

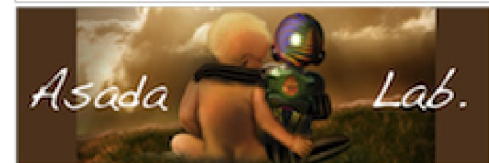
# Visual Attention by Audiovisual Signal-Level Synchrony

**Matthias Rolf, Minoru Asada**

Specially Appointed Researcher

[matthias@ams.eng.osaka-u.ac.jp](mailto:matthias@ams.eng.osaka-u.ac.jp)

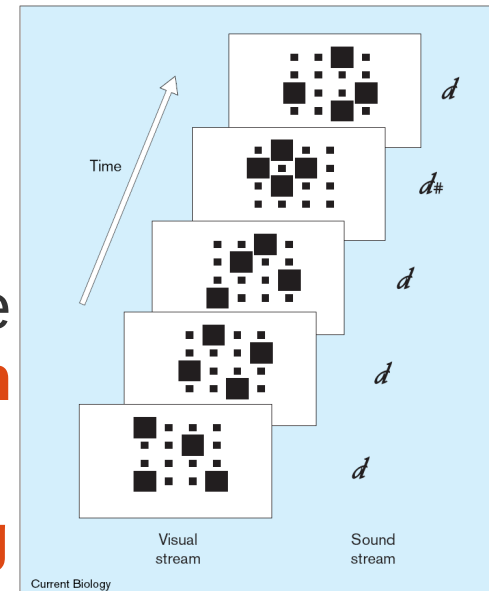
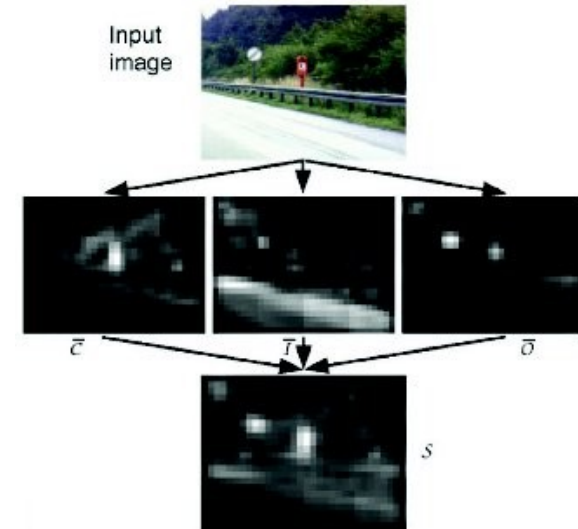
[matthias-rolf.blogspot.com](http://matthias-rolf.blogspot.com)



# Attention is more than visual saliency



- Robotics naturally focused on vision
  - E.g. saliency maps
- Possible agglomeration with audition
  - “multimodal” saliency maps
  - [Ruesch 2008, Schauerte 2011]
- Missing: **cross-modal** relations
- Important in humans: e.g. **synchrony**
  - E.g. cross-modal pop-out [Vroomen 2000]
  - Very important in early childhood:  
Infants prefer looking to synchronous stimuli
- Why? Because synchrony tells about the cross-modal **binding** and **segmentation** of events
- Important and yet difficult when **learning** about the world

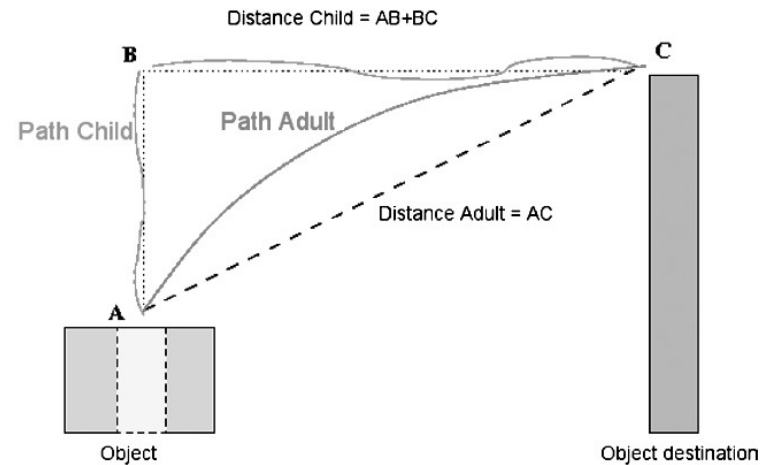


Current Biology

# Caregivers cues towards children



- Adults modify their actions when tutoring children
- Supposed to help structuring sensory stream
  - Highlighting of relevant stimuli and their relation
- Motherese [Fernald 1984]
  - Change of prosody
  - Attracts infants' selective attention [Fernald 1985]
- Motionese [Brand 2002]
  - Changed movements
- Cross-modal **synchronization**
  - Shown with manual event coding scheme [Gogate 2000]
  - Helps for learning [Gogate 2001]



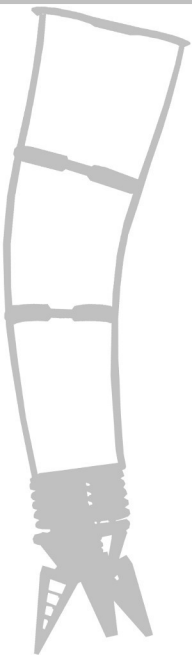
Rohlfing 2006



# Caregivers cues towards children



Original video corpus from [Rohlfing 2006]

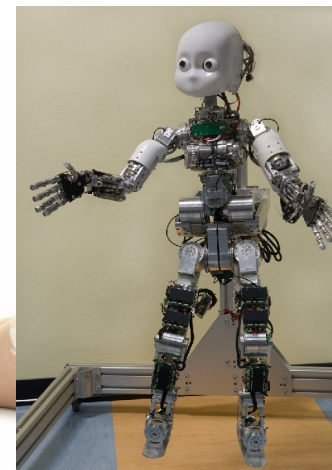


- Symbiotic interaction cycle [Rohlfing 2006]
  - Cross-modal parental cues
  - Infant attention & learning
  - Well tuned to each other
- How to benefit from that?
  - Robots take the infants' place?
- Tutoring cues also observed during human-robot interaction [Vollmer et al, 2009]

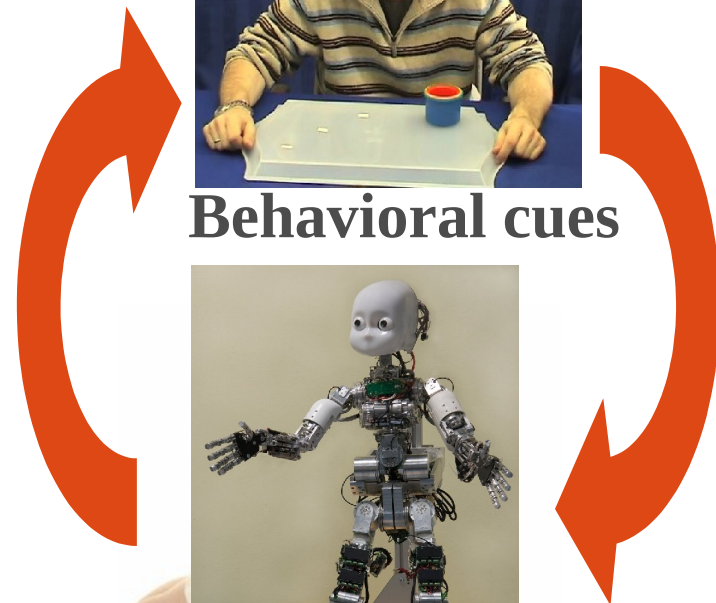
**Goal:** Make robot receptive to cross-modal synchronization cues during human tutoring.



**Behavioral cues**

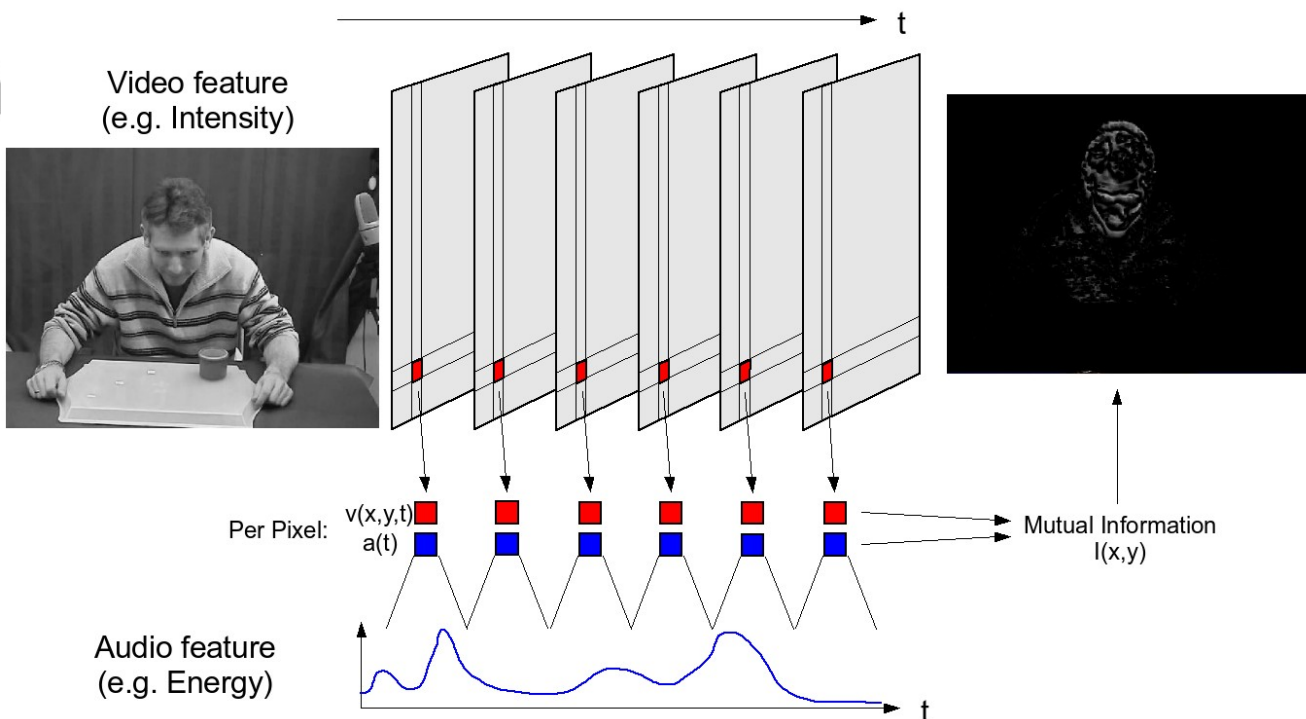
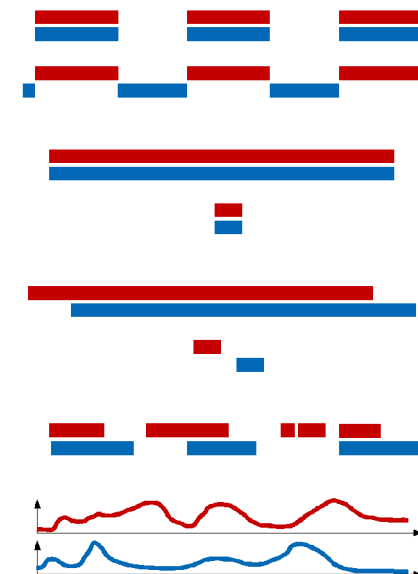


**Tuned attention**



# Modeling Cross-Modal Synchrony

- How to model synchrony?
  - Manual event-overlap coding [Gogate 2000]
- Try less pre-structured approach
- Signal-level synchrony
  - Mutual information, correlation [Hershey 2000]
    - few modifications, see [Rolf 2009]



**Synchrony “map”**  
→ **Attention**  
→ **Relevance (?)**

# Data set

- Video corpus from [Rohlfing 2006]
  - Subset: 184 videos
- Parents demonstrating tasks:
  - Towards their child, 8-30m (AC)
  - Towards their partner (AA)
- Four different tasks

**Key question:** can the system detect the parental cues?

- First step: assess overall synchrony of a demonstration
- Measure: average mutual information
  - Baselined against synchrony with audio white noise
- Direct comparison AC ↔ AA!

**Hypothesis:**  $\text{sync}(\text{AC}) > \text{sync}(\text{AA})$



(a) Cup stacking



(b) Wooden bricks

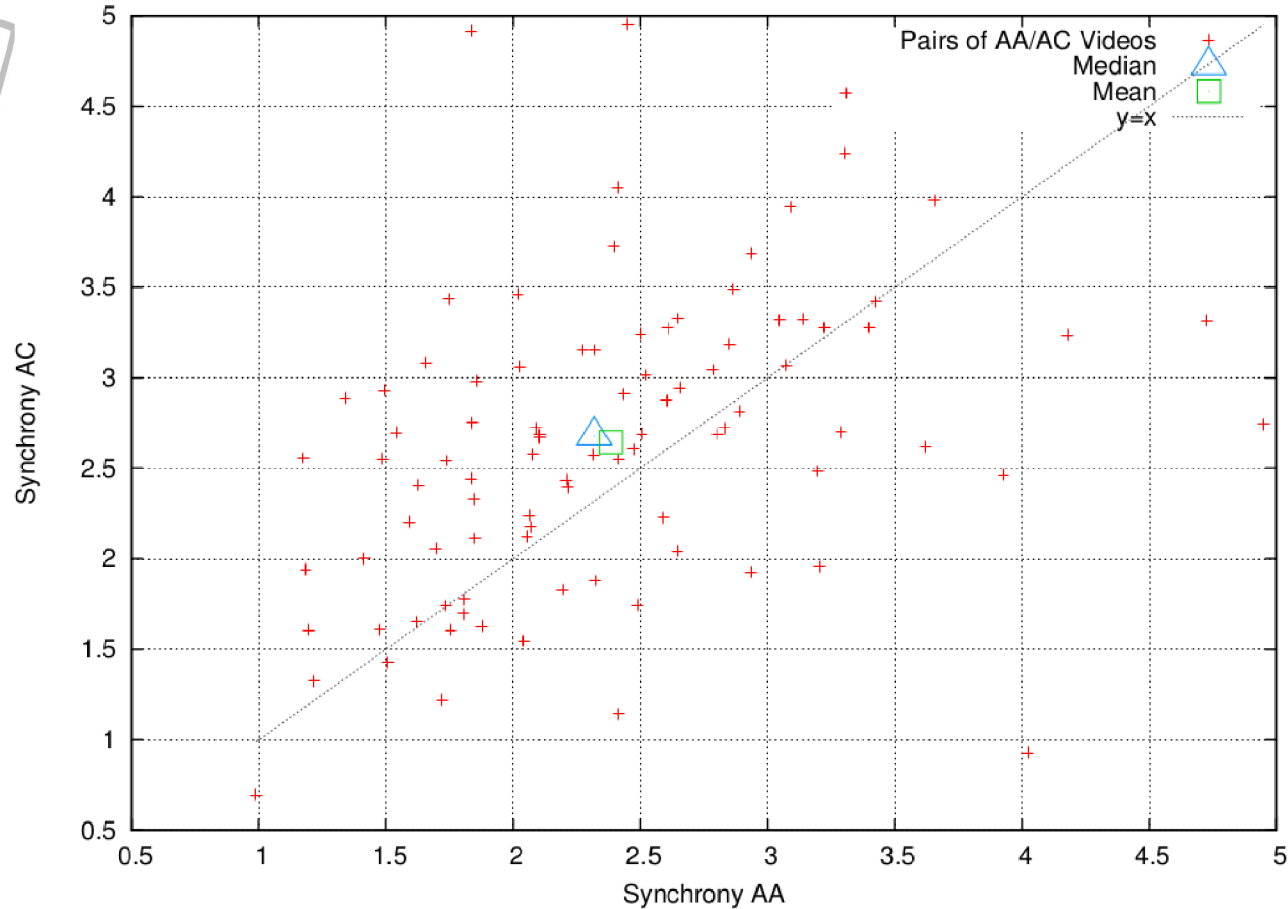


(c) Bell



(d) Salt shaker

# Results



Single point:  
AC synchrony  
vs. AA synchrony  
for same person

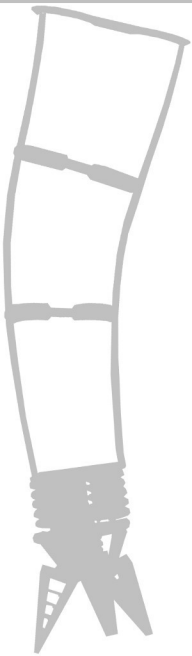
Above diagonal:  
 $\text{sync}(\text{AC}) > \text{sync}(\text{AA})$

~~$H_0: P[\text{sync}(\text{AC}) > \text{sync}(\text{AA})] = 0.5$~~   $p < 0.001$ , 2-tailed sign test

- System is receptive to parents' action synchronization
- First verification of such behavior with comp. model

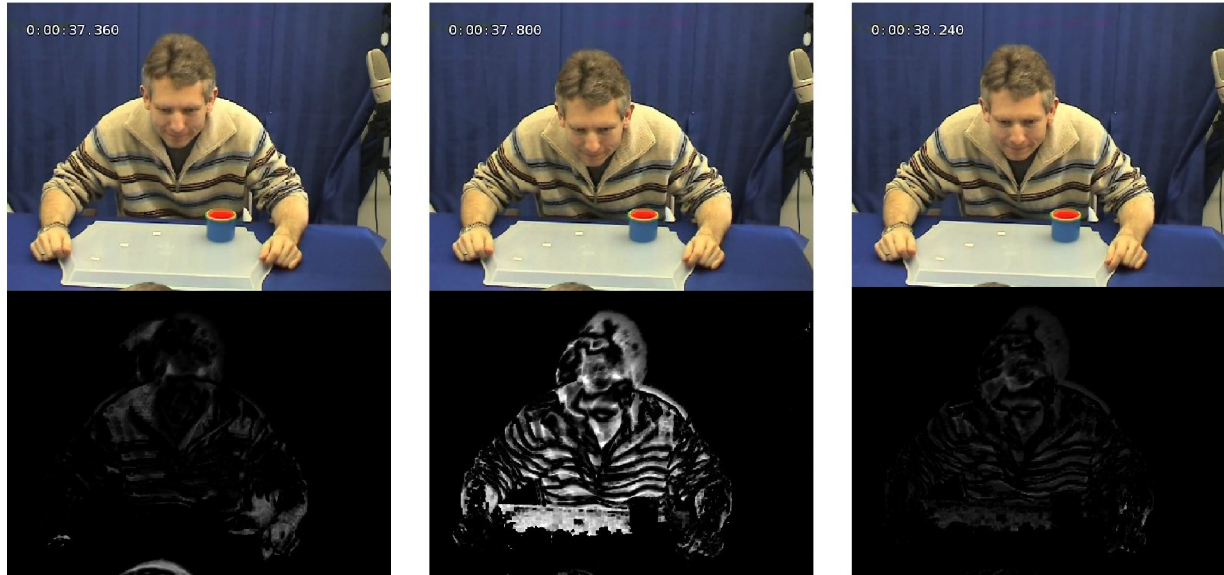


# Synchrony in Space

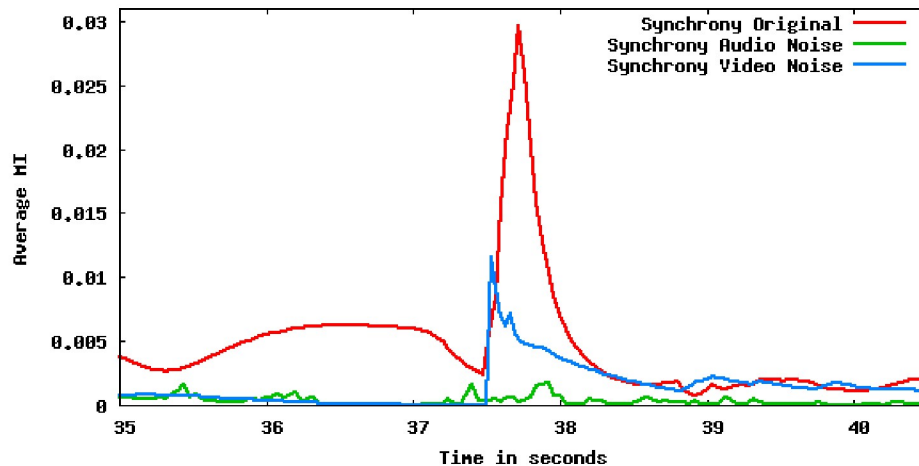


- Overt attention (“focus” of attention)
- Where is the maximum point of synchrony □?
  - Sound source: mouth / head
  - Synchronization with shown objects!
- Comparison: pure visual saliency □
  - Often distracted by simple contours

# Synchrony in Time



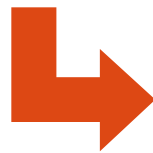
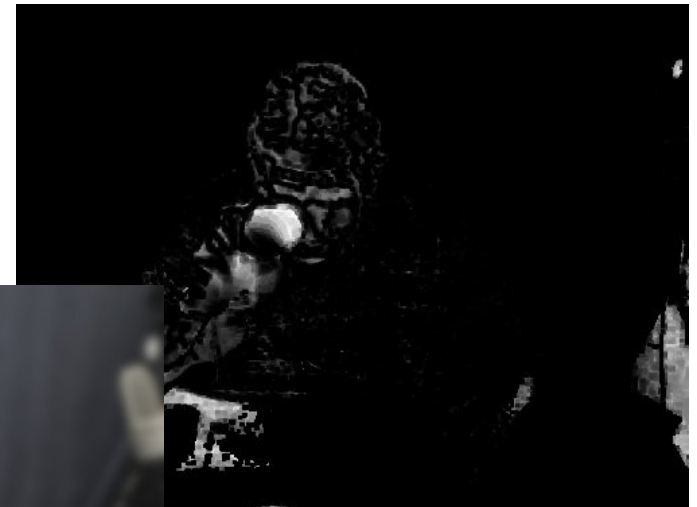
“hmm??”



- Infants sometimes get distracted.....
- One strategy: arouse with large synchronous move
- Peak synchrony

# Assistance Systems for ASD patients

- Current application idea: an assistive device for Autistic Spectrum Disorder patients
- Characterized by impaired attentional skills, in particular in social situations
- Idea: integrated display with highlighting and fadeout



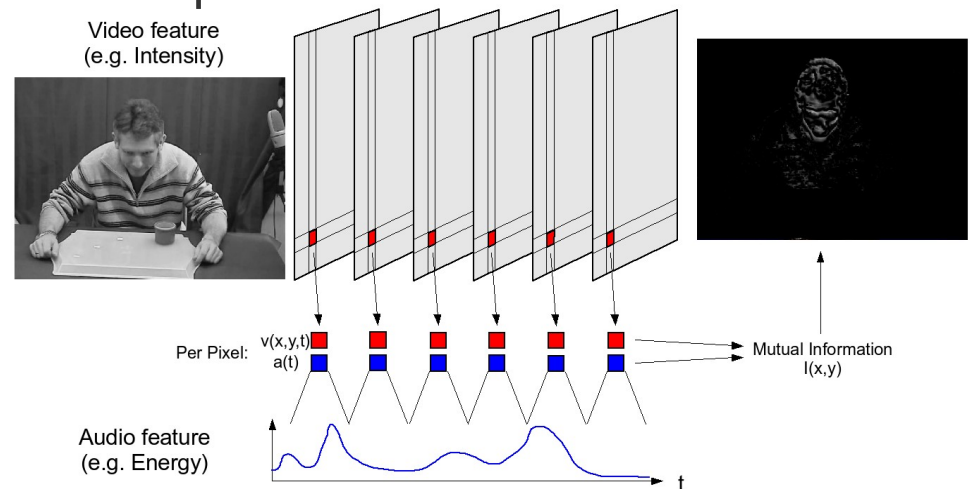
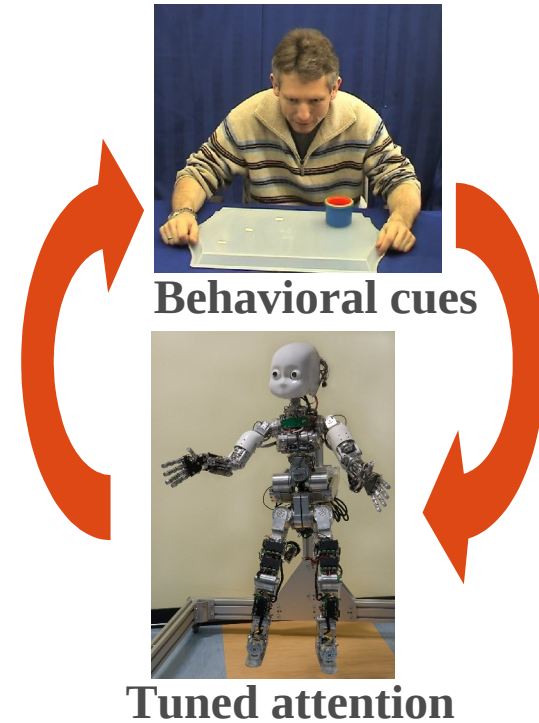
# The synchronous part of visual data...



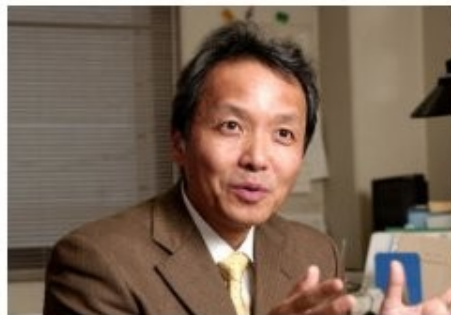
Assistive display prototype. Also an infant's view?

# Discussion

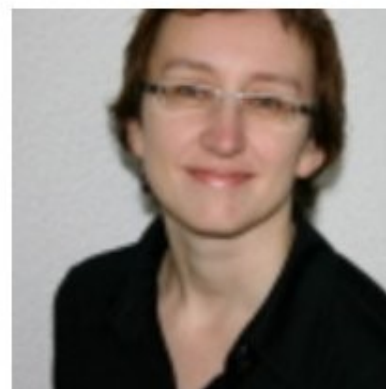
- Cross-modal attention goes beyond multi-modality, e.g. by synchrony
- Interaction cycle parent/infant
- Computational model:
  - Signal-level A/V synchrony
  - Synchrony guides visual attention
- Receptive to tutoring cues
  - First comp. verification of increased synchronization during human tutoring
  - Relevant patterns detected in space
- Applications
  - Find training data for learning [Grahl 2012]
  - Assistive systems?



# Acknowledgements



Minoru Asada, Yukie Nagai,  
Lars Schillingmann



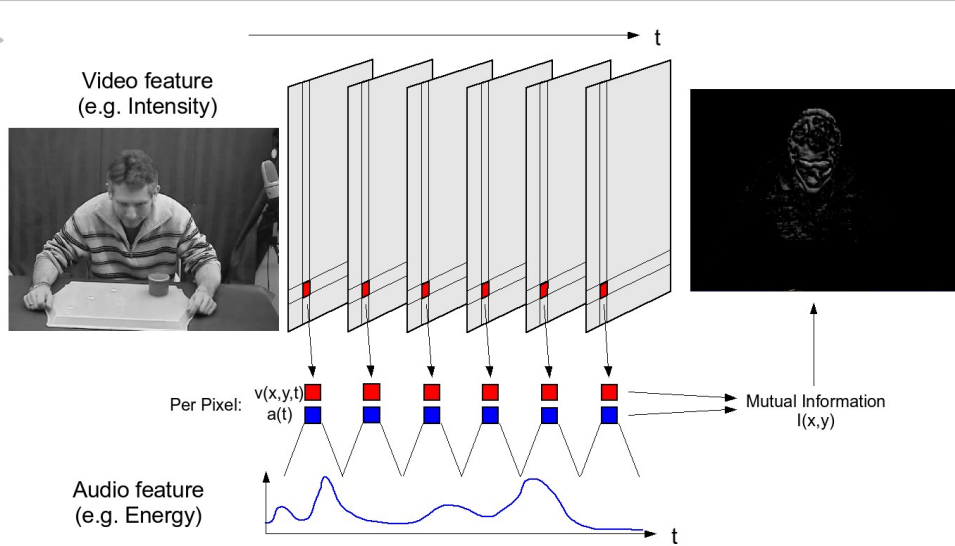
Katharina Rohlfing, Britta Wrede



Shinichiro Kumagaya, Satsuki Ayaya



Marc Hanheide  
UNIVERSITY OF  
LINCOLN



# Thank you for your (cross-modal?) attention!

matthias-rolf.blogspot.com

- M. Rolf, M. Hanheide, and K. Rohlfing. Attention via Synchrony: Making Use of Multimodal Cues in Social Learning. *IEEE Trans. Autonomous Mental Development*, 1(1):55–67, 2009.
- M. Rolf. Audiovisual attention via Synchrony. Master's thesis, Bielefeld University, 2008.
- L. Schillingmann, M. Rolf, S. Kumagaya, S. Ayaya, and Y. Nagai. Assistance for autistic people by segmenting and highlighting cross-modal perceptual information. In *Annual Conference of the Robotics Society of Japan (RSJ)*, 2013.
- J. Hershey and J. Movellan. Audio-vision: Using audio-visual synchrony to locate sounds. *Advances in Neural Information Processing Systems*, 12:813–819, 2000.
- K. Rohlfing, J. Fritsch, B. Wrede, and T. Jungmann. How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Advanced Robotics*, 20(10):1183–1199, 2006.
- L. J. Gogate and L. E. Bahrick. Intersensory Redundancy and 7-Month-Old Infants' Memory for Arbitrary Syllable-Object Relations. *Infancy*, 2(2):219–231, 2001.

