



Source: <https://www.youtube.com/watch?v=h1E-FlguwGw&t=24s>

## Lecture 7

# “Show me your moves.”

### Nonverbal communication

Kim Baraka  
Assistant Professor  
Social AI group

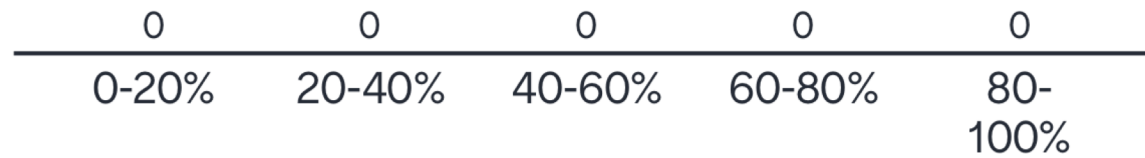
# Learning goals

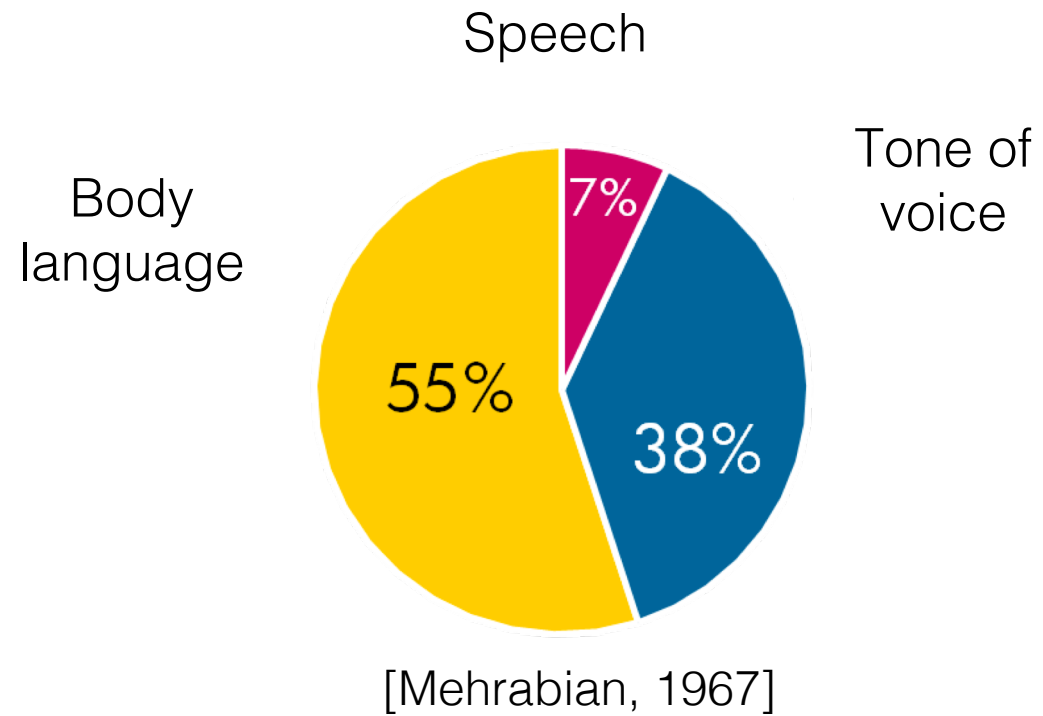
- Get familiar with some *modalities* of human-human non-verbal communication
- Apply and expand on these modalities in the context of HRI
- Design a user study to *evaluate* the effectiveness of non-verbal robot behavior

Go to [www.menti.com](https://www.menti.com) and use the code 4913 0463



## How much of human communication is non-verbal?





Go to [www.menti.com](http://www.menti.com) and use the code 4913 0463

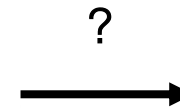
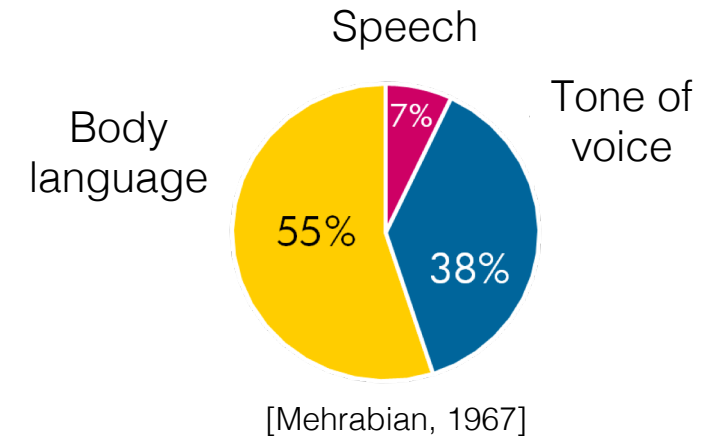
## What are some non-verbal modalities in humans?

<https://www.mentimeter.com/s/337c49f761744a84a58553d3d8eca9b2/fb7c58897234/edit>



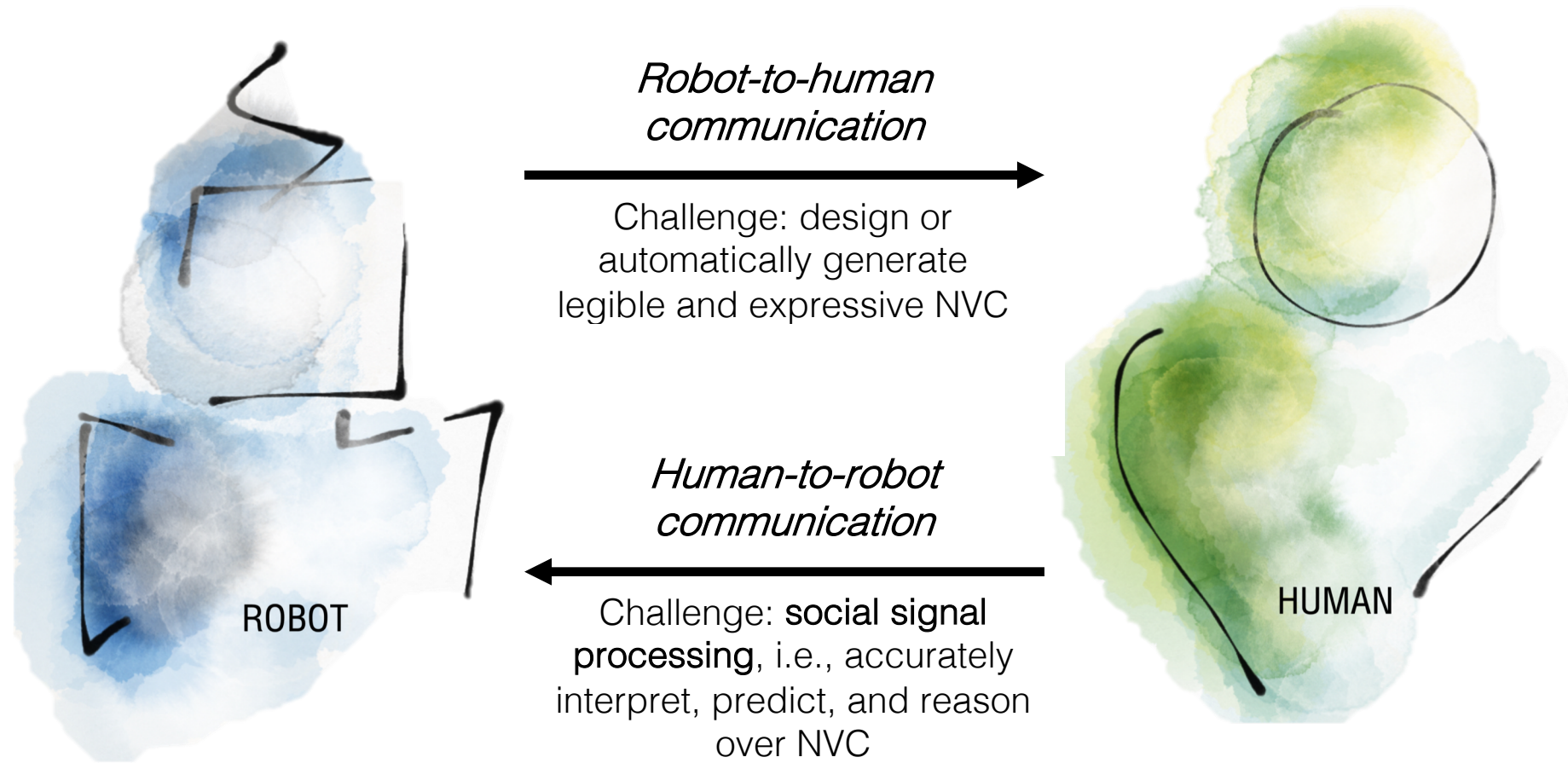
# Non-verbal communication (NVC)

- Speech is only a small percentage of human communication
  - NVC *modalities* : different ways in which a social agent communicates information without words  
→ social cues
  - Challenges for robot NVC
    - Translate principles of human NVC to robots
    - Different embodiment than humans
    - Unique modalities (e.g., lights, sound)
- Embodied interaction: “occurring in real time and real space” (Dourish, 2001)



Sphero robot

# Communication is always a two-way street



# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities



# Gaze

- Gaze = where and how one looks
- People are uniquely sensitive to gaze – important to get it right on robots
- For non-anthropomorphic robots, gaze doesn't necessarily have to involve realistic-looking eyes
- Three types of gaze
  - Mutual gaze (eye contact)
  - Deictic (“pointing” with your eyes)
  - Joint attention
- Types of eye movement
  - Fixation
  - Saccades
  - Smooth pursuits



JIBO “looking” at a cup

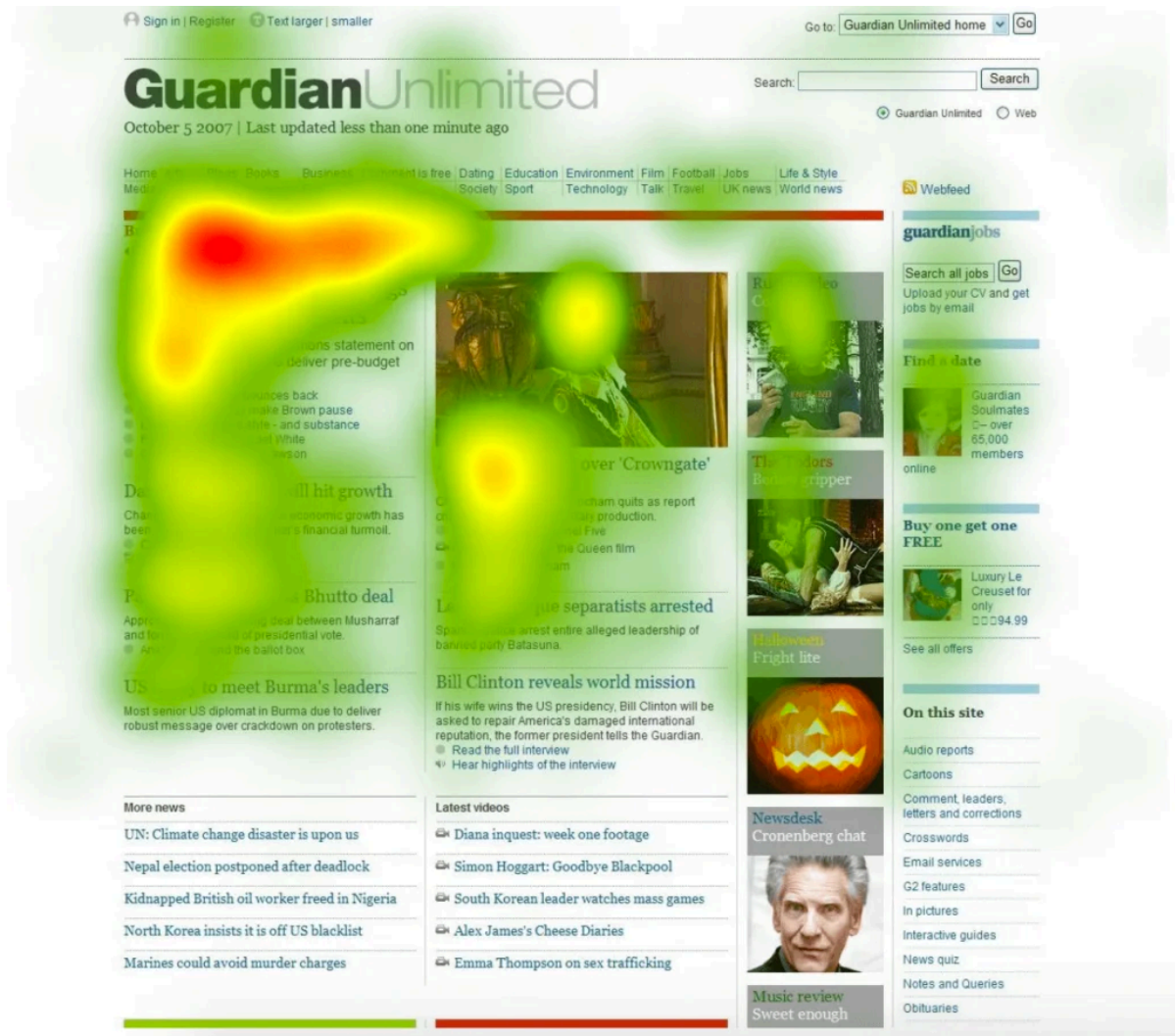
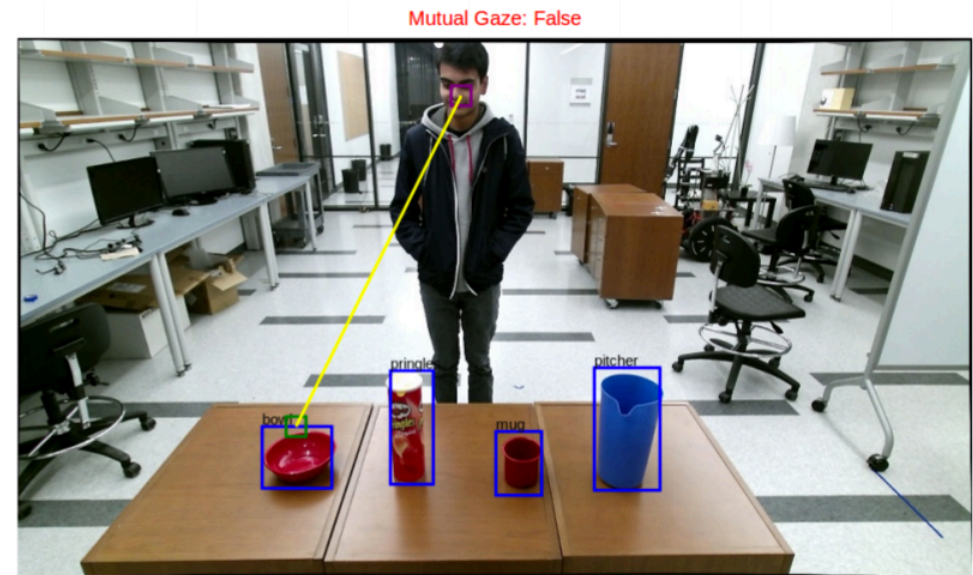


Image Source: talkroute



Saran et al. (2018)

# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities



Gustavo Dudamel (source: YouTube)

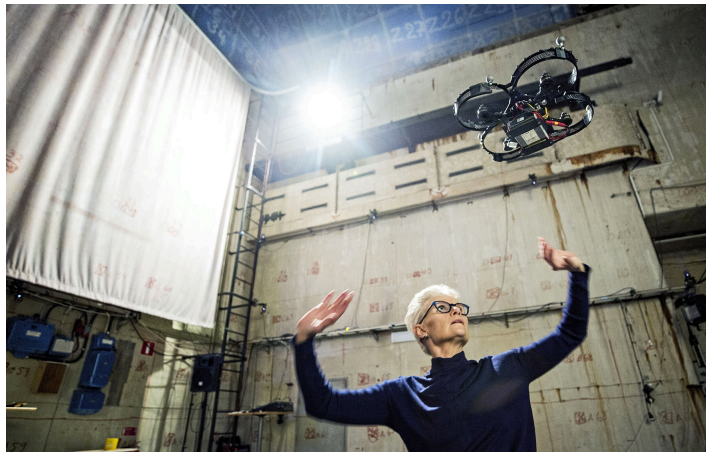
# Gestures

A movement usually, of the body or limbs, that expresses or emphasizes an idea, sentiment, or attitude

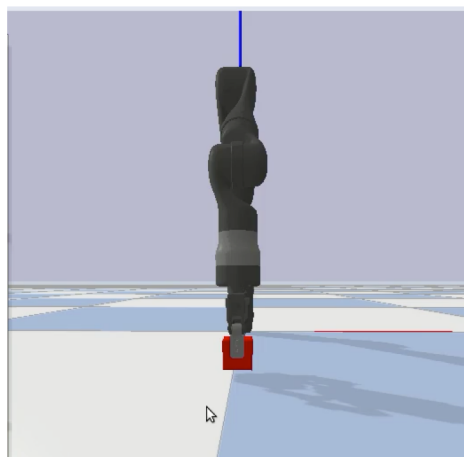
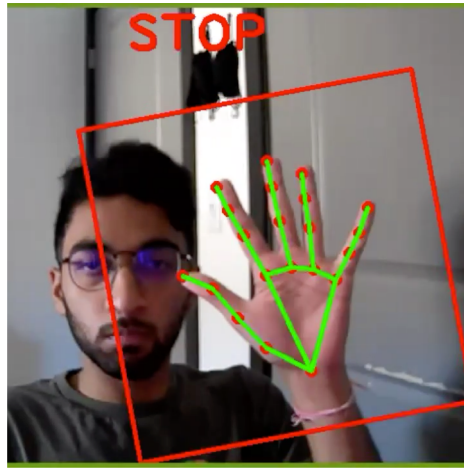
- Iconic – represent concrete objects or actions
- Metaphoric – representation abstract concepts
- Deictic – references entities or locations in space
- Beat – follows speech rhythms

Gestures are strongly culturally-dependent

# Gesture-controlled robots



Opera Mecatronics, Aerial robotic choir (KTH)

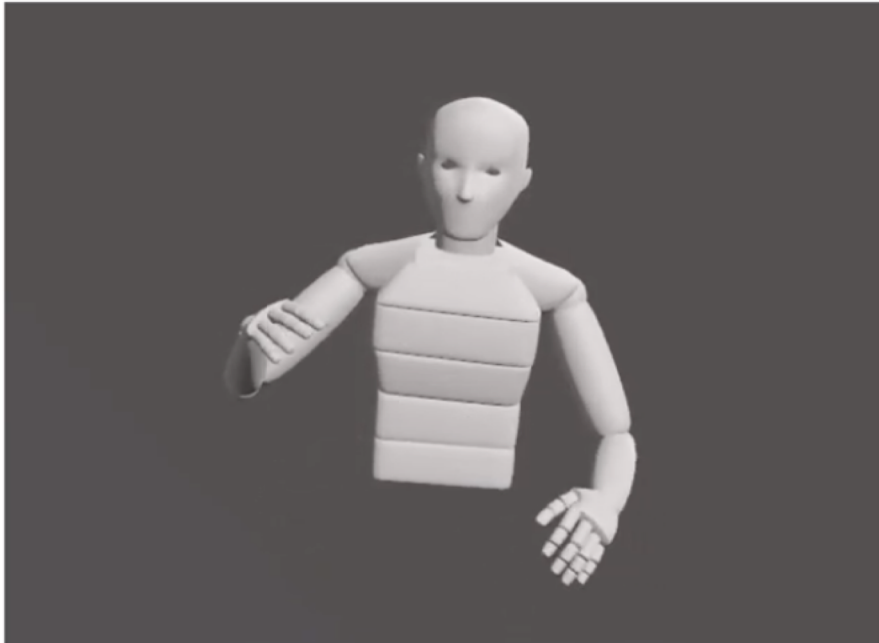


Work in progress in collaboration with UT Austin

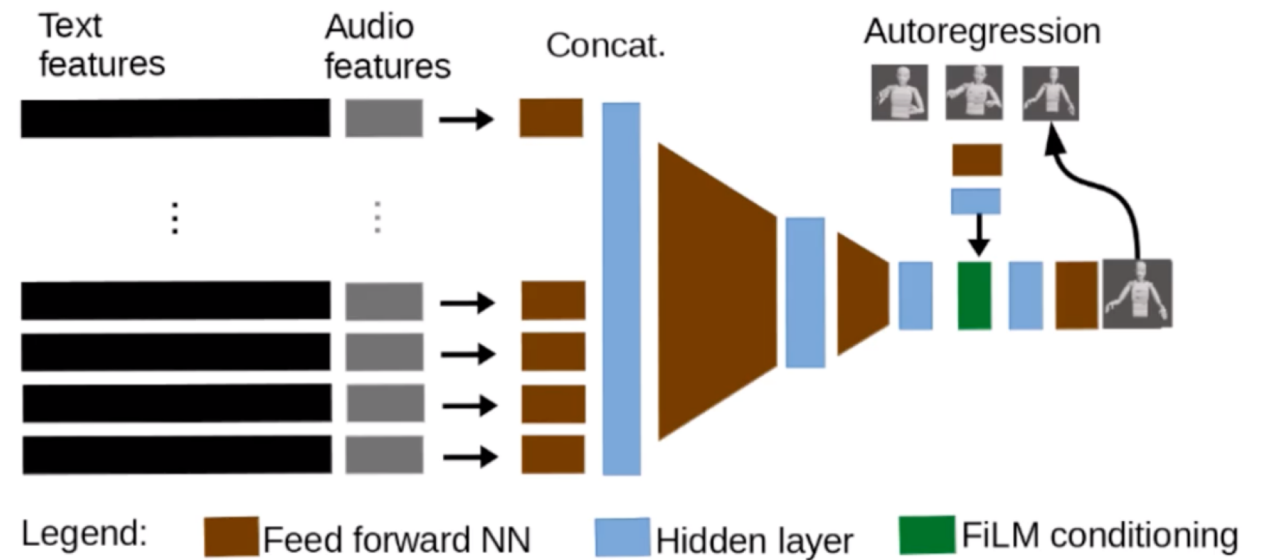
## Some methods for gesture classification

- RGB vision-based
- Depth vision-based (Leapmotion)
- Motion-capture (wearable sensors, or marker-enabled)
- Glove-enabled (Pison <https://www.youtube.com/watch?v=bsF7be6wBrg>)

# Automatic generation of speech-accompanying gestures



## Gesture Generation Framework



Gesticulator: A framework for semantically-aware speech-driven gesture generation (Kucherenko et al. 2020)

# Activity

*For people in the classroom:*

Student 1: explain to student 2 to how to get from here to the lab.

Student 2: Observe NVC, especially types of gaze and gesture

*For people at home:*

If you have company, ask them to tell you how to get to the nearest supermarket

## *Cheat sheet*

### Gaze

- Mutual gaze
- Deictic
- Joint attention

### Gestures

- Iconic
- Metaphoric
- Deictic
- Beat



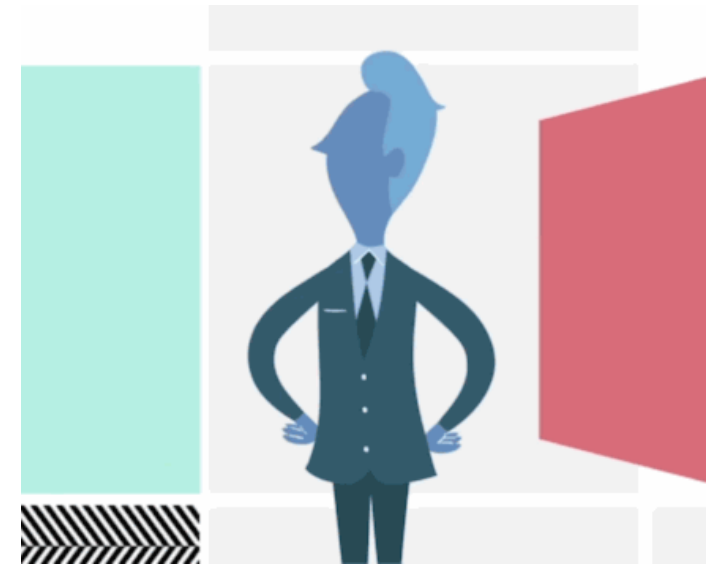
What kind of NVC did you use?

# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities

# Expressive motion

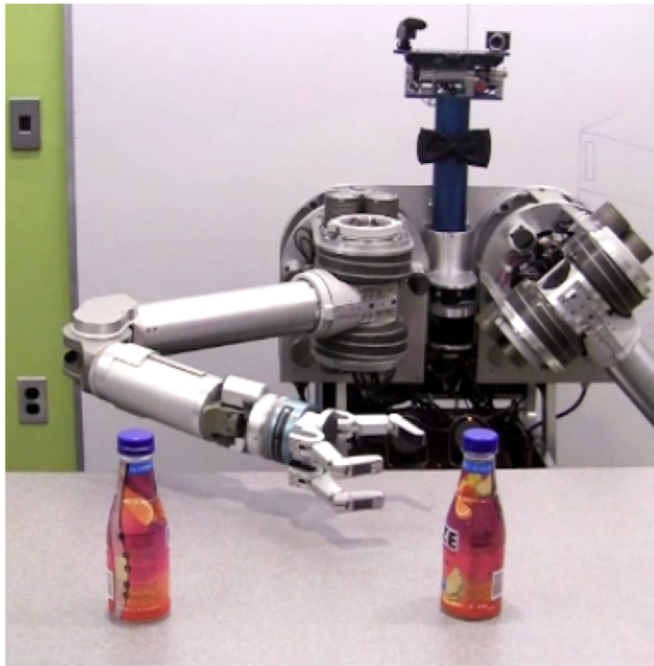
- Motion with a primarily functional purpose, modulated for a communicative purpose
- Expresses a hidden component of the robot's program (e.g., state, goal, intent, performance, affect, etc.)
- Can interfere with robot operation or performance



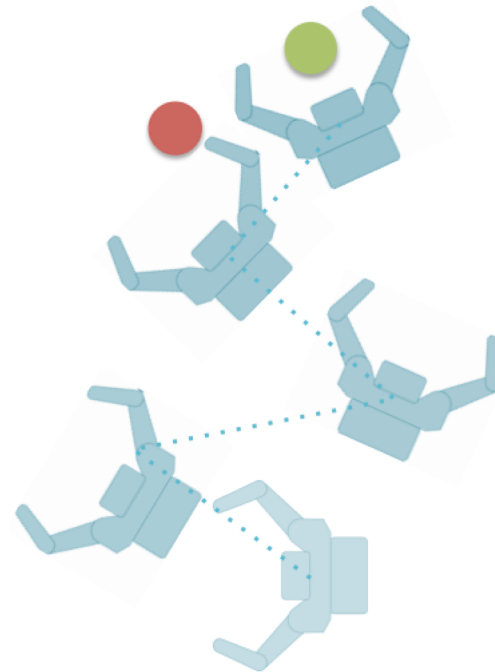
Scandis (IdeaRocket) (Source: Vimeo)

# Expressive motion (manipulator)

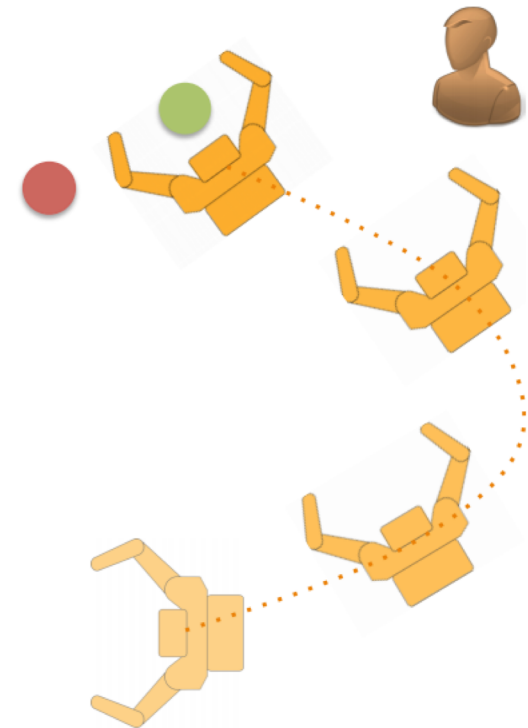
[Dragan, 2015]



HERB manipulator

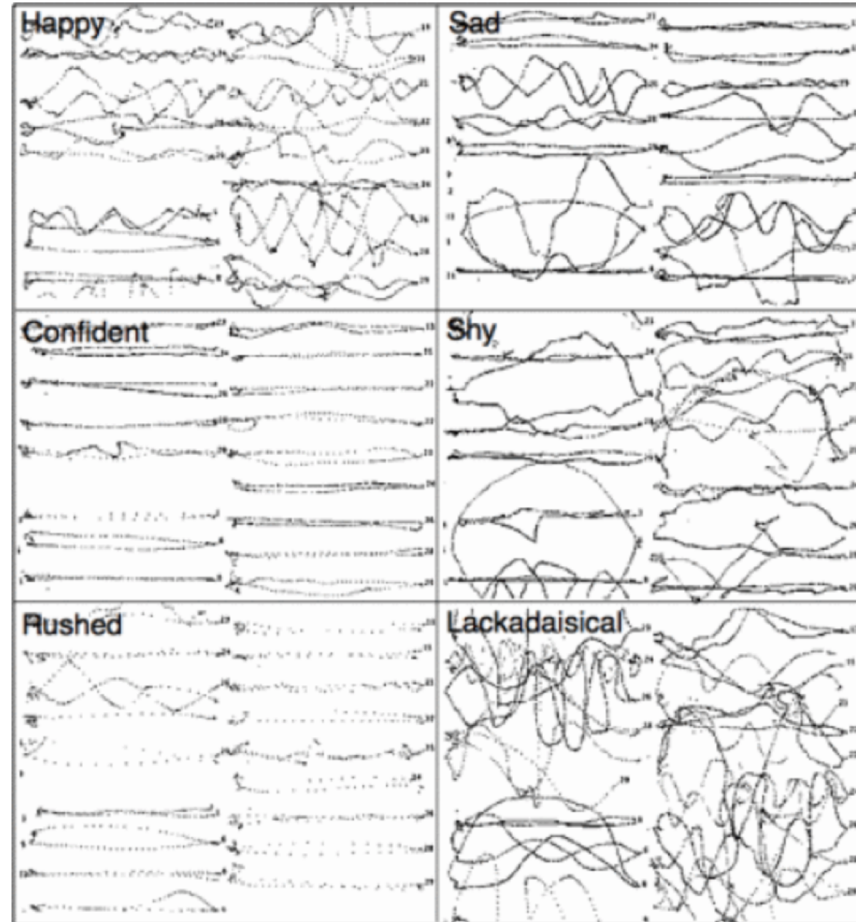
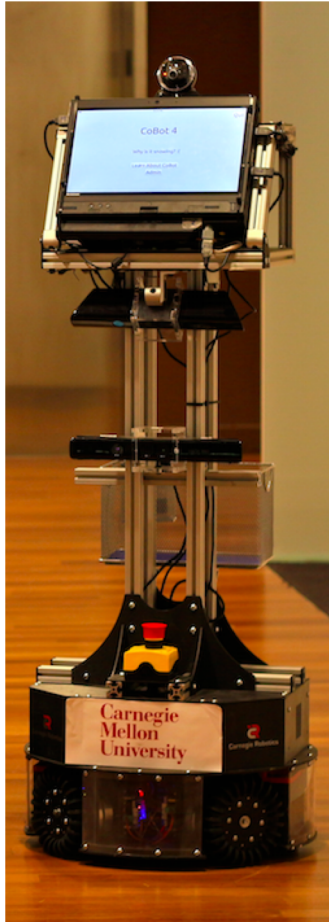


Functional motion



Legible motion

# Expressive motion (mobile robot) [Knight et al., 2014]



- Motions in x, y, and theta inspired by the Laban effort framework (dance)
- Design of trajectories involved the participation of actors

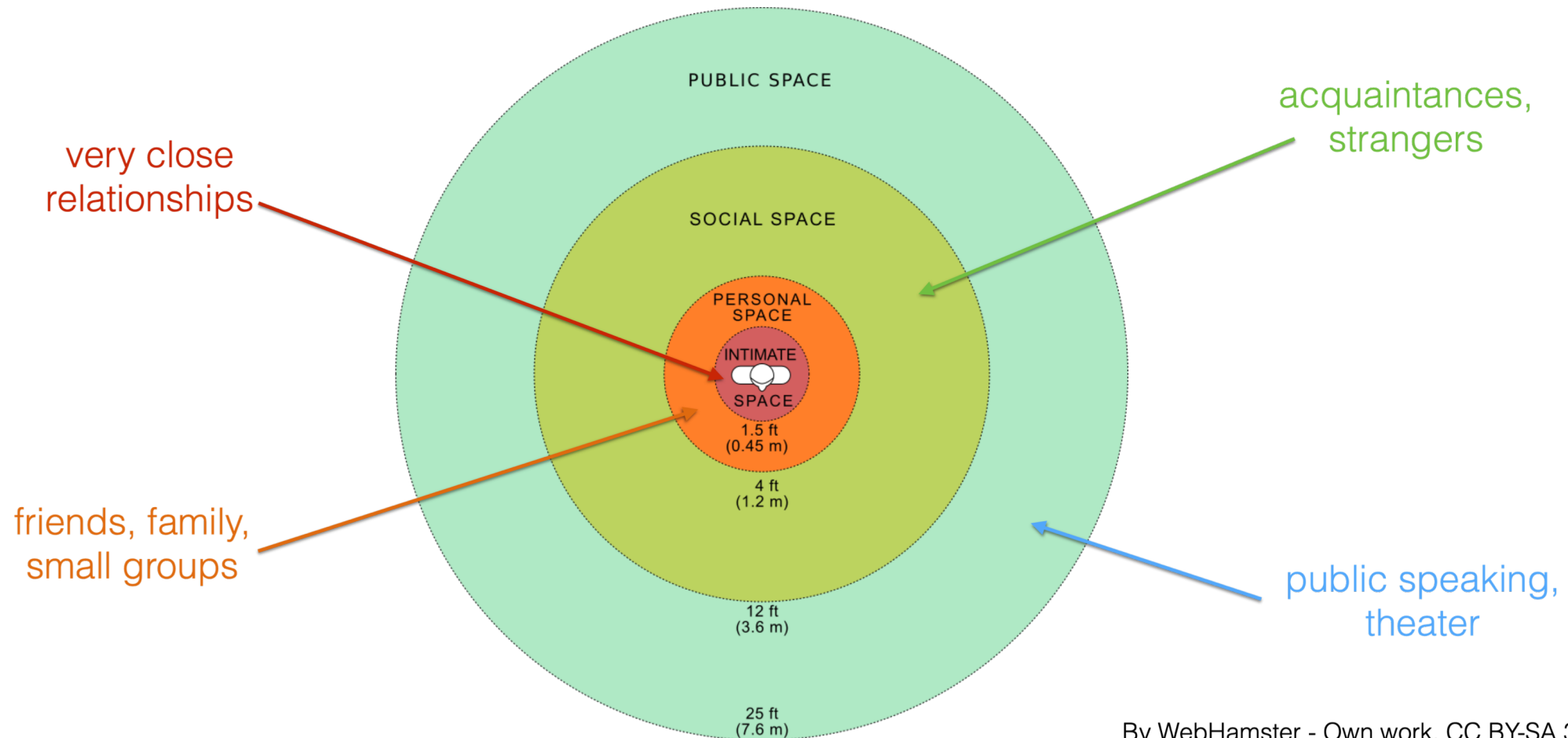
# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities

# Proxemics

- "the study of spatial distances individuals maintain in various social and interpersonal situations." — Rios-Martinez et al. (2015)
- First introduced by Edward T. Hall in 1966
- A key part of interpersonal communication
- Strongly influenced by
  - personal factors (age, gender)
  - environmental factors (task, noise, context)
  - societal factors (culture)

# Hall's interpersonal distances

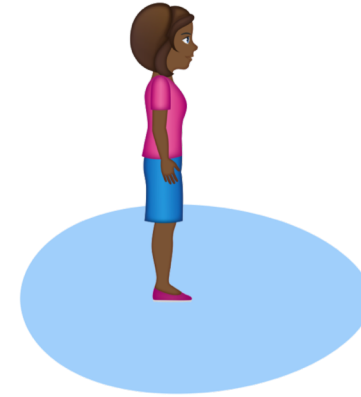


By WebHamster - Own work, CC BY-SA 3.0  
<https://commons.wikimedia.org/w/index.php?curid=6147809>



# Personal space

- Personal space is egg-shaped, longer in front than in the rear (Hayduk et al., 1981)
- When physical proximity is out of sync with social connection, people diffuse the discomfort by avoiding eye contact and conversation (Shaver, *Principles of Social Psychology*, 2015)



Quora user Ciril J Thundiylil

# Outline

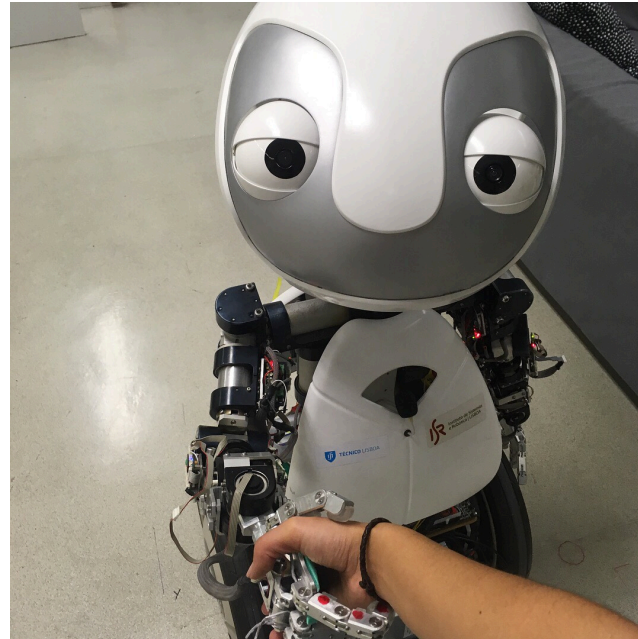
- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities

# Haptics

Touch, texture, pressure, force, etc.



PARO being touched.  
Credits: New York Times



Avelino et al. (2018)



Granados et al (2017)



Contact improvisation with a Baxter robot:  
communication through weight sharing,  
negotiation of pressure, and exploring  
leader-follower dynamics

(in collaboration with Isabel Valverde,  
Ana Moura, and Nuno Leite)



Collaborative painting through shared control  
of a paintbrush

(in collaboration with Robert Zacharias, Yeliz Karadayi,  
and Su Baykal)

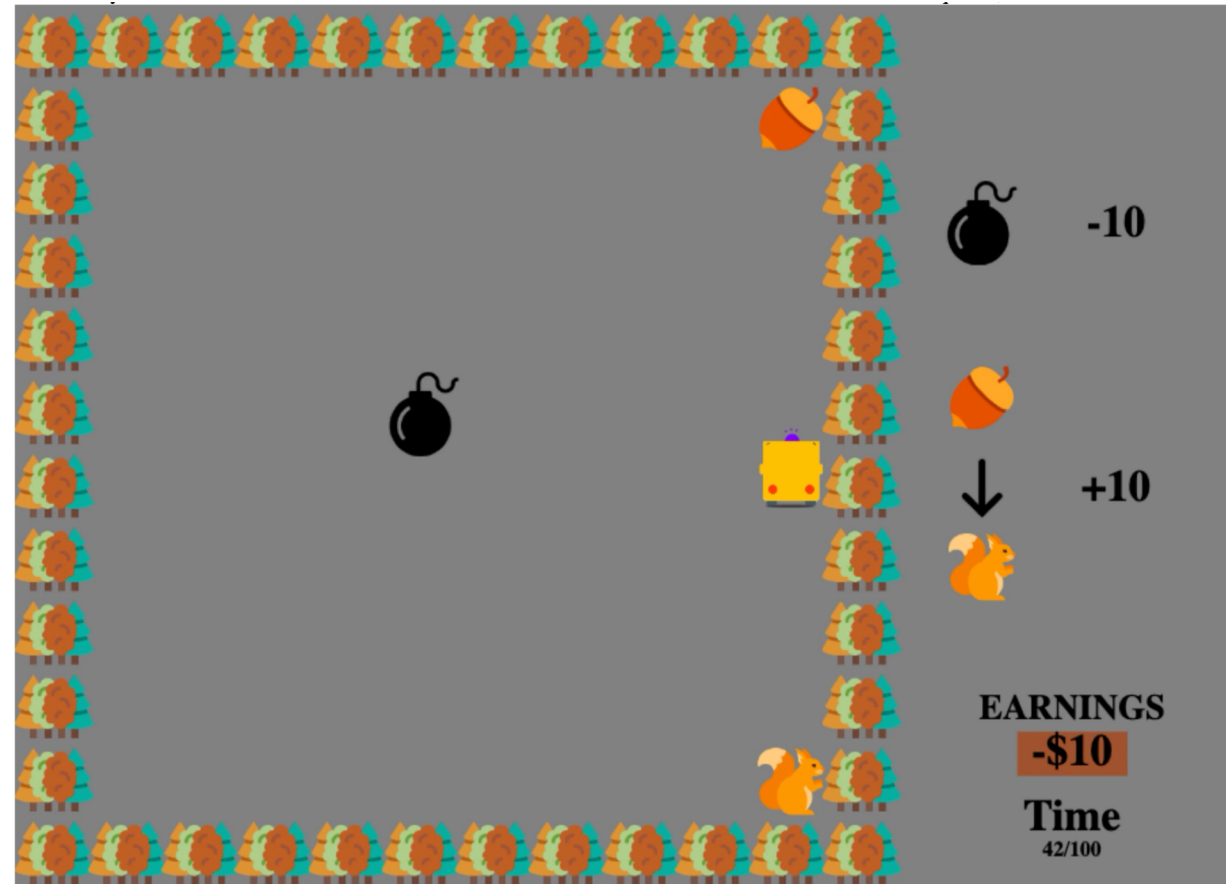
# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities

# Prosody

“The rhythm, stress, and intonation of speech” (APA)  
Not **what** is being said, but **how** it is being said

- Appropriately modulating prosody on a robot is difficult
- Extracting prosodic cues from human voice can be used to extract more task-relevant or internal state information



Prosody sensitive learning: Work in progress in collaboration with UT Austin

# Outline

- Gaze
- Gestures
- Expressive motion
- Proxemics
- Haptics
- Prosody
- Robot-specific modalities

Go to [www.menti.com](https://www.menti.com) and use the code 5711 7498

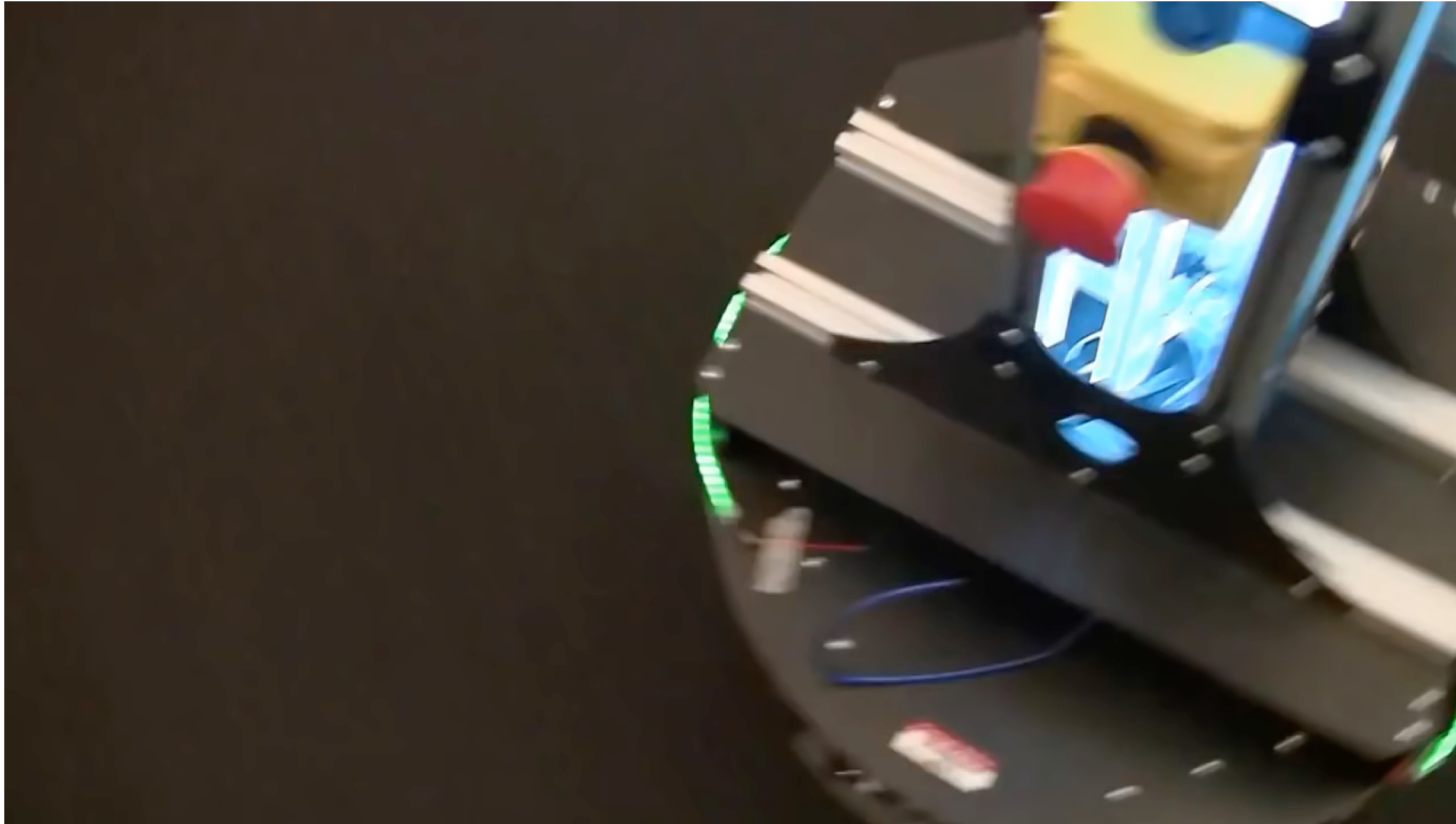
# Robot-specific NVC modalities?





# Expressive lights

[Baraka et al., 2018]



- Increases state and action *transparency* (What is the robot “thinking”? Why did the robot do that?)
- Contributes to predictability and trust

How to evaluate that a given non-verbal robot behavior communicates effectively?

# Evaluation methods

- Identify the intended *goal* of the communication (e.g., reveal internal state, communicate goal or intent, express affect, etc.)
- Select appropriate *measures* for its success (people guess right, people adapt their behavior, people trust more, etc.)
- Show the robot performing the same task, under two *conditions*:
  - (C1): With NVC
  - (C2): Without NVC
- *Compare* the measures under both conditions using statistical tests

# Example: expressive lights

[Baraka et al., 2016]

*Goal of communication:*

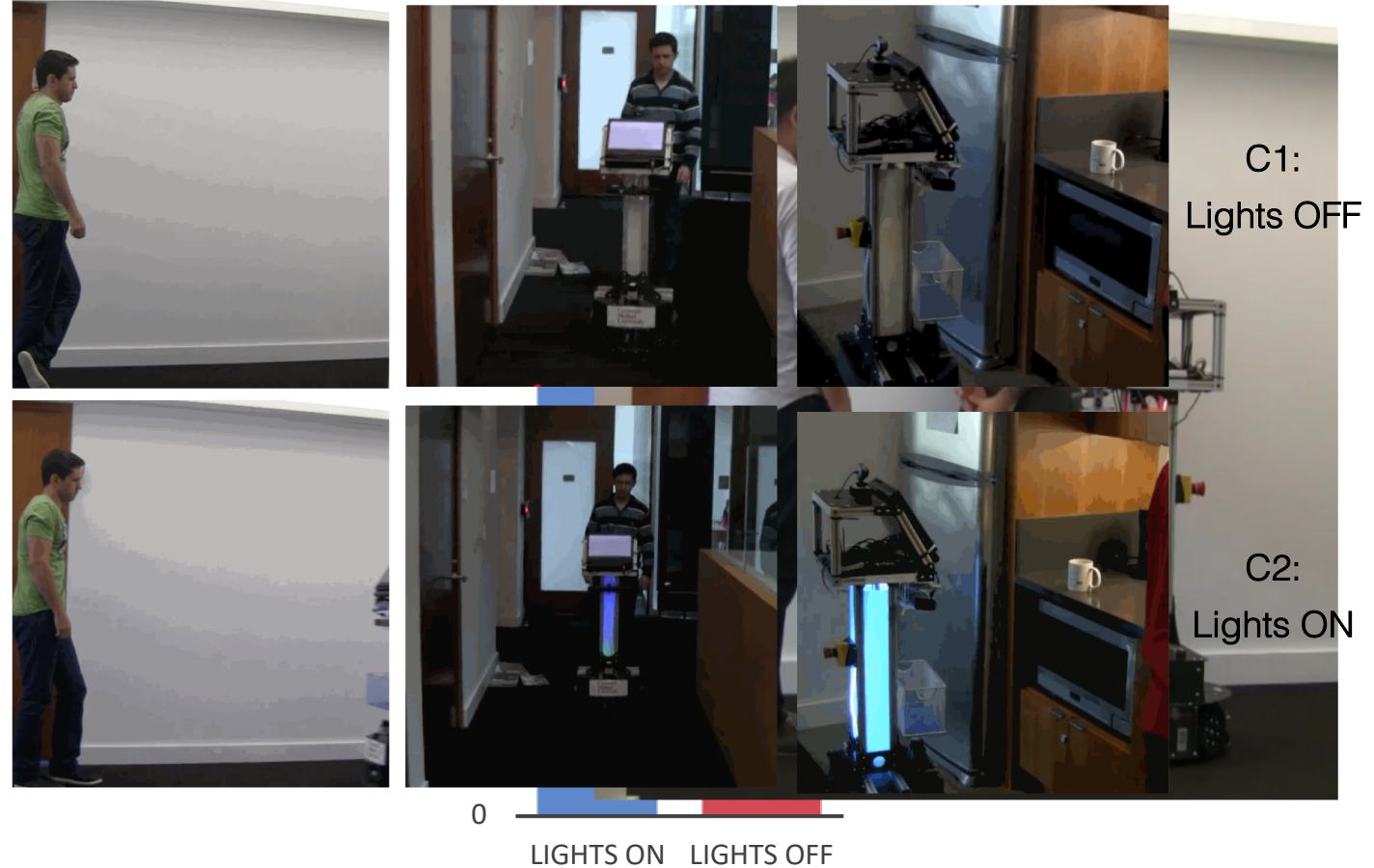
Reveal internal state

*Measure:*

*Accuracy* on answering questions about robot state (multiple choice), e.g.:

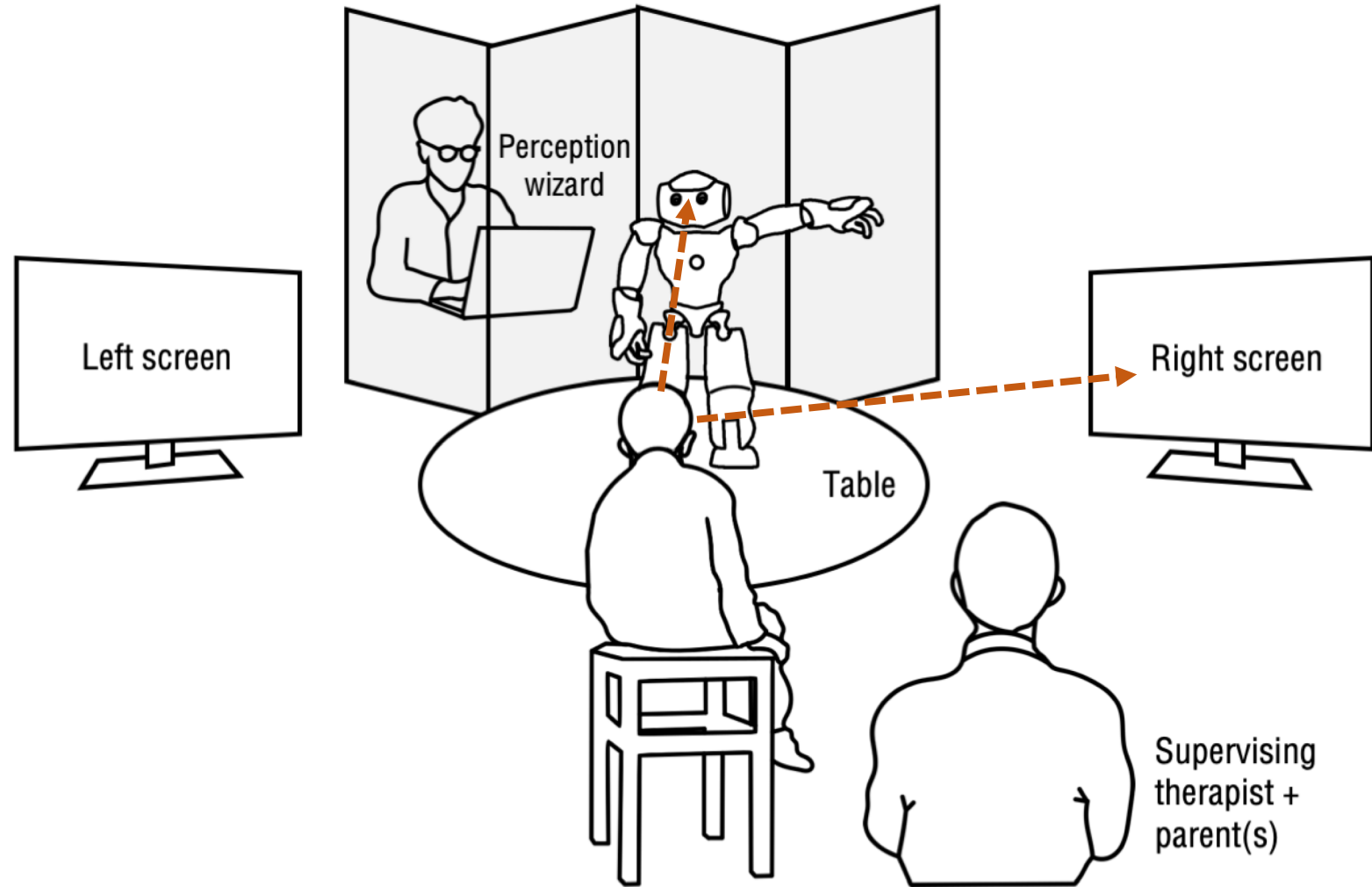
- “What is the robot doing?”
- “Why did the robot...?”

Accuracy (%)



# Multi-modal NVC

- Multi-modal NVC usually means richer NVC
- Challenges of interpreting multi-modal human NVC signals?
- Challenges of producing robot NVC?
- What are some cases of undesirable multi-modal NVC?

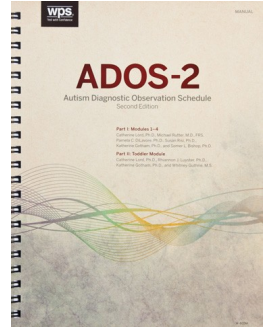


“An Optimization Approach for Structured Agent-Based Provider/Receiver Tasks” (Baraka et al., 2019)

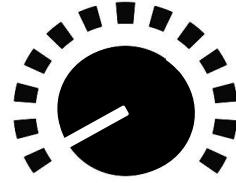
# Action hierarchies

Increasingly assistive ↓

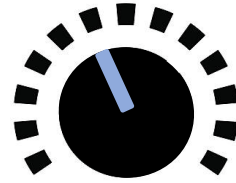
Level	JATT	NAME
1	Speech+gaze	Brief speech
2	Speech+gaze+pointing	Speech
3	Speech+gaze+pointing+video	Speech+lights
4	Speech+gaze+pointing+video+sound	Speech+lights+motion



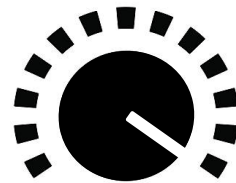
“An Optimization Approach for Structured Agent-Based Provider/Receiver Tasks” (Baraka et al., 2019)



Under-assist



Just-right



Over-assist



# Summary

- NVC is essential for *embodied interaction* with humans  
Contributes to fluidity, transparency, trust,...
- NVC modalities include (but are not limited to): gaze (mutual gaze, deictic, joint attention, ...), gestures (iconic, metaphoric, beat, ...), expressive motion, proxemics (interpersonal distances and personal space), haptics (physical HRI), and prosody
- Robot-specific modalities include lights, sound, and robot-specific motion
- NVC is still a vast open area of research in HRI