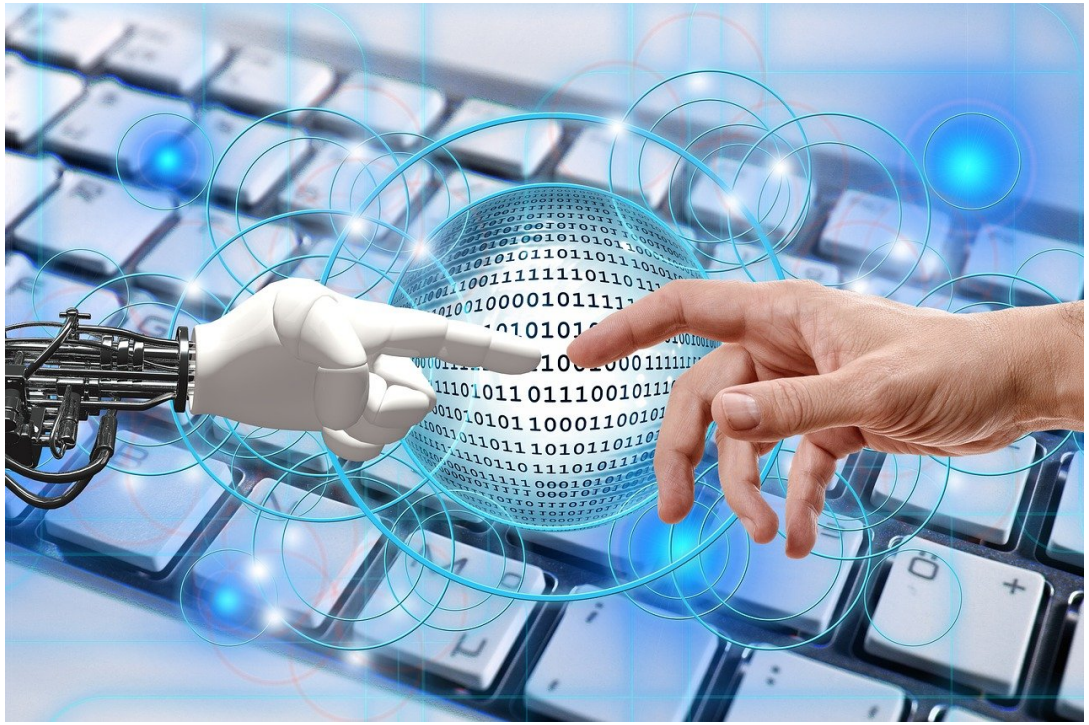
Lecture 03 - Nonverbal Communication in Humans and Robots



Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

Department of Computer Science
Hochschule Bonn-Rhein-Sieg
Sankt Augustin

25th April 2024

Learning Goals

- At the end of today's lecture, you will be able to:
  1. List the communication modalities in humans and robots.
  2. Identify the key elements and concepts of nonverbal communication.
  3. Elucidate the technical aspects of realizing multimodal communication between humans and robots.

Why do We Communicate?

- Communication is about sending and receiving messages.
- We communicate:
  - To exchange information.
  - To understand each other's behaviors, goals, intentions, mental states, opinions, perspectives, preferences, etc.
  - To coordinate or synchronize actions between multiple individuals.
  - To operate cohesively within groups.
  - To satisfy own as well as shared goals.
  - To survive.
  - ...

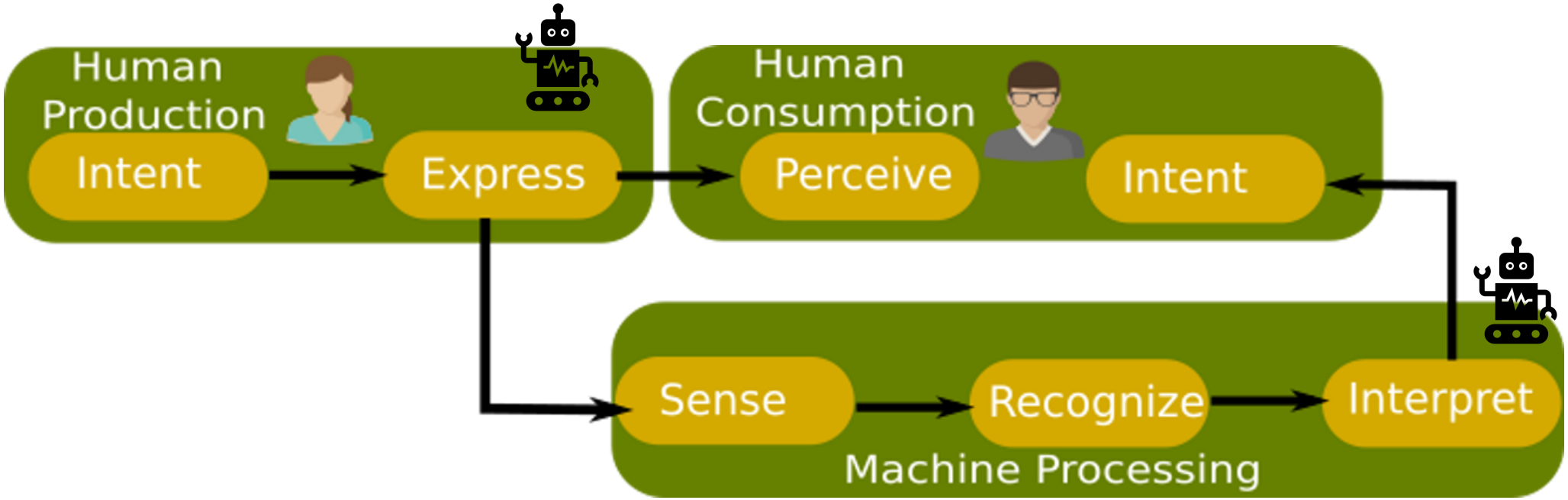
Is Communication Easy?

https://youtu.be/dBT6u0FyKnc

- Obviously not!
- Researchers have looked at how humans communicate, in order to identify:
  - The **elements** of communication
  - The **modes** of communication
  - The **processes** involved in communication
  - The **factors** that influence communication
- .. and to create **a model for interhuman communication and comprehension**...
- … which could be used to study communication problems as well as to design artificial communicative agents.

Human Communication (1/2)

- Involves two broad categories of processes:
  1. Perceiving and interpreting signs
  2. Generating and expressing signs



Recreation of Fig. 3 in (Narayanan and Georgiou, 2013)

Human Communication (2/2)

- Has two components:
  1. Verbal
     - Involves the use of a language
     - Discrete signals (phonemes, words)
     - E.g. spoken language, sign language
  2. Nonverbal
     - Involves movement of different parts of the body, often in sync with verbal component
     - Continuous signals
     - E.g. intonation, muscle or joint movement.

Nonverbal Communication

https://youtu.be/ilmDN2e_FIc

What nonverbal behaviours do you observe?

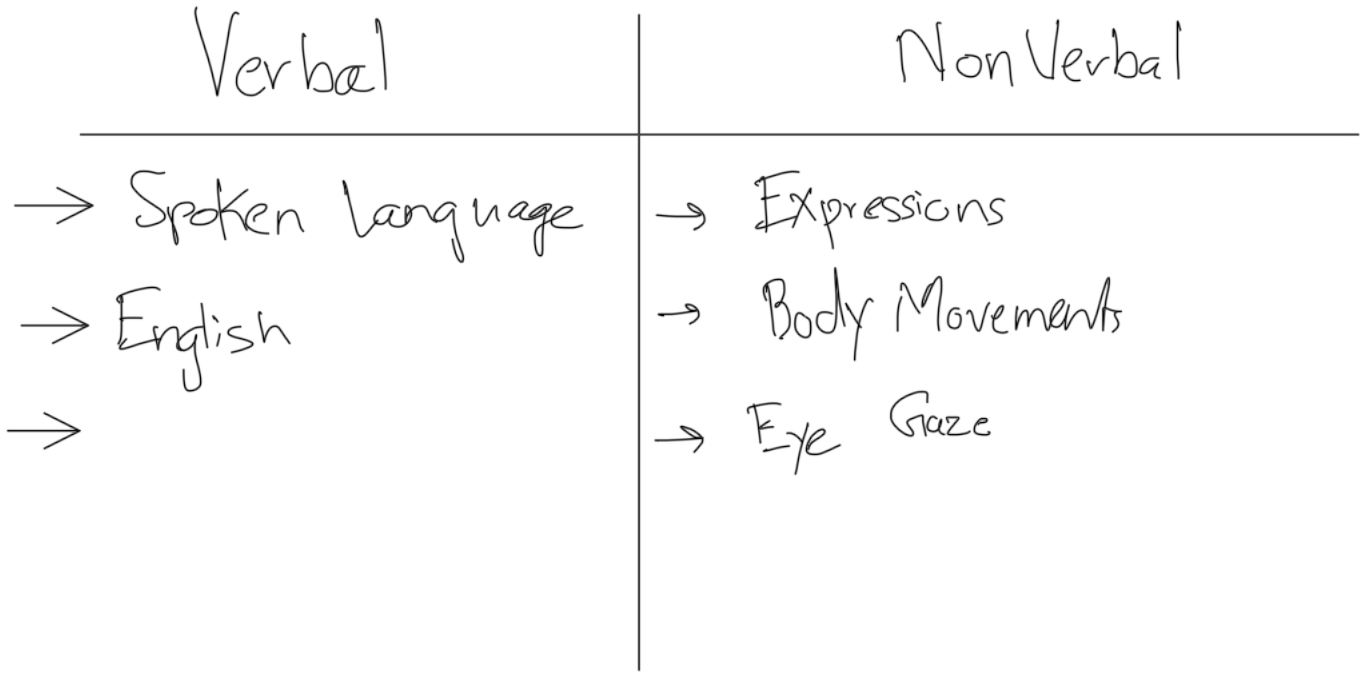
- Raising or dropping of ears
- Blinking of eyes
- Opening and closing of mouth
- Eye movements
- Head, hand, body movements
- ….

- Multiple modalities and all are synchronized to convey emotional responses to objects
  - Liking, happiness
  - Repulsion, fear

Think-Pair-Share

- Watch the videos on slides 4 and 7.
- Write down the verbal and nonverbal components of communication.
- Explain the behaviour generation and behaviour consumption processes in both cases.
- Time: 10 minutes

- Summary: 5 Minutes



Nonverbal Communication Modes

- Kinesics
  - Involves movements and positioning of the robot's body.
  - E.g. **Facial expressions**, hand and arm **gestures**, head and body movements, **eye gaze**
- Proxemics
  - The physical distance between agents influences communication.
    - Hall's four proxemic zones: **intimate < personal < social < public**. (Hall, 1966)
    - Also indicates the relationship between the agents.
- Haptics
  - Physical feedback to interaction partner; e.g. vibrations, touch
- Vocalics
  - Nonverbal parts of speech or other vocalizations; e.g. pitch, loudness, tempo, etc.
- Chronemics
  - How time-related aspects influence communication.

Source: https://open.lib.umn.edu/communication/chapter/4-2-types-of-nonverbal-communication/

Facial Expressions: Recognition and Generation

Facial Expressions

- Reflect our state of mind: emotions, pain, distraction, interest, …
  - (Darwin, 1872; Kunz and Lautenbacher, 2014; Gjoreski et al. 2020; Yeasin et al., 2006)
- Formed by facial muscle movements:
  - **Facial Action Units (AUs)** defined in the Facial Action Coding System (Ekman & Friesen, 1978)

Facial Action Coding System (FACS)

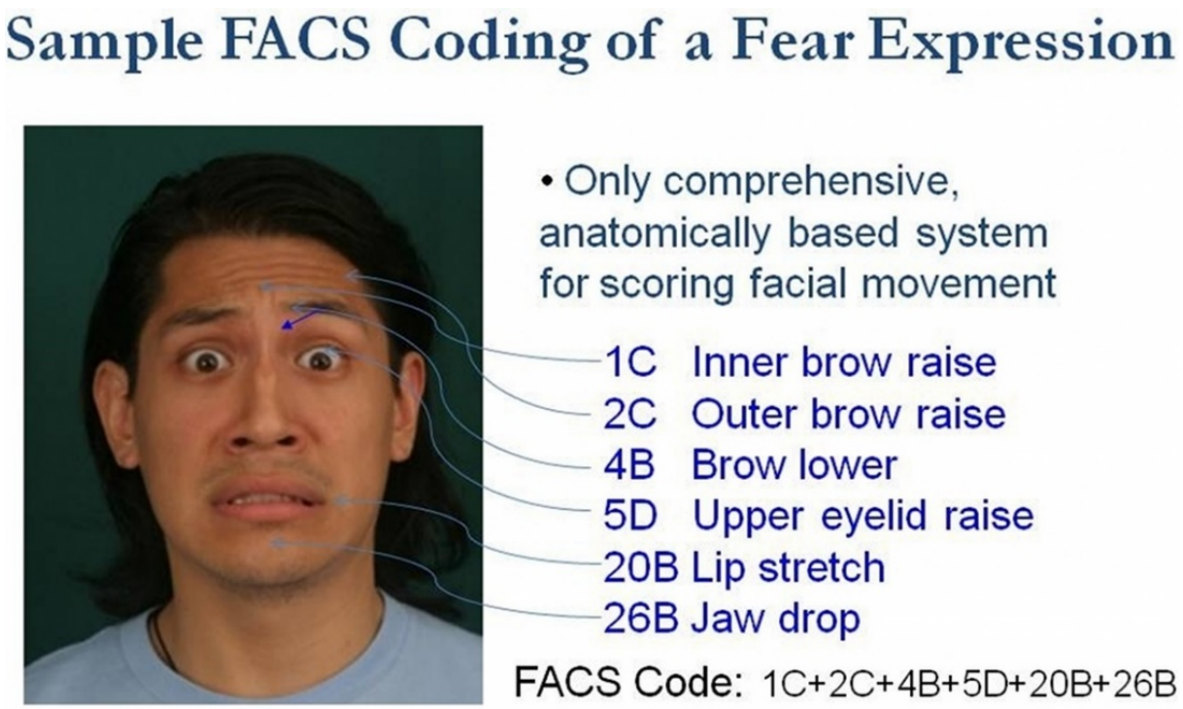
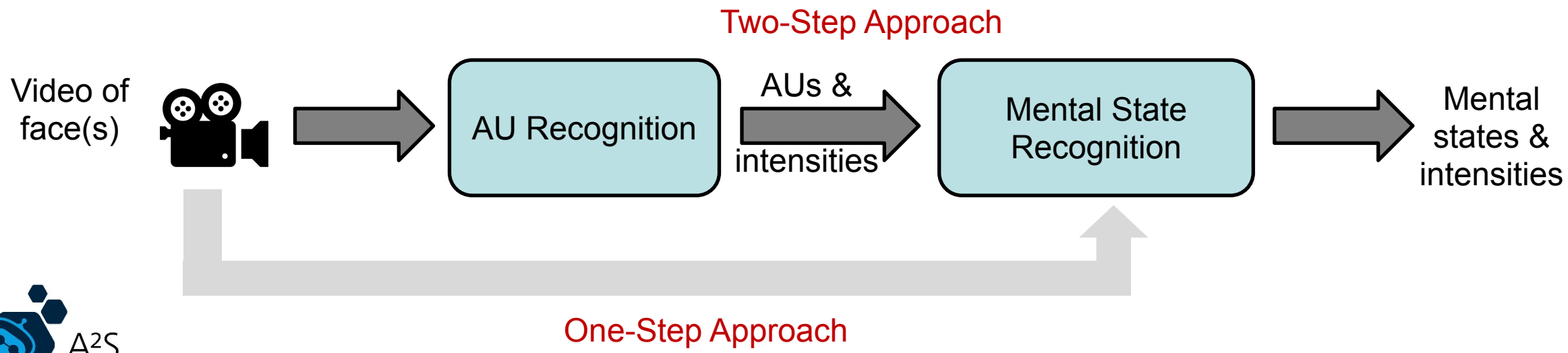


Image source: David Matsumoto and Paul Ekman (2008), Scholarpedia, 3(5):4237
License: Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License

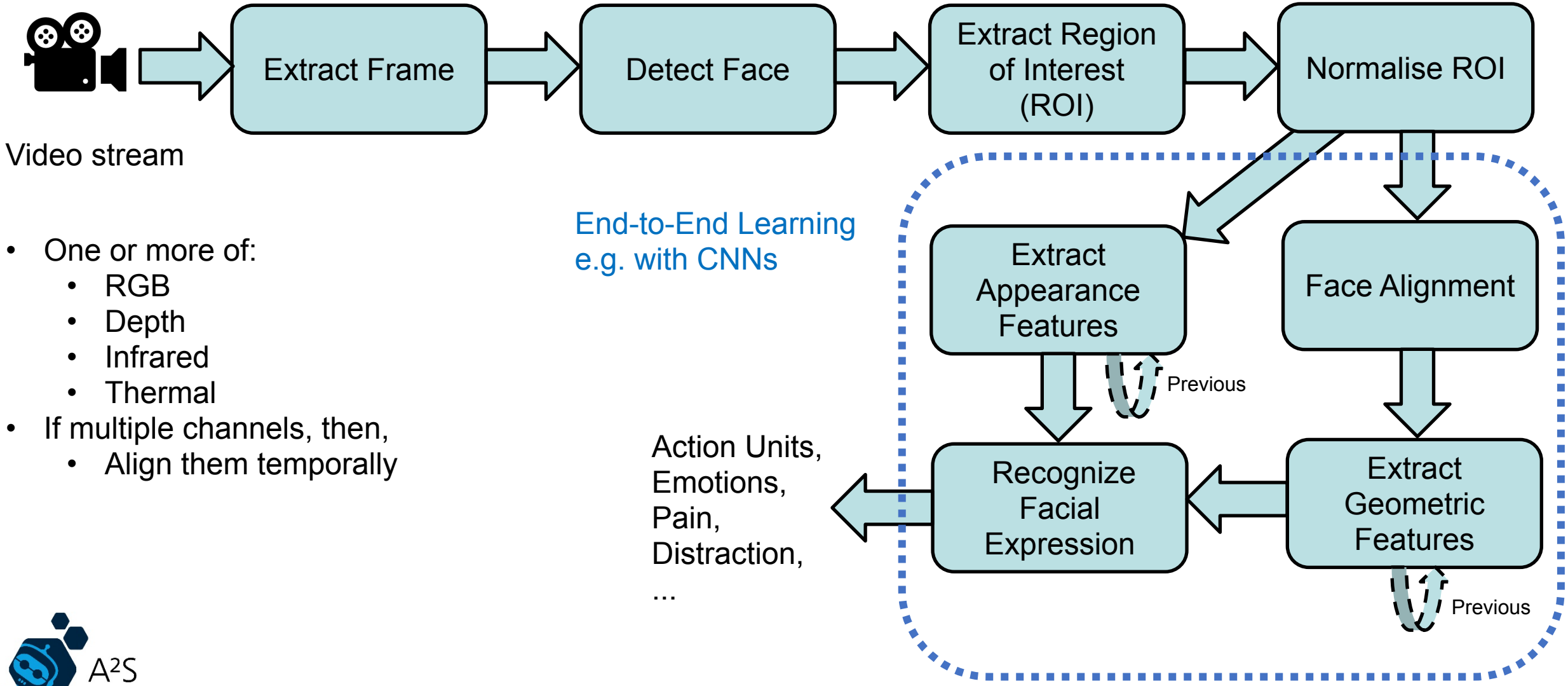
- 44 different AUs
- Produced by individual muscles or groups of muscles
- Can be visually differentiated from one another.
- Can describe any facial expression
- Coding of facial expression
  - AU label (1 to 44)
  - AU intensity (absent or [A, E])

Facial Expression Recognition

- Sign judgment
  - Detect AUs and their intensities
- Message judgment
  - Recognize the mental state that is being conveyed by the facial expression
    - e.g. pain, emotions, distraction, engagement, etc.
  - One-step versus two-step approaches (Hassan et al., 2021)

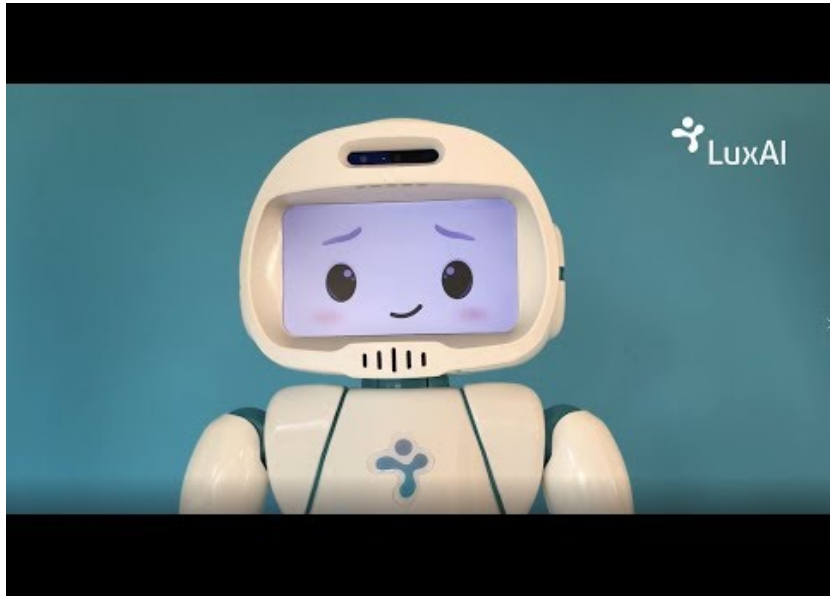


Features for Facial Expression Recognition



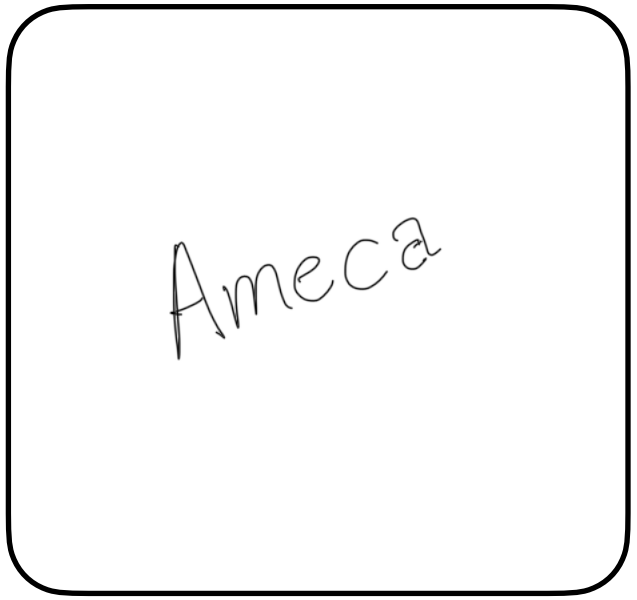
- One or more of:
  - RGB
  - Depth
  - Infrared
  - Thermal
- If multiple channels, then,
  - Align them temporally

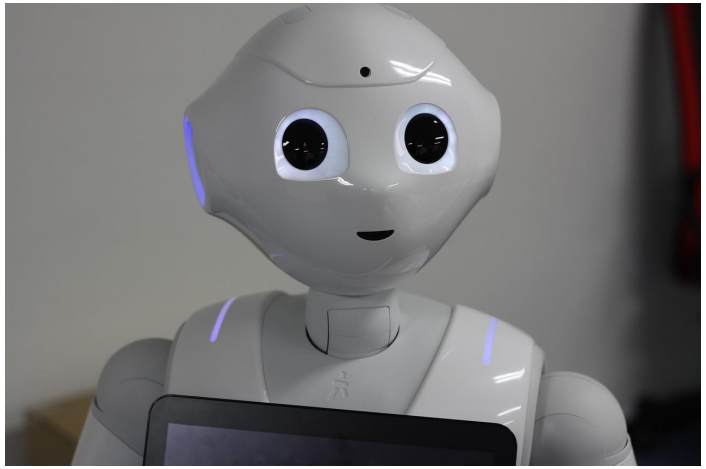
Facial Expression Generation



Virtual, human-like expressions
https://youtu.be/QnTbtbZupWE

Physical, human-like expressions
https://youtu.be/dG7kNhxrOG8





Artificial / Technical modality – LEDs around eyes

Realization of facial expressions (e.g. smile) is embodiment-specific and differs from one robot to another.

- True also for other nonverbal cues such as gaze.

Facial Expression Generation

- Option 1:
  - Create templates for **facial expressions of emotions** based on the embodiment, with the option to vary the intention of the expression.
  - Limited no. of expressions possible.
- Option 2:
  - Create templates for **components of facial expressions** (similar to AUs), and combine several components at various intensities.
  - Greater variety of expressions possible.

Group Discussion

- Form groups of 4

- Skim through the reading assignment from last week.
- Collect 5 examples of the influence of facial expressions of robots on humans.

- Time: 10 minutes

- Summary: 5 Minutes

Eye Gaze: Recognition and Generation

Role of Gaze in Communication

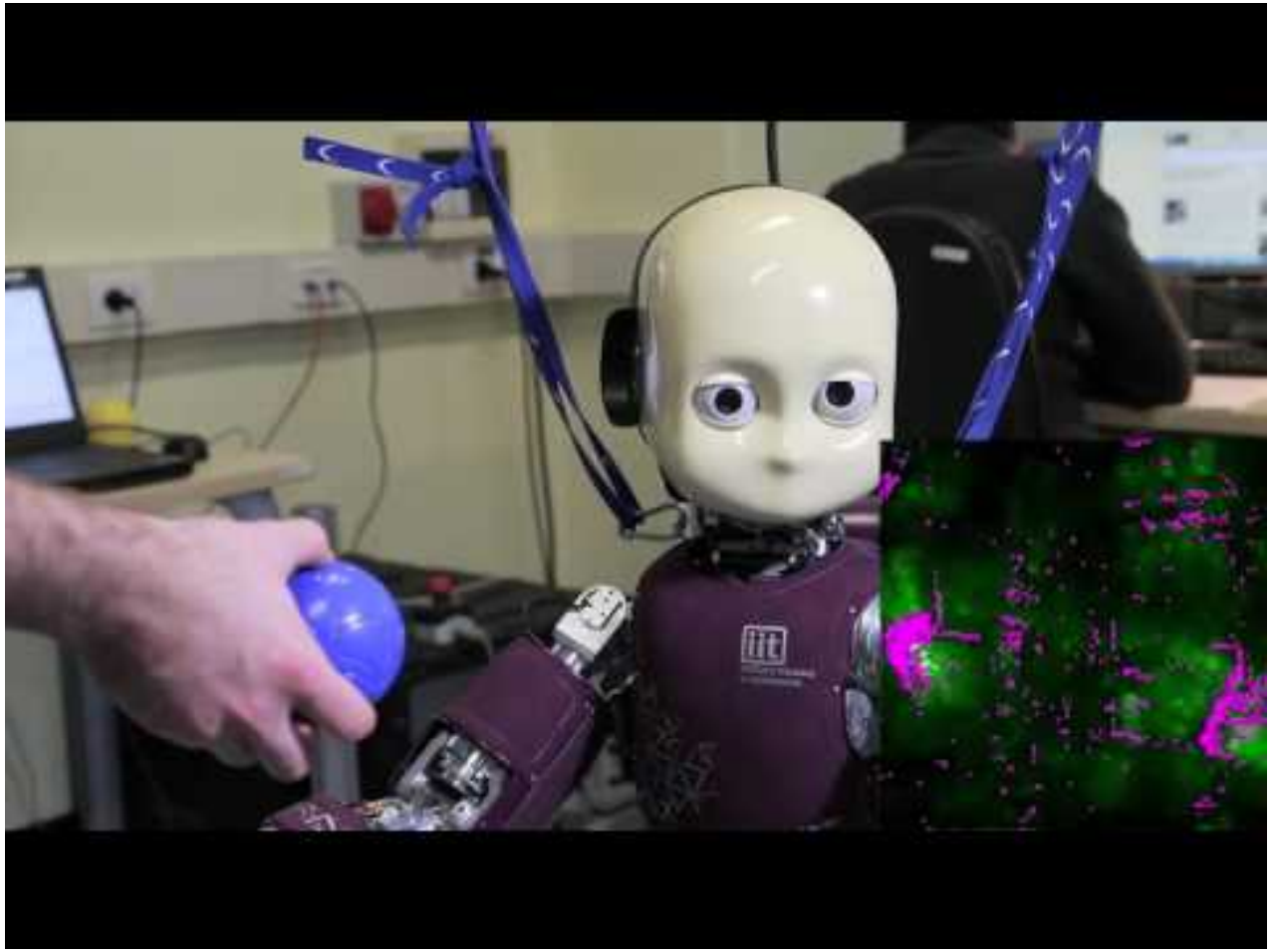
- Eyes play a crucial role in social communication.
- Eye gaze helps to communicate our intent, emotions, engagement, interest, attention, etc. to others. E.g.
  - What is a person looking at?
  - What might the person want to do?
    - E.g. pick up an object; seize conversation floor; cross a street, ...
  - How might the person be feeling?
- Gaze direction, gaze duration, gaze direction changes – all convey different messages during interaction.
  - Look at interaction partner now-and-then while speaking or listening.
    - But, avoid staring for a long time.
  - Change gaze to guide the attention of listener(s) to an object of interest (*joint attention*).

Eye Movements

Movement Type	Description	Functionality	Application in Interaction
Fixations	Eyes stationary	Information acquisition, attention, cognitive processing	Reading, scene perception
Saccades	Rapid eye movements between fixations	Switch attention targets	Visual search
Smooth pursuit	Eyes fixed on a moving object	Follow a moving target	Tracking a moving object, gaze-based steering
Scanpath	A sequence of alternating short fixations and saccades	Scanning	Analyse behavior of agent
Gaze duration	Sum of all fixations (and proportion of time spent) in an area of interest before the eye leaves that area	Convey intent, cognitive processing	Level of engagement or interest, object of interest
Blinks	Rapid closing and opening of eyelids	Indicate behavioral states	Indicate liveliness, stress, fatigue, etc.
Pupil diameter	Size of the pupil	Convey cognitive effort	Cognitive load, fatigue

Based on (Kar and Corcoran, 2017)

Gaze Control in Robots



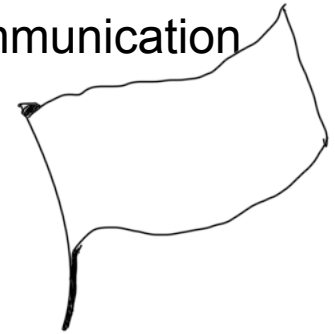
https://youtu.be/n6qTkw5U7YI

Which parts (i.e. joints) of iCub were used to indicate its gaze?

- Eyeballs
- Head / neck
- Body

- Gaze shift requires trajectory planning involving several joints.

- **Central question:** How to plan the trajectory such that it appears naturalistic and suits the interaction and communication context?



Group Discussion

- Form groups of 4

- Skim through the reading assignment from last week.
- Collect 5 examples of the influence of robots' gaze on humans.

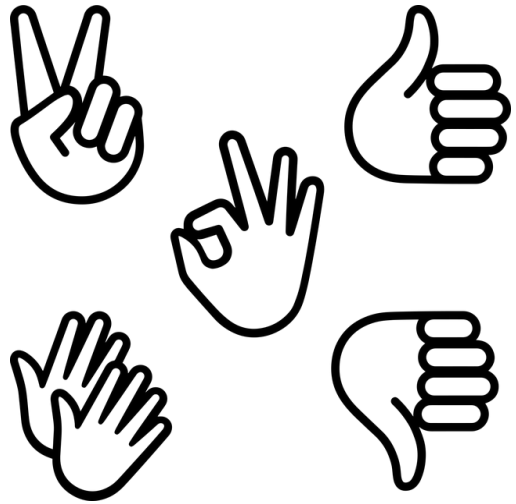
- Time: 10 minutes

- Summary: 5 Minutes

Hand Gestures: Recognition and Generation

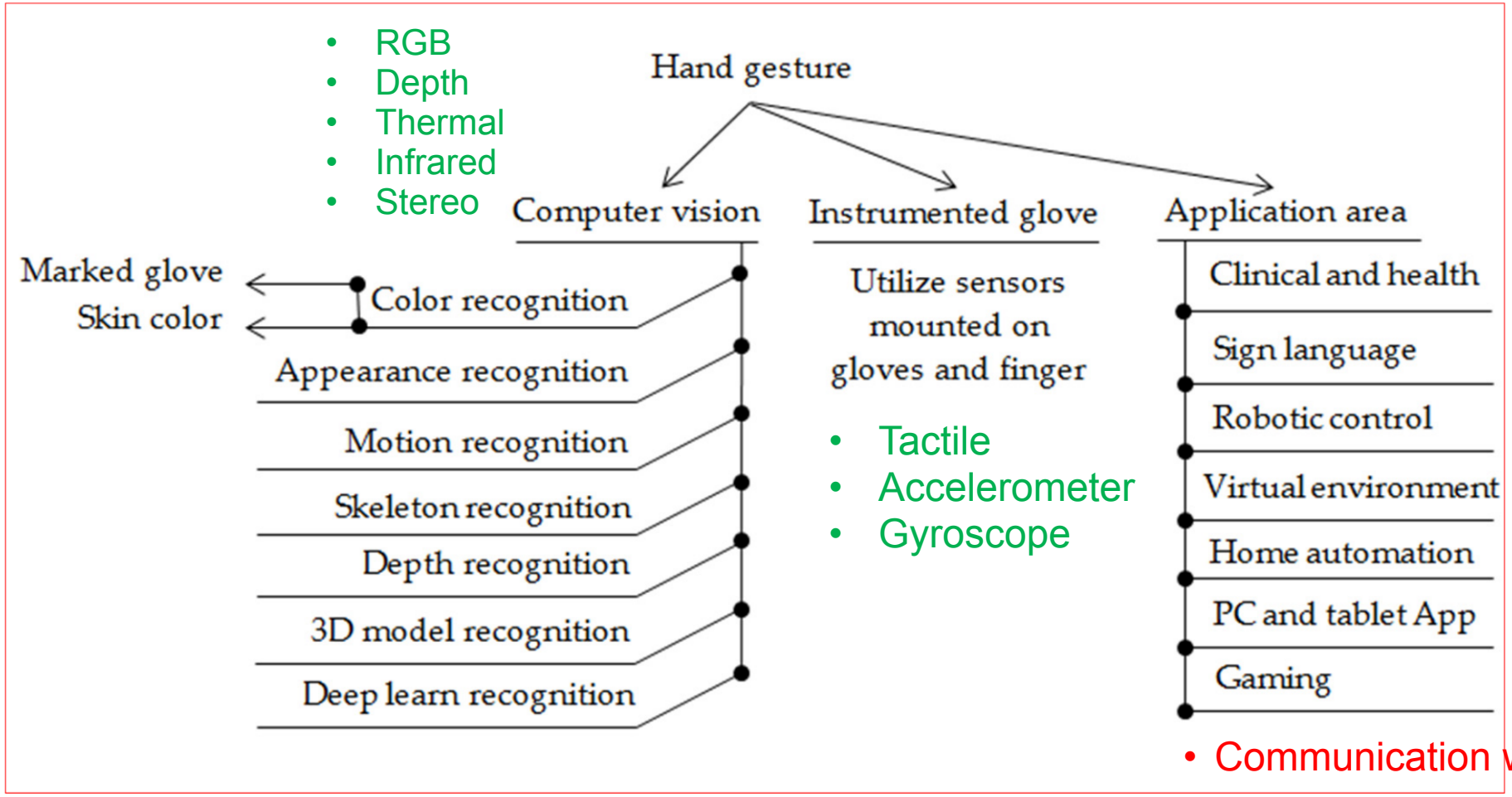
Gestures: Types

- Adaptors
  - Gestures that indicate mental states like anxiety, uneasiness, arousal, boredom, etc.
    - Fidgeting with fingers, clicking pens, shaking of legs, …
    - To release (excessive) energy.
- Emblems
  - Have a specific meaning that has been agreed-to by specific communities.
    - Thumbs-up, victory sign, claps, etc.
  - Culture-specific: Same sign can have different meanings in different cultures or societies --> Use carefully!
    - E.g. the V-sign in Britain
- Illustrators
  - Linked to speech and generated more subconsciously.



Source: https://open.lib.umn.edu/communication/chapter/4-2-types-of-nonverbal-communication/

Hand Gesture Recognition



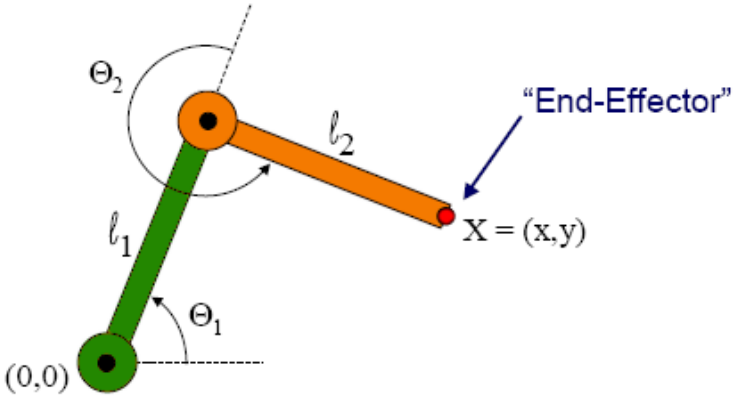
Source: Figure 2 in (Oudah et al., 2020) extended with info from Figure 1.

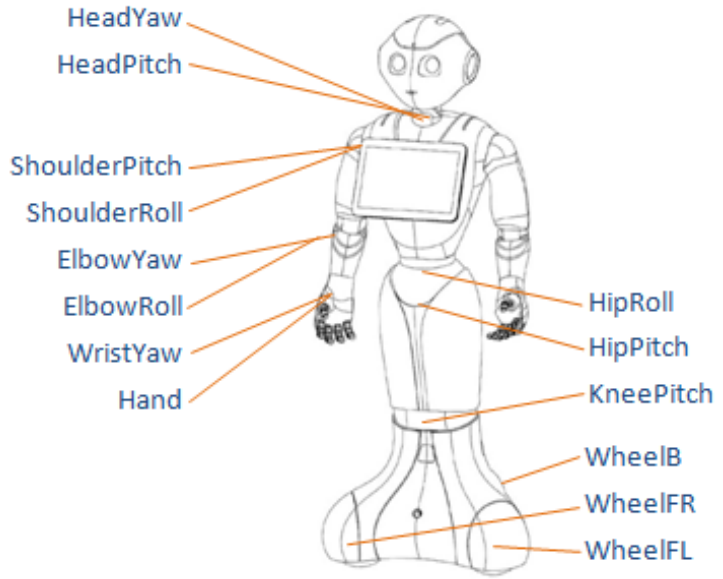
Gestures Accompanying Speech

- “... Speakers gesture more when they talk about spatial topics than when they talk about abstract or verbal ones” (Alibali, 2005, p. 313).
- Bielefeld Speech and Gesture Alignment (SaGA) (Lücking et al., 2010)
  - Spatial communication tasks: give direction and describe a scene
  - Includes following gestures:
    - Indexing (pointing), placing, shaping, drawing, posturing, sizing, counting, hedging.
  - Morphology and movement:
    - Handshape: similar to American Sign Language
    - Palm (back of palm) orientation: Up, down, left, right, forward, backward
    - Wrist position and distance: Relative to the body of the gesturing person
    - Movement direction: Combinations and sequences of six cardinal directions
    - Movement trajectory: linear, curved

Gesture Generation in Robots

- Trajectories in joint space:
  - A sequence of waypoints describing (position, velocity and acceleration) for each (relevant) joint.
  - In joint space, positions are angles, and velocities and accelerations are time derivates of it.
- Could also be specified in Cartesian space:
  - Use inverse kinematics to compute trajectory in joint space.
- Pre-defined trajectories are tuned through trial and error & stored in a gesture library.
  - Adapted on-the-fly e.g. tempo.



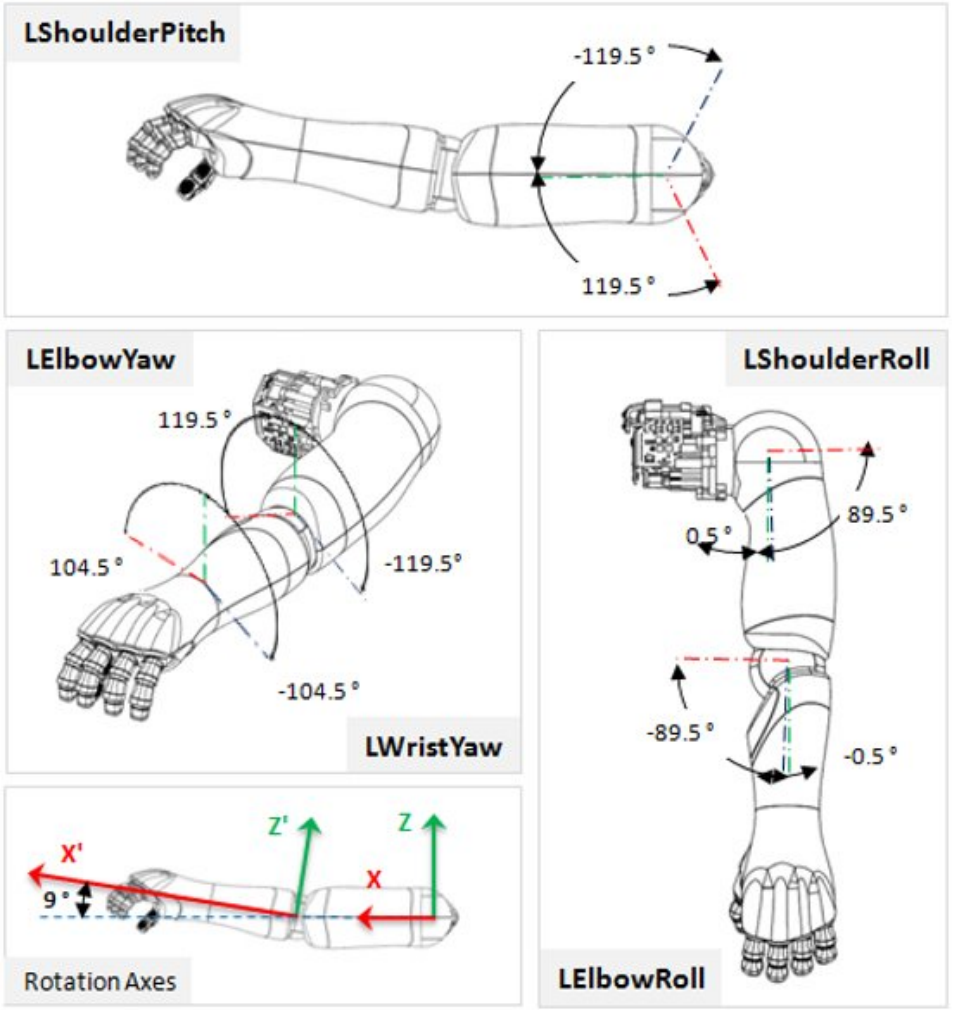


http://doc.aldebaran.com/2-4/family/pepper_technical/motors_pep.html

Gesture Generation in Robots

- Nowadays, machine learning methods are used to learn trajectories from human demonstrations.
  - Annotated videos
  - Human movement data

- But, direct 1-1 transfer to robots unlikely due to structural and kinematic differences.



http://doc.aldebaran.com/2-4/_images/joint_left_arm.png

Group Discussion

- Form groups of 4

- Skim through the reading assignment from last week.
- Collect 5 examples of the influence of robots' gestures on humans.

- Time: 10 minutes

- Summary: 5 Minutes

Summary

Nonverbal Communication

- The nonverbal modalities that are available for communication depend on the physical embodiment.
  - For humans: Depends on the anatomical and muscular functioning.
  - For robots: What actuators and sensors are the robot built with?
- Expected nonverbal communication abilities depend on appearance and framing.
  - This affects how delays and errors in nonverbal behaviour influence human perception and acceptance of robots.
- How does it influence humans when robots show these nonverbal behaviors? (Saunderson and Nejat, 2019).
  1. Shift cognitive framing
  2. Elicit emotional responses
  3. Trigger specific behavioral responses
  4. Improve task performance

Recap Nonverbal Communication

- Nonverbal communication modes: Kinesics, proxemics, haptics, vocalics
- Nonverbal signals:
  - Facial expressions: Signs, messages
  - Eye movements: Fixations, gaze, saccades, smooth pursuits, scanpaths, blinks, pupil size
  - Gestures: Adaptors, emblems, illustrators
- Recognition of nonverbal signals
  - Mainly based on machine learning and mathematical models applied to sensor data
- Generation of nonverbal signals
  - Predefined trajectories or animations
  - Recently, machine learning models to predict robot poses and trajectories
  - Embodiment-specific adaptations essential

Multimodal Nonverbal Behaviour

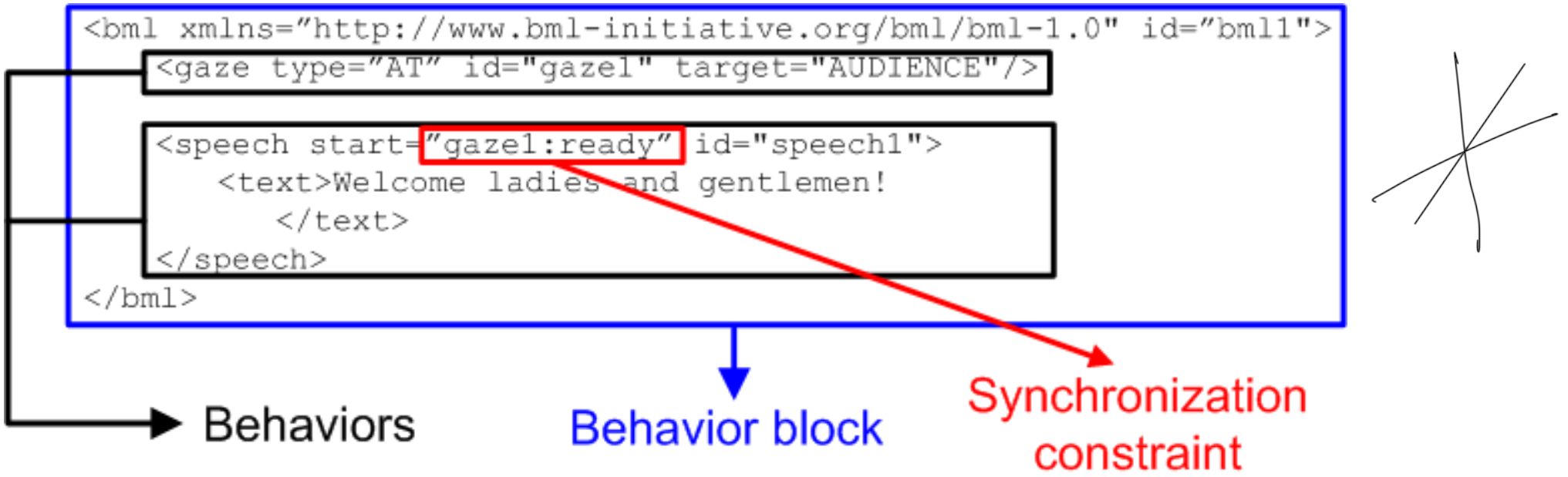
Temporal Alignment

- Signals coming through different channels should be aligned in order to interpret its meaning correctly.
  - Speech
  - Hand gestures
- Alignment
  - Start
  - End
  - Duration
  - Rhythm

- Multimodal behaviour recognition
  - Communicative role of each modality
  - Use multiple information fusion strategies, different timescales
- Fluent generation of multimodal behaviour
  - What are needed?
    - Feedback on execution status
    - Predictions about execution
    - Behaviour planning
    - Status monitoring and scheduling
    - Incremental update
    - Failure handling

Behavior Markup Language (BML) (Vilhjálmsson et al., 2007)

- An XML description language for specifying multimodal (verbal and nonverbal) behaviours that should be expressed by an embodied conversational agent (a virtual agent or a physical robot).
- A BML Realizer understands BML requests and realises these behaviours on the agent.



Example of a BML Request [Source: Figure 1 from https://projects.cs.ru.is/projects/behavior-markup-language/wiki]

Conclusion

- In today's lecture, you learnt to:
  1. List the communication modalities in humans and robots.
     - Verbal and nonverbal modalities
  2. Identify the key elements and concepts of nonverbal communication.
     - Kinesics, proxemics, haptics, vocalics
  3. Elucidate the technical aspects of realising multimodal communication between humans and robots.
     - Recognition and generation of nonverbal behaviours
       - Facial expressions, eye gaze, hand gestures
     - Synchronising verbal and nonverbal behaviours using BML

References and Recommended Reading

1. Saunderson, S., Nejat, G. **How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction**. *Int J of Soc Robotics* **11,** 575–608 (2019). https://doi.org/10.1007/s12369-019-00523-0
2. Doncieux, S., Chatila, R., Straube, S. *et al.* Human-centered AI and robotics. *AI Perspect* **4,** 1 (2022). https://doi.org/10.1186/s42467-021-00014-x

[References 1 & 2 would be important for the exam.]

Next Session

- Lecture on “Verbal communication in humans and robots”
  - Thursday, 02.05.2024 at 9 am
  - In: B060, Grantham Allee 20, 53757 Sankt Augustin

- Reading assignment on model of communication between humans: See LEA

References (1/2)

- M.W. **Alibali**. 2005. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. Spatial Cognition and Computation, 5:307–331.
- Pablo **Cañas**, David Griol, Zoraida Callejas, "Towards versatile conversations with data-driven dialog management and its integration in commercial platforms", Journal of Computational Science, Volume 55, 2021.
- C. **Darwin**. The expression of the emotions in man and animals. Original work published 1872. New York: Oxford University Press, 1998.
- **Doncieux**, S., Chatila, R., Straube, S. et al. Human-centered AI and robotics. AI Perspect 4, 1 (2022). https://doi.org/10.1186/s42467-021-00014-x
- **Ekman**, P., & Friesen, W. V. (1978). Manual for the Facial Action Code. Palo Alto, CA: Consulting Psychologist Press.
- M. **Gjoreski**, M. Ž. Gams, M. Luštrek, P. Genc, J. Garbas and T. Hassan, "Machine Learning and End-to-End Deep Learning for Monitoring Driver Distractions From Physiological and Visual Signals," in IEEE Access, vol. 8, pp. 70590–70603, 2020.
- E. T. **Hall**, The Hidden Dimension, 1st ed. Garden City, N.Y: Doubleday, 1966.
- T. **Hassan** et al., "Automatic Detection of Pain from Facial Expressions: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 6, pp. 1815-1831, 1 June 2021, doi: 10.1109/TPAMI.2019.2958341.
- A. **Kar** and P. Corcoran, "A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms," in IEEE Access, vol. 5, pp. 16495-16519, 2017, doi: 10.1109/ACCESS.2017.2735633.
- M. **Kunz** and S. Lautenbacher. “The faces of pain: a cluster analysis of individual differences in facial activity patterns of pain.” In: European Journal of Pain 18.6 (2014), pp. 813–823.
- **Lücking** A, Bergmann K, Hahn F, Kopp S, Rieser H. The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In: Kipp M, Martin J-P, Paggio P, Heylen D, eds. LREC 2010 Workshop: Multimodal Corpora–Advances in Capturing, Coding and Analyzing Multimodality. 2010: 92-98.
- S. **Narayanan** and P. G. Georgiou, "Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language," in Proceedings of the IEEE, vol. 101, no. 5, pp. 1203-1233, May 2013, doi: 10.1109/JPROC.2012.2236291.

References (2/2)

- **Oudah**, M.; Al-Naji, A.; Chahl, J. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. J. Imaging 2020, 6, 73. https://doi.org/10.3390/jimaging6080073
- **Saunderson**, S., Nejat, G. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. Int J of Soc Robotics 11, 575–608 (2019). https://doi.org/10.1007/s12369-019-00523-0
- **Vilhjálmsson**, H. et al. (2007). The Behavior Markup Language: Recent Developments and Challenges. In: Pelachaud, C., Martin, JC., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds) Intelligent Virtual Agents. IVA 2007. Lecture Notes in Computer Science(), vol 4722. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74997-4_10
- M. **Yeasin**, B. Bullot and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," in IEEE Transactions on Multimedia, vol. 8, no. 3, pp. 500–508, June 2006.