

CVPR 2018
SALT LAKE CITY • JUNE 18-22

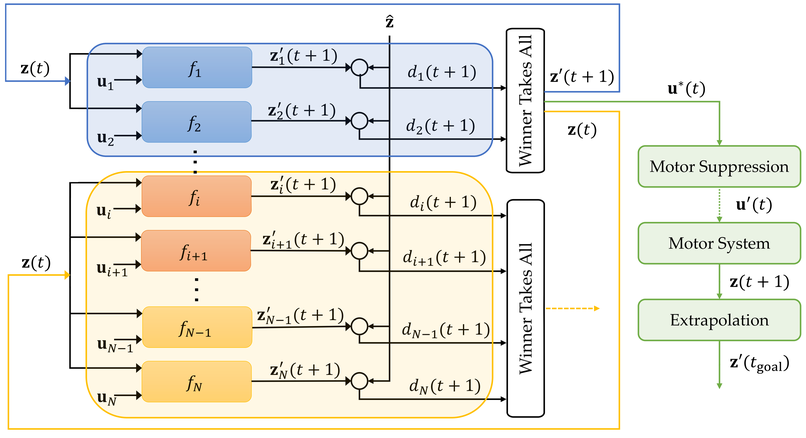
Introduction and Background:

"Perspective taking is the ability to assume another person's visual and/or cognitive viewpoint"

- Applications for perspective taking include **viewpoint invariant action recognition** and more natural **human-robot interaction**.
- Kessler & Thomson (2010) propose the embodied transformation account, where **perspective taking is the mental simulation of the physical movement** necessary to acquire another perspective.
- In Fischer & Demiris (ICRA2016) we suggested a **perspective taking framework for humanoid robots**.
- Here, we introduce a **computational model** to systematically test this account. We advocate that **perspective taking is governed by a competition process for visual attention** between multiple forward models, where execution of recent models is preferred.



Architecture Overview:



- Forward models $\{f_i(\mathbf{z}(t), \mathbf{u}_i(t)) \mid f_i \in \mathbf{F}\}$ provide an estimate of state $\mathbf{z}(t+1)$ given current state $\mathbf{z}(t)$ and motor command $\mathbf{u}_i(t)$.
- Forward model $f^*(t)$ that minimizes distance d is executed:

$$\mathbf{u}^*(t) = \underset{\mathbf{u}_i}{\operatorname{argmin}}\, d\big(f_i(\mathbf{z}(t), \mathbf{u}_i(t)), \hat{\mathbf{z}}\big)$$

- Whereby:

$$d(\mathbf{z}(t), \hat{\mathbf{z}}) = \underbrace{d_s(\mathbf{z}(t), \hat{\mathbf{z}})}_{\text{Euclidian distance}} + \underbrace{d_\theta(\mathbf{z}(t), \hat{\mathbf{z}})}_{\text{Angular disparity}}$$

- **Attentional component** selects subset $\mathbf{A}(t) \subset \mathbf{F}$ of forward models to be executed. $\mathbf{A}(t)$ contains $f^*(t)$ and other forward models based on similarity of $\mathbf{u}_i(t)$ and $\mathbf{u}^*(t)$.

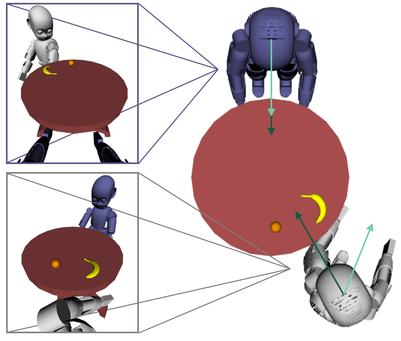
Paper & Code Download:

www.imperial.ac.uk/PersonalRobotics

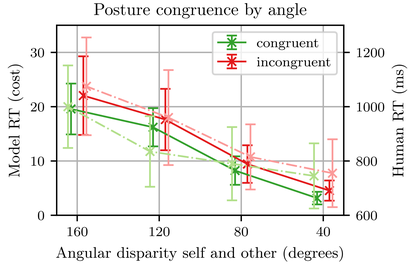
www.tobiasfischer.info

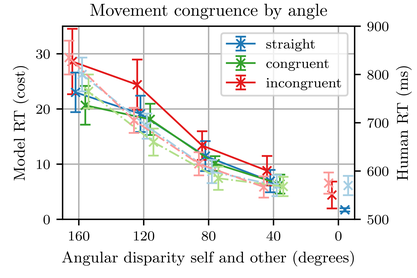
Experimental setup:

- Two simulated iCub humanoid robots performing a visual perspective taking task.
- One robot (self agent) has to decide whether a given object is to the left or right of the target agent.
- Posture and movement congruence are systematically varied.



Experimental results:





Solid, dark lines: model results; dashed, bright lines: human data

Conclusions:

- The model response time is similar to those of humans only if attentional component is employed. **We argue that humans implement an attentional mechanism similar to that of our model.**
- **Testable predictions: 1) Habituation effect** where response time depends on the congruence between the required movement direction of the current and previous mental simulations; **2) forced early response** leads to bias towards the egocentric perspective.