Towards Action Representation based on Acoustic Packages

Britta Wrede, Lars Schillingmann, Katharina J. Rohlfing

Overview

- **Cues for action segmentation in tutoring situations**
  - Background on Acoustic Packaging [Brand et al., 2007]
  - Computational Model of Acoustic Packaging and Evaluation
- **Action Learning in infants**
  - Inferences about other‘s goals [Gergeley, 2003]
- **AP for learning and representing actions**

Acoustic Packaging

How to associate information in different modalities for language and action learning?

- **Synchrony** [Zukow-Goldring, 1997] [Matatyaho, Mason & Gogate et al., 2007]
  - Synchronous object movement and verbal labeling enhances object learning
  - More low-level synchrony in ACI than in AAI [Rolf et al., 2009]

- **Acoustic Packaging** [Brand et al, 2007]
  - Synchrony between language and events helps to divide sequence of events into units [Hirsh-Pasek & Golinkoff, 1996]
  - Speech segment determines perceived (end of) action

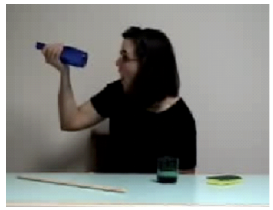
Acoustic Packaging [Brand et al., 2007]

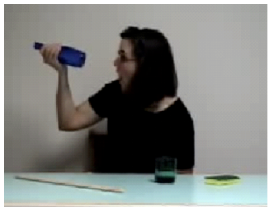
Question: Does speech influence how action is structured by infants?

Experiment: 32 Infants of 7.5 – 11.5 months of age; Preferential Looking

Familia-rization

Video





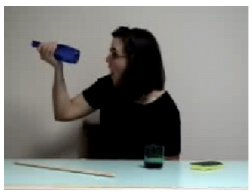
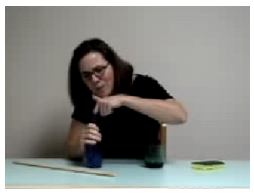

A B A C

Audio

Wow! Do you see what she's doing? She's blixing!

Test: Split Screen

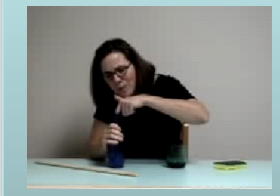
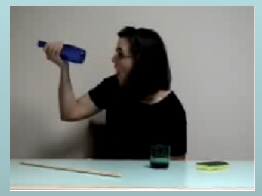
Video

A B

Preferred Sequence

B A

Non-packaged sequence perceived as new

⇒ **Speech structures action !**

Computational Model of AP

[Schillingmann et al., 2009]

Long term goals

- Temporal segmentation of actions
- Generating appropriate feedback
- Integration with action and speech learning approaches

Evaluation

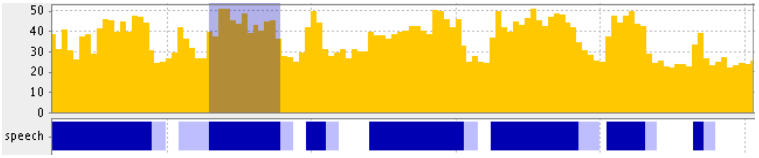
- Does model reflect structural properties of tutoring behavior?

Computational Model of AP

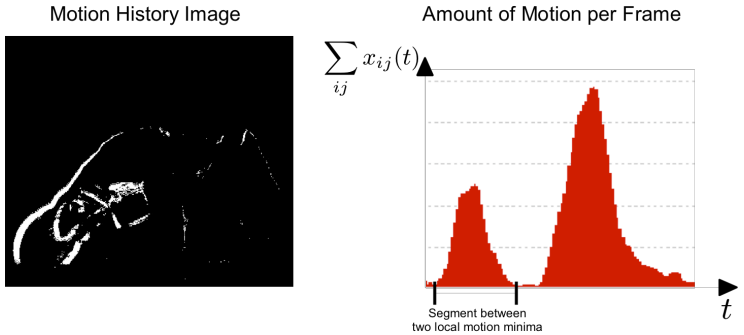
[Schillingmann et al., 2009]

Segmentation

Speech: by ASR (ESMERALDA)

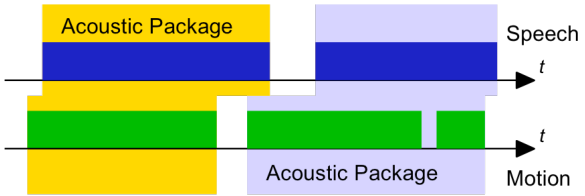


Action: by motion history images



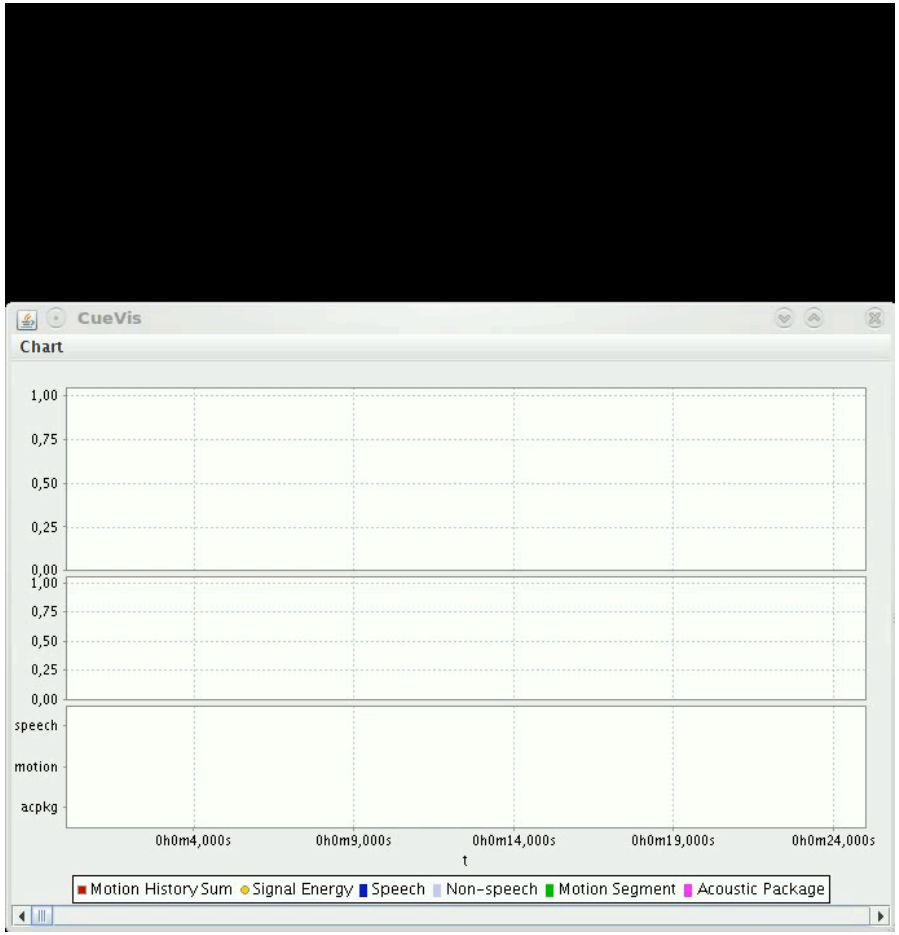
Temporal Association

Acoustic Package created if segments overlap



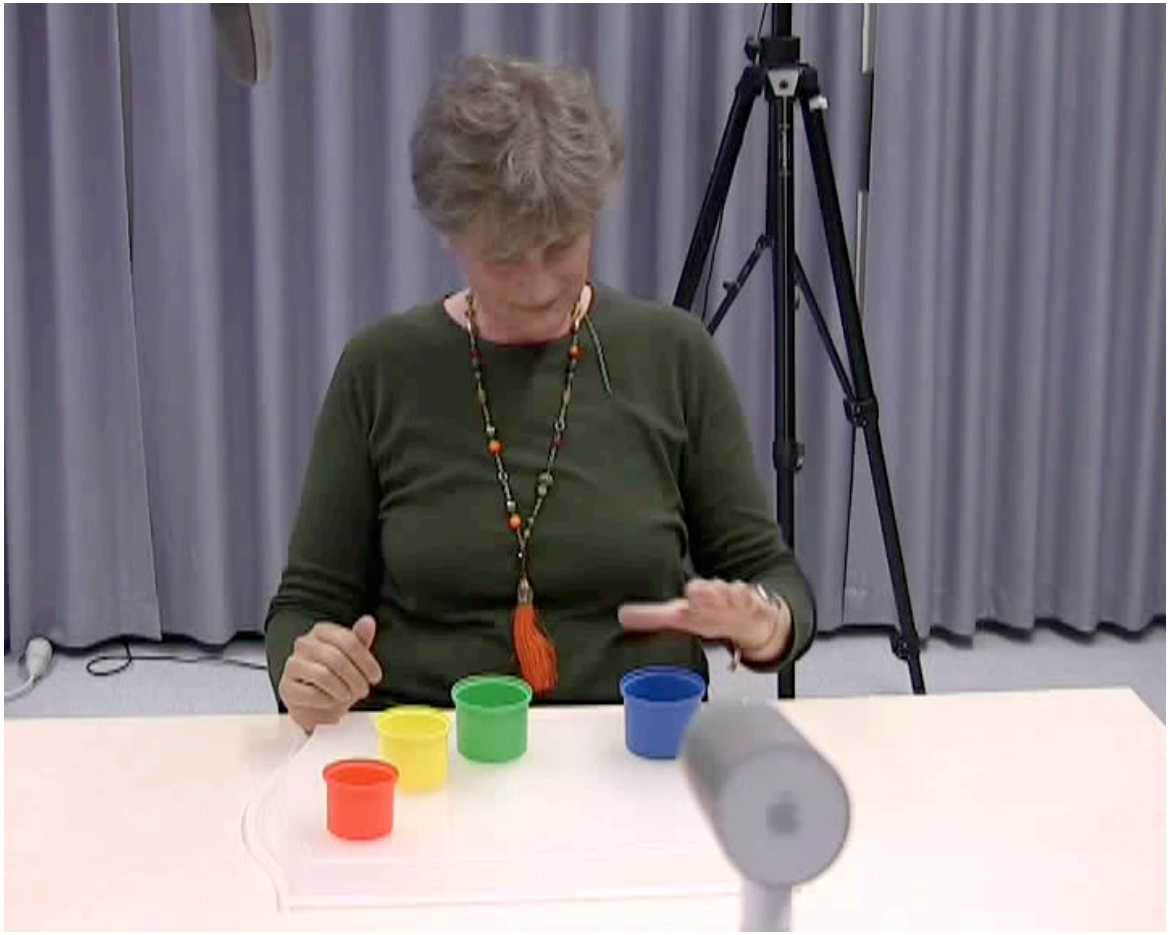
Computational Model of AP

[Schillingmann et al., 2009]



Computational Model of AP

[Schillingmann et al., 2009]



Evaluation

Data

- Videos from Motionese corpus (11 AAI, 11 ACI) and from babyface study (11 ARI)
- Task: stacking cups

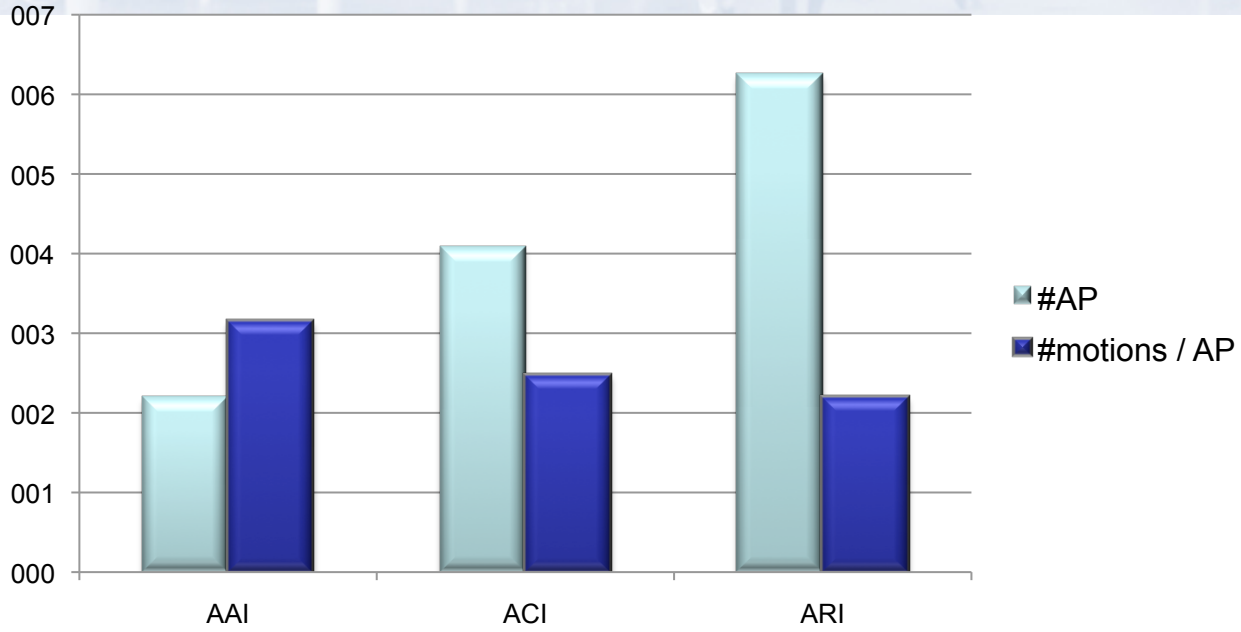
Analysis

- Automatic detection of Acoustic Packages
- Measurements:
  - number of Acoustic Packages (#AP)
  - mean number of motions per Acoustic Package (#motions / AP)

Hypothesis

- ACI more structured than AAI
- More #AP and less #motions / AP in ACI

Results



- Sig. more Acoustic Packages in ACI and ARI
- Sig. less Motions per Acoustic Packages in ACI and ARI

⇒ Hypothesis confirmed

⇒ Automatically detected Acoustic Packages find more structure in ACI and ARI

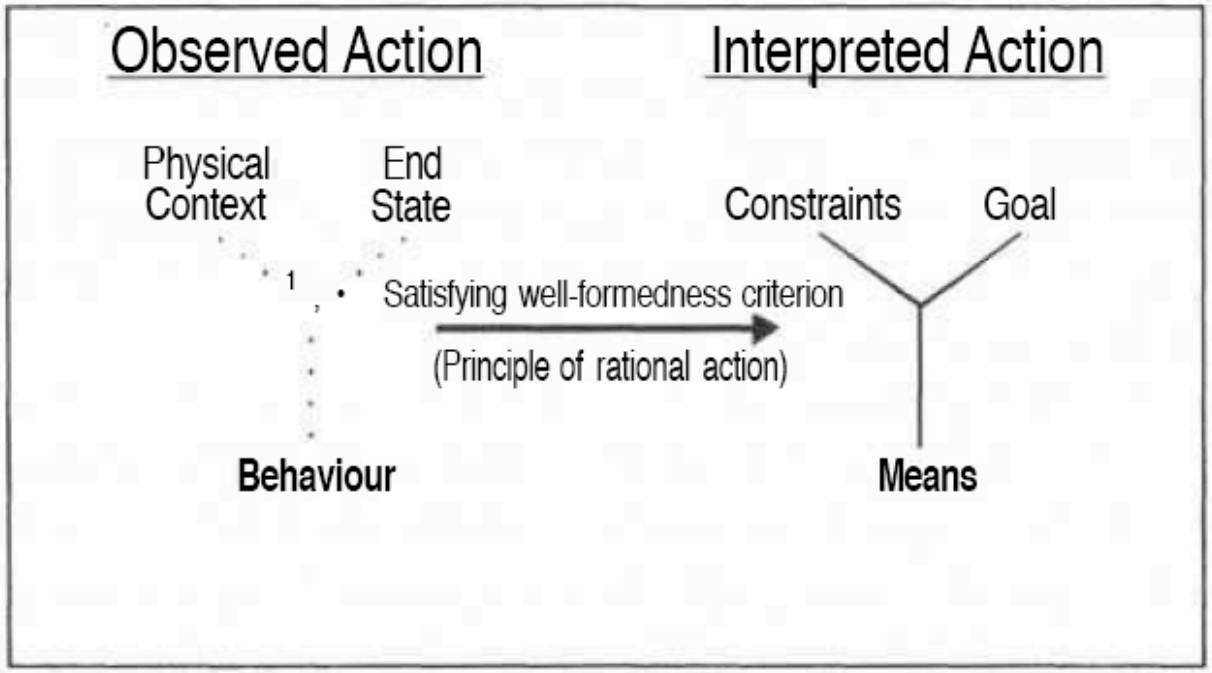
⇒ Acoustic Packages as basis for Action Representation?

Overview

- **Cues for action segmentation in tutoring situations**
  - Background on Acoustic Packaging [Brand et al., 2007]
  - Computational Model of Acoustic Packaging and Evaluation
- **Action Learning in infants**
  - Inferences about other‘s goals [Gergeley, 2003]
- **AP for learning and representing actions**

How to draw inferences about other's goal directed actions?

[Gergeley, 2003]



Assumption (well-formedness criterion):

Observed behavior

– will bring about goal state

– is most efficient means to reach goal

Support for Interpreting Action in IDS

Constraints

Goal

Means

Action learning in infants

Top-down processes: Language (syntactical constructions) can help to determine goal of action (e.g. path vs goal-oriented)

Goal-oriented

- „look the frog jumps **to the leave**“
- „look the yellow cup goes **into the red one**“

Path-oriented

- „look how the frog **jumps**“
- „look how you can **turn** the cup upside down“

Action learning in infants

Interaction can help to determine goal of action (e.g. path vs goal-oriented) – Hypothesis!

Goal-oriented

- Tutor: „look the frog jumps **to the leave**“
- (Infant lets the frog jump around)
- Tutor: „no no, the frog wants to go **to the leave**“

Path-oriented

- Tutor: „look how the frog **jumps**“
- (infant moves frog to the leave)
- Tutor: „no, it doesn‘t go like this, look how it **jumps**“

Overview

- **Cues for action segmentation in tutoring situations**
  - Background on Acoustic Packaging [Brand et al., 2007]
  - Computational Model of Acoustic Packaging and Evaluation
- **Action Learning in infants**
  - Inferences about other‘s goals [Gergeley, 2003]
- **AP for learning and representing actions**

Discussion – Acoustic Packages

- Acoustic Packages as a learner-oriented **segmentation** of the action
- Multi-modal binding
  - AP contain specifically chuncked structure (tying verbal constructions to visual movements or series of movements)
- Interaction
  - AP segmentation will differ depending on learner feedback

Discussion - Representation

- Multi-modal:
  - Verbal (lexical, syntactic constructions) (interpretation of observed behavior)
  - Visual (e.g. scene changes -> end state -> goal(s))
  - Trajectories (e.g. hand movements -> physical context -> constraints; hand movements -> behavior -> means)
  - Multi-modal structure (means, constraints, goal(s))
- Dynamic in nature:
  - Representation will change over duration of interaction
  - Representation will change over different interactions (and learning of other actions)

Thank you for your attention!