

# Show Me The Way

A GROUNDED FRAMEWORK FOR GESTURES AND ITS APPLICATIONS

SE367

Debidatta Dwibedi

# Objective

- A framework that is able to:
  - ▣ Learn gestures
  - ▣ Take descriptions while performing gestures as input
  - ▣ Associate gestures with words
- Use this framework to:
  - ▣ generate gestures helpful in route descriptions using an *embodied cognition agent*(ECA) i.e. a robot or a simulation of one.

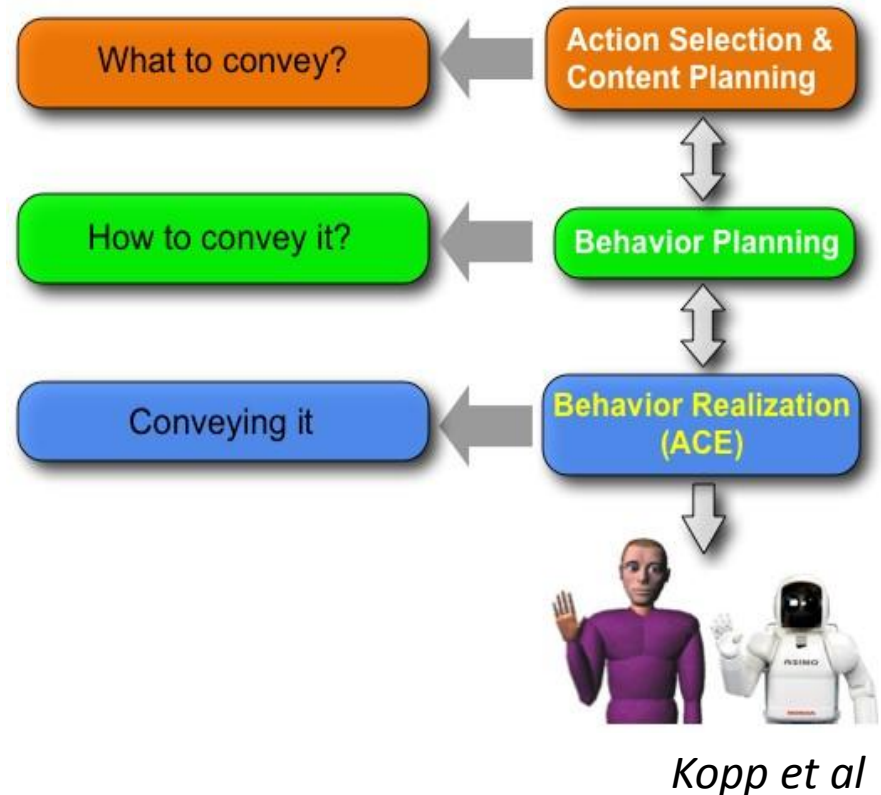
# Applications



Robot giving directions in a shopping mall

# Related Work

- ▣ Knowledge representation for generating locating gestures in route directions [Striegnitz et al]
- ▣ Integrated model of speech and gesture in robots [Kopp et al]
- ▣ Studied models of utterance, gesture and timing to better facilitate the human-robot interaction [Okuno et al]
- ▣ A joint model of language and perception for grounded attribute learning [Matuszek et al]



# Recording Gestures

- Used a Kinect to record gestures
  - ▣ Easy to get coordinates of joints and hands
  - ▣ Egocentric coordinates by subtracting hip coordinates
  - ▣ Used Frechet distance to compare query gesture with recorded gestures
  - ▣ One shot learning of gestures.
    - No training dataset or programming required
    - Humans can teach gestures without programming

# Extracting words and phrases from route/assembly descriptions

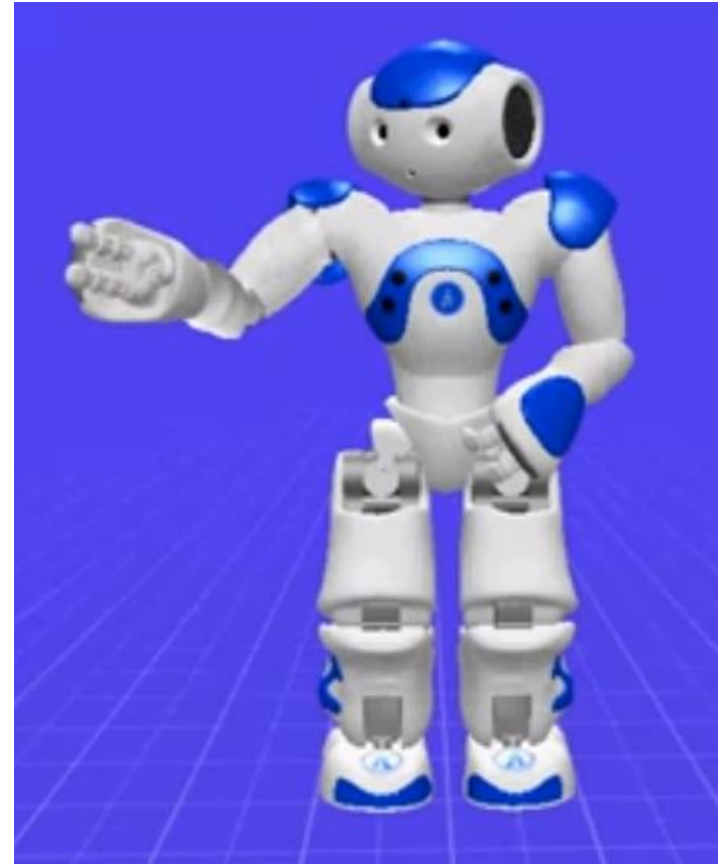
- Collected route descriptions to discover words and phrases that might need gestural representation
  - 37 route descriptions collected
  - Words and phrases discovered:
    - right, left , turn, straight, hall, road, walk, building
    - ('take', 'right'), ('take', 'left'), ('go', 'straight'), ('turn', 'left'), ('turn', 'right'), ('right', 'turn')
- Collected assembly instructions of a TV stand
  - Words and phrases discovered:
    - shelf, frame, glass, place, top, bottom, bolts
    - ('allen', 'wrench'), ('shelf', 'frame'), ('bottom', 'shelf') , ('glass', 'shelf'), ('top', 'shelf')

# Associating Words with Gestures

- Conditional probability to find most probable word associated with that gesture
- Maximise ratio of  $P(\text{Gesture} | \text{Word})$  to  $P(\text{All other gestures} | \text{Word})$

# Transferring Gestures to ECA

- Using Choreographe to simulate gestures
  - ▣ Nao robot is the chosen ECA
    - <http://youtu.be/VKo1L9OzB2c>
    - Easily transfer these gestures to a robot
    - Cannot orient body in direction of turn
- Map coordinates from the recorded gesture to joint angles of the robot



“Take Right”



# References

- Tversky, Barbara, et al. "Explanations in gesture, diagram, and word." *Spatial Language and Dialogue, Coventry, KR, Tenbrink, T., and Bateman, J.(Eds.), Oxford University Press, Oxford (2009)*.
- Kopp, Stefan, Kirsten Bergmann, and Ipke Wachsmuth. "Multimodal communication from multimodal thinking—towards an integrated model of speech and gesture production." *International Journal of Semantic Computing 2.01 (2008): 115-136*.
- Okuno, Yusuke, et al. "Providing route directions: design of robot's utterance, gesture, and timing." *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on. IEEE, 2009*.
- Striegnitz, Kristina, et al. "Knowledge representation for generating locating gestures in route directions." *Proceedings of Workshop on Spatial Language and Dialogue (5th Workshop on Language and Space). 2005*.
- Matuszek, Cynthia, et al. "A Joint Model of Language and Perception for Grounded Attribute Learning." *arXiv preprint arXiv:1206.6423 (2012)*.