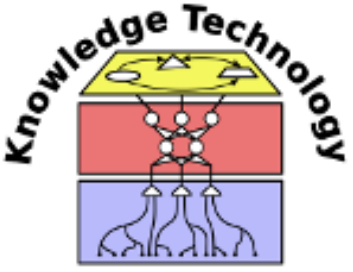
Hey Robot, Why Don’t You Talk To Me?

Hwei Geok Ng, Paul Anton, Marc Brügger, Nikhil Churamani, Erik Fließwasser, Thomas Hummel, Julius Mayer, Waleed Mustafa, Thi Linh Chi Nguyen, Quan Nguyen, Marcus Soll, Sebastian Springenberg, Sascha Griffiths, Stefan Heinrich, Nicolás Navarro-Guerrero, Erik Strahl, Johannes Twiefel, Cornelius Weber and Stefan Wermter

Knowledge Technology
Department of Informatics
Universität Hamburg



www.knowledge-technology.info

Motivation

- Human-centric environment → sociable and interactive
  [Brooks et al., 1999] [Breazeal, 2003]
- Improves user's perception → overall competence [Duffy, 2003]
- Model effective and engaging interactions [Trajkovski and Collins, 2009]
- To present an interaction scenario with the NICO robot holding an engaging conversation with the users
  - Autonomous interaction
  - Personalization → increase likeability [Dautenhahn, 1995]
  - Object learning scenario: Humanoidly Speaking
    [Hinaut et al., 2015] [Twiefel et al., 2016]

Interaction Video (Part 1)

Outline

- About NICO
- Face Detection and Tracking
- Person Identification
- Speech Processing
- Conversation and Modeling

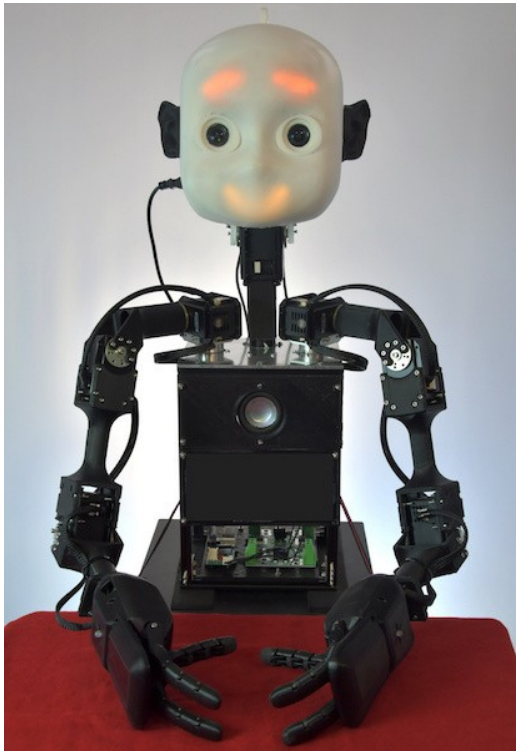


Figure: NICO robot

NICO

- Neuro-Inspired COmpanion Robot
  [Kerzel et al., 2017]
- Built for neuro-cognitive research
- Multi-modal capabilities:
  - Kinetic arms
  - Stereo vision
  - Speech
  - LED facial expressions
  - External microphone

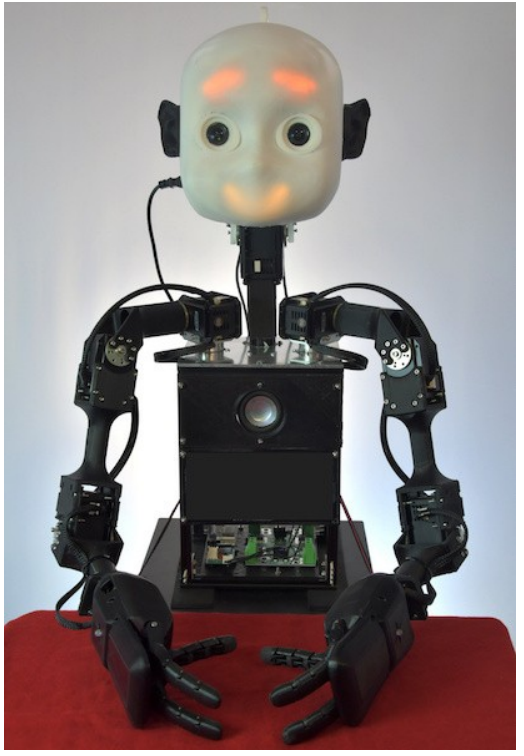
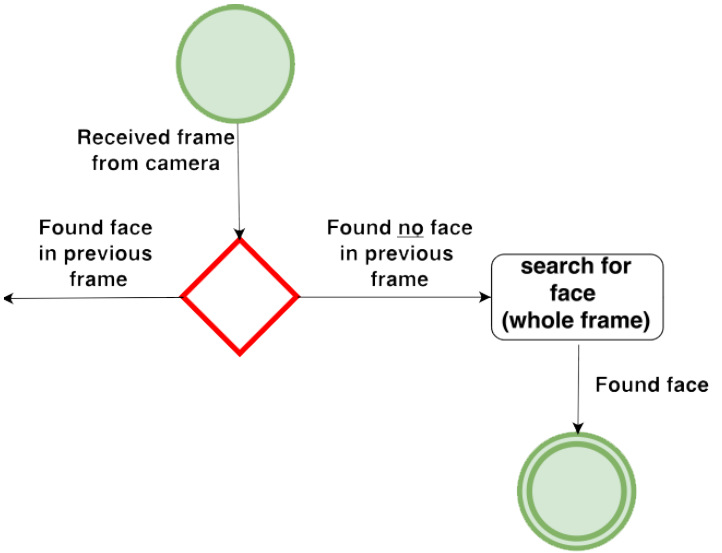


Figure: NICO robot

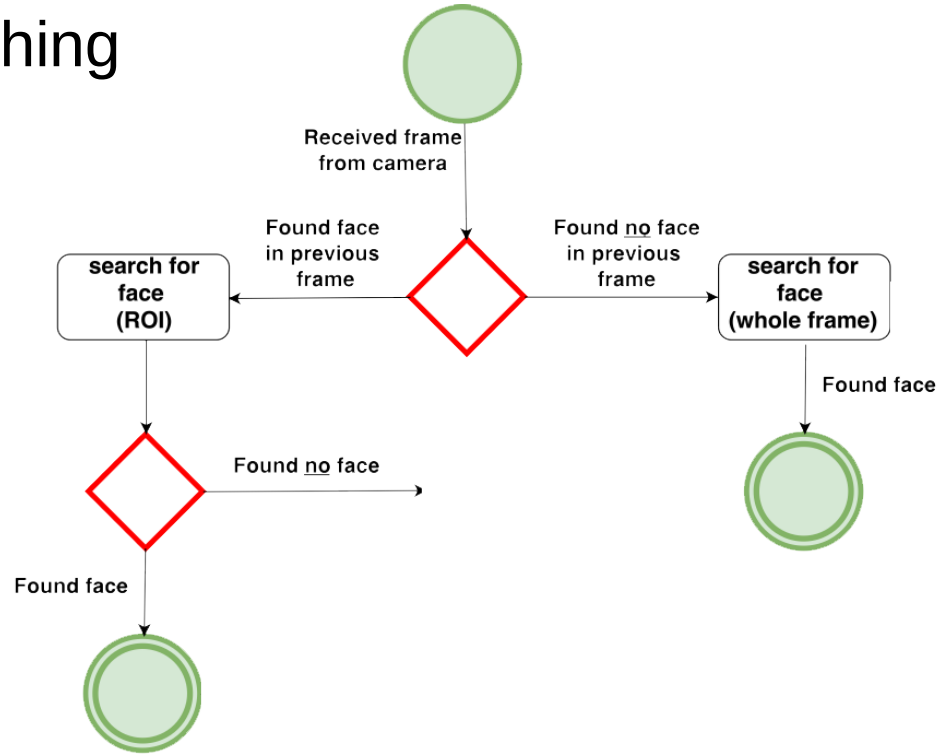
Face Detection and Tracking

- Haar-like cascades based face detection [Viola and Jones, 2001]
- Extended with template matching



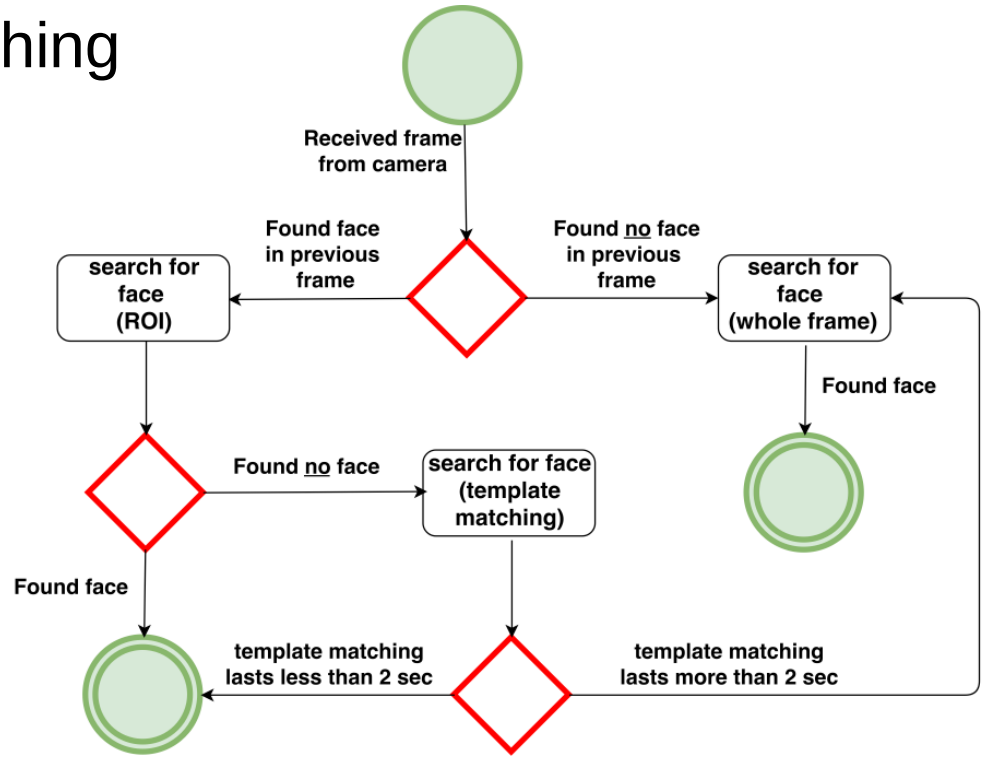
Face Detection and Tracking

- Haar-like cascades based face detection [Viola and Jones, 2001]
- Extended with template matching



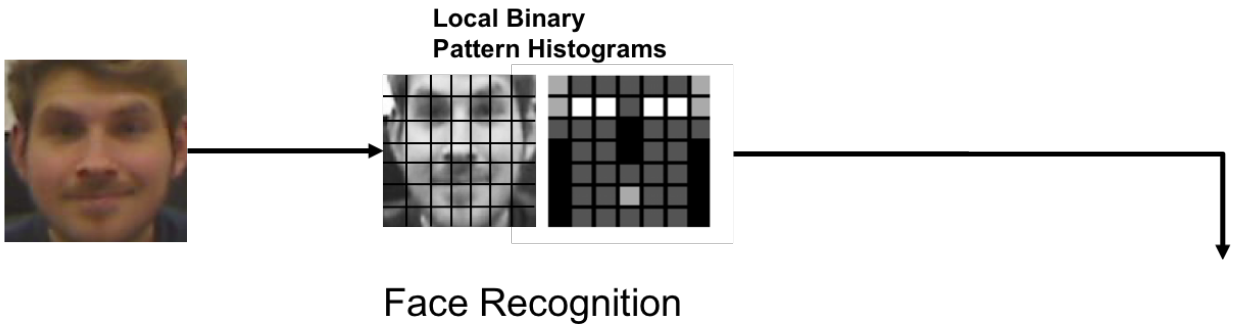
Face Detection and Tracking

- Haar-like cascades based face detection [Viola and Jones, 2001]
- Extended with template matching



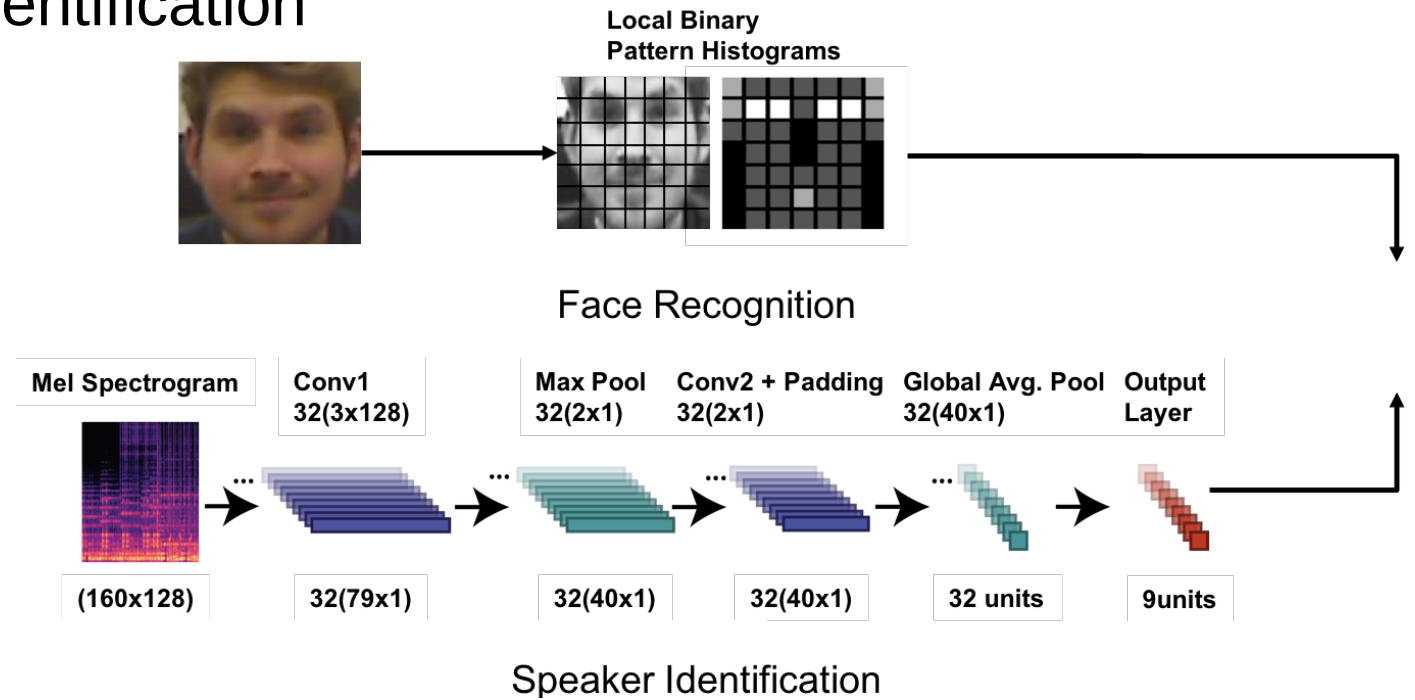
Person Identification

- Face recognition



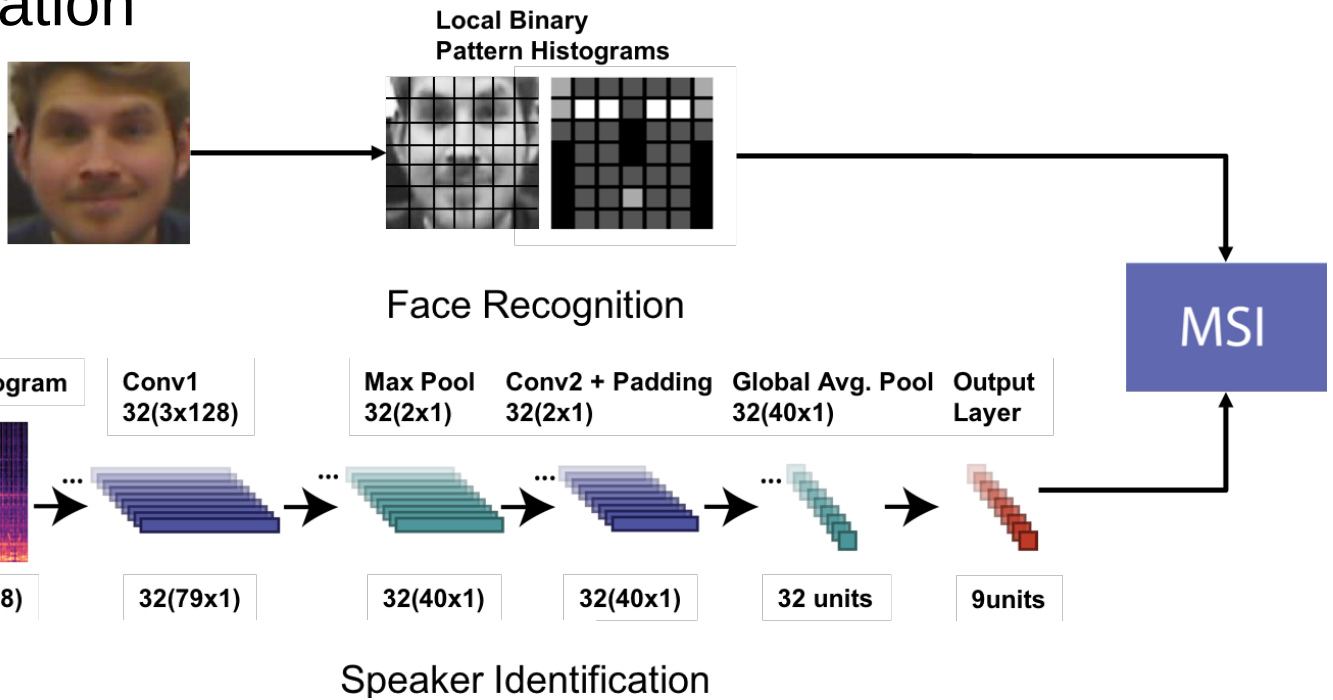
Person Identification

- Face recognition
- Speaker identification



Person Identification

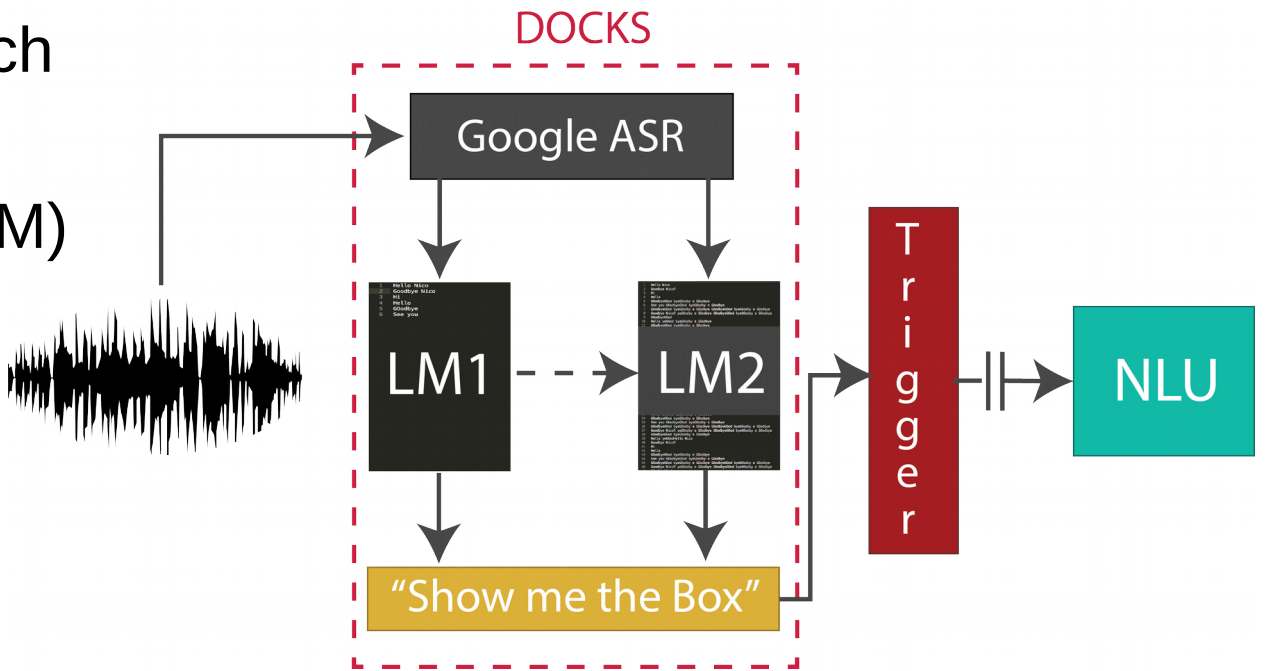
- Face recognition
- Speaker identification
- Multi-sensory integration



Speech Processing

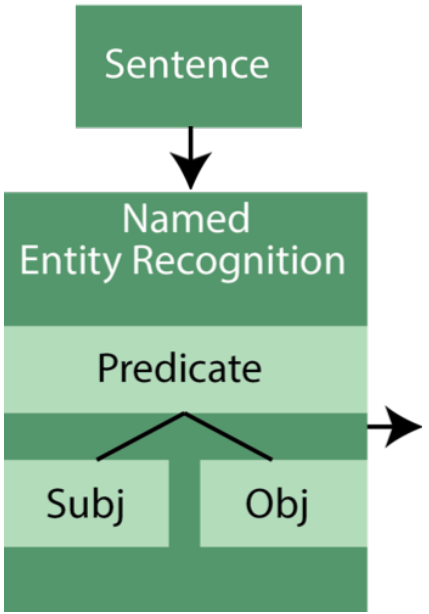
- Speech recognition using the DOCKS framework (DOmain- and Cloud-based Knowledge for Speech recognition) with Language Models (LM)

  [Twiefel et al., 2014]



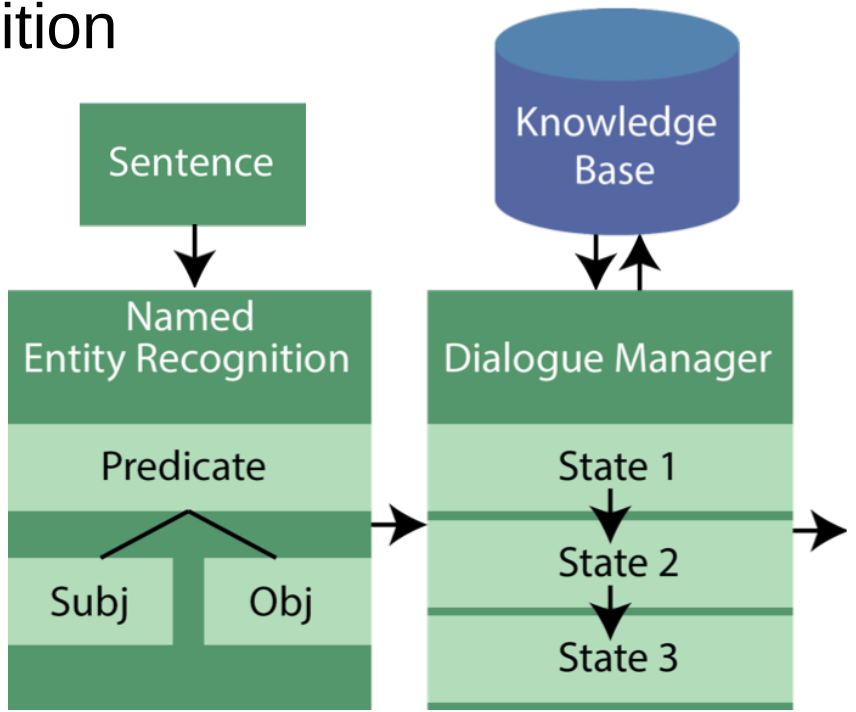
Conversation Modeling

- Natural Language Understanding:
  Named Entity Recognition



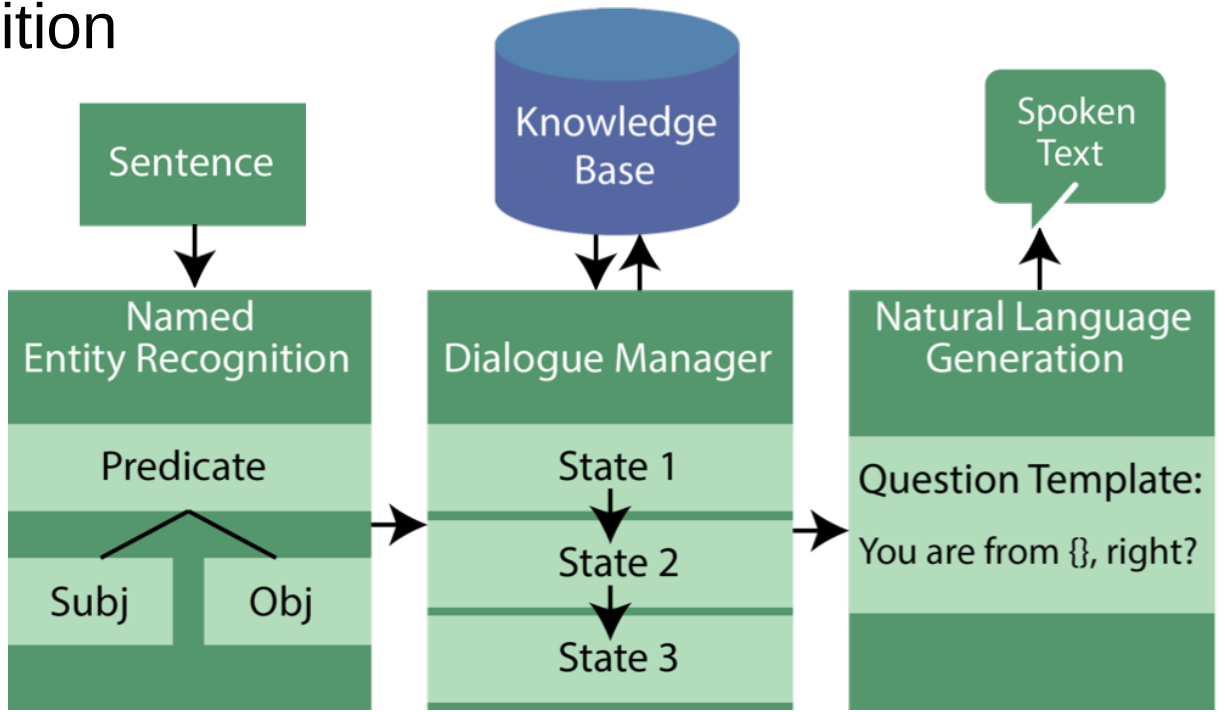
Conversation Modeling

- Natural Language Understanding: Named Entity Recognition
- Dialogue Manager
- Knowledge Base

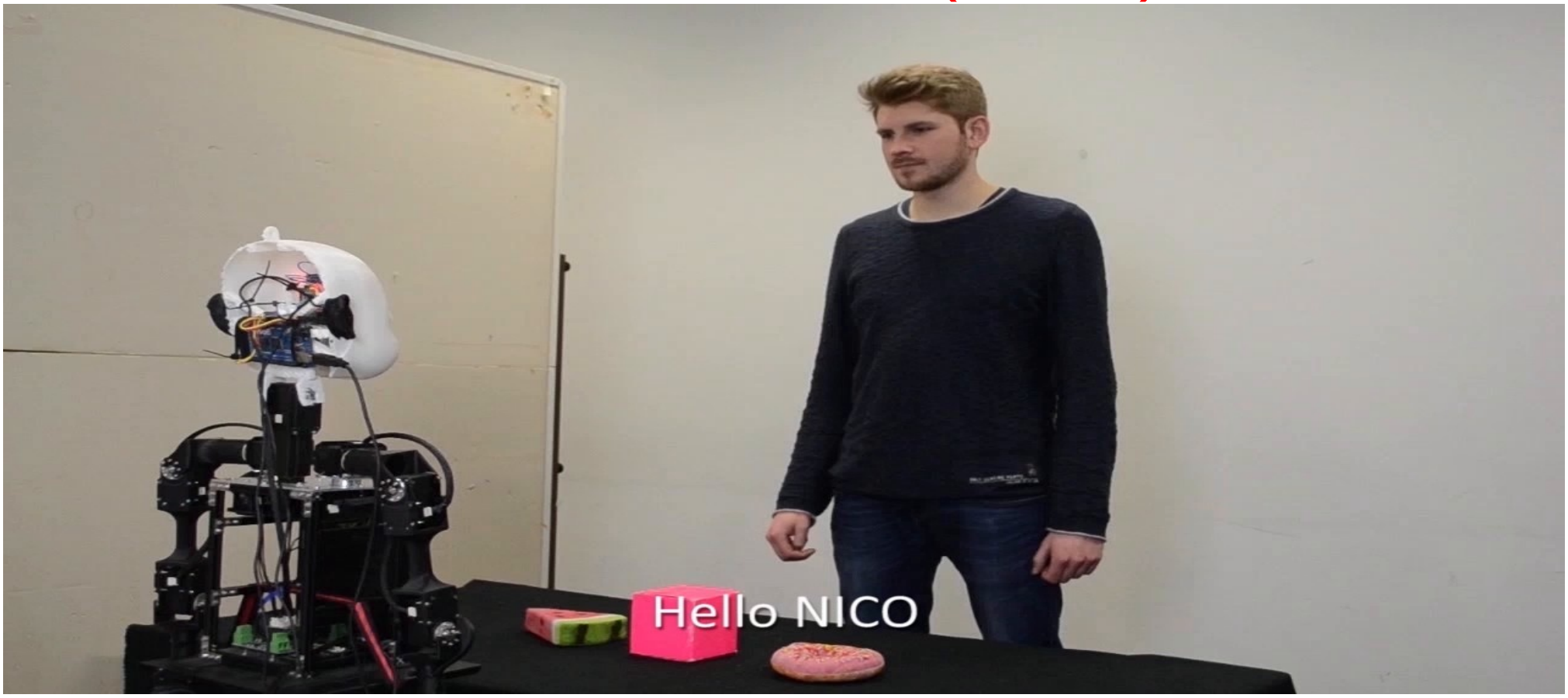


Conversation Modeling

- Natural Language Understanding: Named Entity Recognition
- Dialogue Manager
- Knowledge Base
- Natural Language Generation and text-to-speech synthesis



Interaction Video (Part 2)



Summary

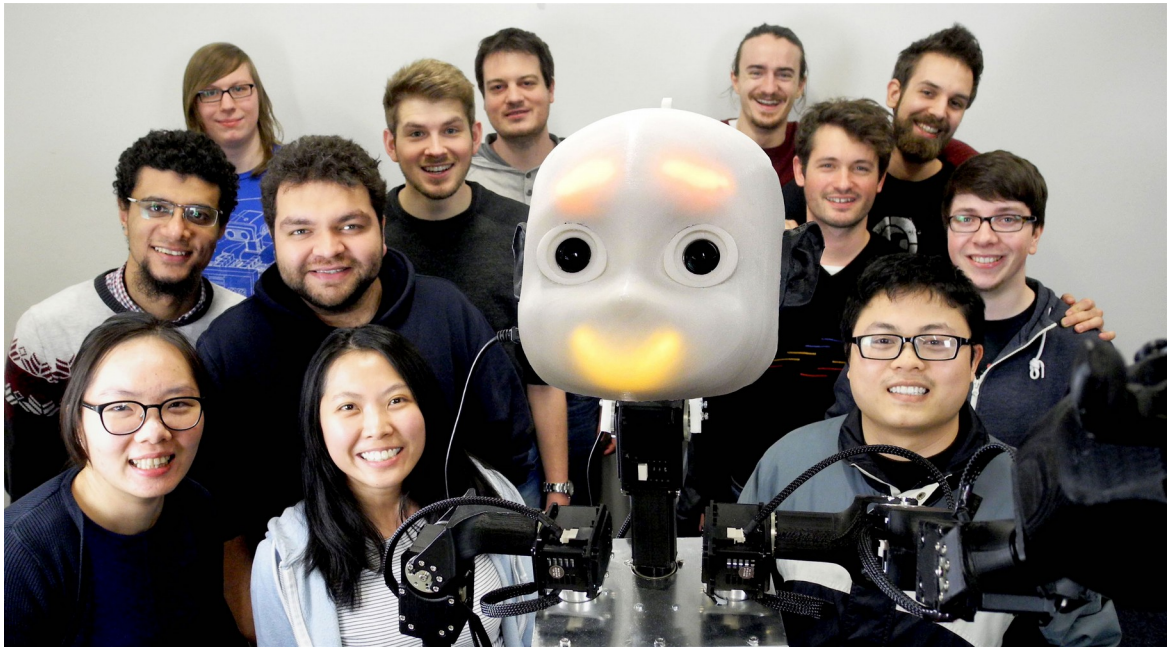
- NICO as a personalised interactive social robot
  - Performs face recognition and tracking
  - Recognizes the user through vision and speech
  - Understands user's natural language
  - Generates replies through conversation modeling
- Autonomous and personalised interaction
- Further experiments presented in:

  "The Impact of Personalisation on Human-Robot Interaction in Learning Scenarios" [Churamani et al., HAI, Bielefeld, Germany, October 2017]

Thank You

www.knowledge-technology.info

phri1617@informatik.uni-hamburg.de



References

C. Breazeal, “Emotion and sociable humanoid robots,” International Journal of Human-Computer Studies, vol. 59, no. 1, pp. 119–155, 2003.

R. A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. M. Williamson, “The Cog project: Building a humanoid robot,” in Computation for Metaphors, Analogy, and Agents, C. L. Nehaniv, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 52–87.

N. Churamani, P. Anton, M. Brügger, E. Fließwasser, T. Hummel, J. Mayer, W. Mustafa, H. G. Ng, T. L. C. Nguyen, Q. Nguyen, M. Soll, S. Springenberg, S. Griffiths, S. Heinrich, N. Navarro-Guerrero, E. Strahl, J. Twiefel, C. Weber, S. Wermter, “The Impact of Personalisation on Human-Robot Interaction in Learning Scenarios”, Proceedings of the Fifth International Conference on Human Agent Interaction, 2017, to appear.

K. Dautenhahn, “Getting to know each other – Artificial social intelligence for autonomous robots”, Robotics and Autonomous Systems 16, 2 (1995), pp. 333 – 356, Moving the Frontiers between Robotics and Biology.

B. R. Duffy, “Anthropomorphism and the social robot,” Robotics and Autonomous Systems, vol. 42, no. 3, pp. 177 – 190, 2003, Socially Interactive Robots.

X. Hinaut, J. Twiefel, M. B. Soares, P. Barros, L. Mici, and S. Wermter, “Humanoidly speaking–learning about the world and language with a humanoid friendly robot.” International Joint Conference on Artificial Intelligence Video competition, 2015.

References

M. Kerzel, E. Strahl, S. Magg, N. Navarro-Guerrero, S. Heinrich, and S. Wermter, “NICO – Neuro-Inspired COmpanion: A Developmental Humanoid Robot Platform for Multimodal Interaction,” in IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Lisbon, Portugal: IEEE, 2017, to appear.

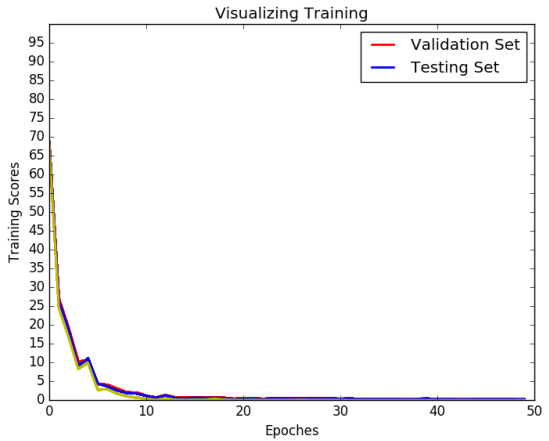
G. Trajkovski and S. G. Collins, “Handbook of Research on Agent-Based Societies: Social and Cultural Interactions”, 1st ed. IGI Global, 2009.

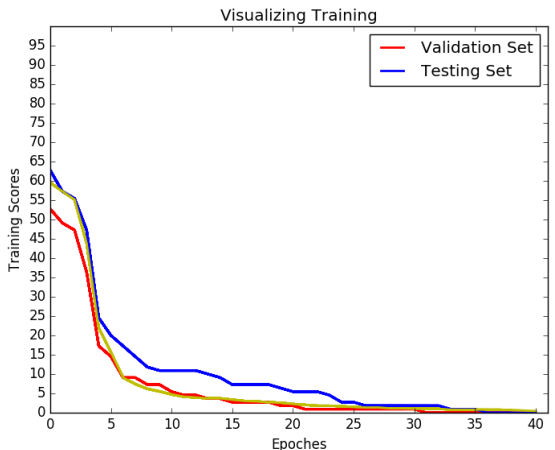
J. Twiefel, T. Baumann, S. Heinrich, and S. Wermter, “Improving Domain-Independent Cloud-Based Speech Recognition with Domain Dependent Phonetic Post-Processing,” in AAAI Conference on Artificial Intelligence, vol. Twenty-Eighth. Quebec City, Quebec, Canada: AAAI Press, 2014, pp. 1529–1535.

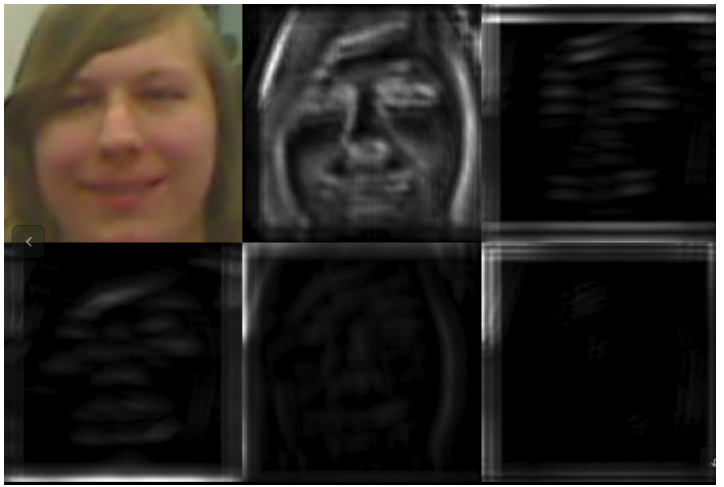
J. Twiefel, X. Hinaut, M. Borghetti, E. Strahl, S. Wermter, “Using natural Language Feedback in a Neuro-inspired Integrated Multimodal Robotic Architecture”, Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 52–57, Aug 2016.

P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1. Kauai, Hawaii, USA: IEEE, 2001, pp. 511–518.

Face Recognition and Tracking using CNN



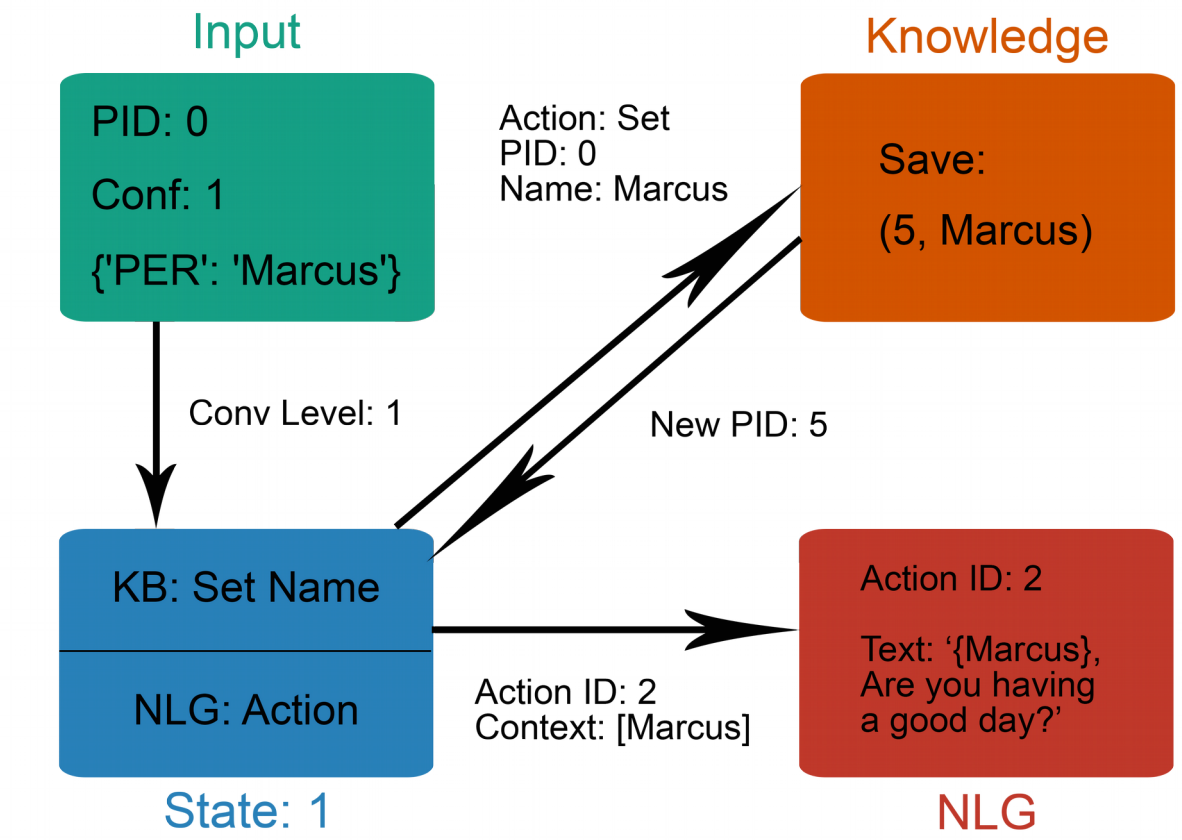




Speech Recognition

Method	Feature	Training Accuracy	Testing Accuracy
CNN	MFCCs	98%	76%
CNN	Mel-Spectogram	96%	94%
GMM using Fischer vectors and LDA	MFCCs	-	96%

Conversation Modeling



Conversation Modeling

User‘s Utterance	Input Received	Executed State	Actions Performed	Transition	Output Utterance
‘Hello NICO.’	ConvLvl: 0 PID: 0 Pred:{‘hello’: ‘’}	Hello	- If PID is unknown, set AID = 2	Set ConvLvl = 1	AID: 2, ‘Hello, what‘s your name?’
‘My name is Erik.’	ConvLvl: 1 PID: 0 Pred:{‘PER’:‘Erik’}	Name	- Set name to KB - Get PID: 1 - Set to known person - Set AID = 5 - NLG args: ‘Erik’	Set ConvLvl = 2	AID: 5, ‘That‘s a good name, {Erik}. I will remember it. Are you having a good day?’
‘Yes, I am.’	PID: 1 Pred: {‘INTERJECTION’:‘YES’}	Day	- If yes, set AID =7 - Save answer to KB	Set ConvLvl = 3	AID: 7, ‘I am so happy to hear that. Is this your first time participating in an experiment?’
…..					
‘Goodbye NICO.’	ConvLvl: 42 PID: 1 Pred:{‘goodbye’: ‘’}	Demo_1	- If goodbye, set AID = 1 - NLG args: ‘Erik’	Set ConvLvl = 0	AID:1, ‘Ok, {Erik}, I really enjoyed our conversation today, but let‘s take a break and I will see you soon.’

Preliminary Results

- Experiments:
  - Perceived intelligence
  - Social acceptance
  - Likeability
- Results:
  - More intelligent
  - More likeable
  - More engaged in conversation