Towards a Computational Model of Child Gesture-Speech Production. How information is spread across modalities in pre-school children

Olga Abramov, Stefan Kopp, Anne Németh, Friederike Kern,
Ulrich Mertens and Katharina Rohlfing

SOCIAL COGNITIVE SYSTEMS
AG Kognitive Systeme und soziale Interaktion

Universität Bielefeld

CITEC

DFG

Abstract: Previous work by [1] studied gesture-speech interaction in adults. [1] focussed on temporal and semantic coordination of gesture and speech and found that while adult speech is mostly coordinated (or redundant) with gestures, semantic coordination increases the temporal synchrony. These observations do not necessarily hold for children (in particular with respect to *iconic* gestures, see [2]), where the speech and gesture systems are still under development. We studied the semantic and temporal coordination of speech and gesture in 4-year old children using a corpus of 40 children producing action descriptions in task oriented dialogues. In particular, we examined what kinds of information are transmitted verbally vs. non-verbally and how they are related. To account for this, we extended the semantic features (SFs) developed in [3] for object descriptions in order to include the semantics of actions. We coded the SFs on the children's speech and gestures separately using video data. In our presentation, we will focus on the quantitative distribution of SFs across gesture and speech. Our results indicate that speech and gestures of 4-year olds are less integrated than those of the adults, although there is a large variability among the children. We will discuss the results with respect to the cognitive processes (e.g., visual memory, language) underlying children's abilities at this stage of development. Our work paves the way for the cognitive architecture of speech-gesture interaction in preschoolers which to our knowledge is missing so far.

Motivation

- How is the meaning transmitted through the two modalities – speech and gesture in preschoolers? Defining a means to measuring speech-gesture redundancy.
- Statistical evidence: which types of gestures occur more frequently in one context than in others?
- Modelling the cognitive capabilities of children (age 4-5).

Related Work

Alibaili et al. (2009)
- narative data 17 children (5-10 years)
- 20 adults

→ children produced more non-redundant combinations than adults

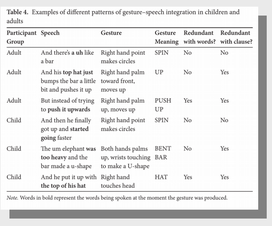


Table 4. Examples of different patterns of gesture–speech integration in children and adults

Participant Group	Speech	Gesture	Gesture Meaning	Redundant with words?	Redundant with clause?
Adult	And there's **a uh** like a bar	Right hand point makes circles	SPIN	No	No
Adult	And his **top hat just** bumps the bar a little bit and pushes it up	Right hand palm toward front, moves up	UP	No	Yes
Adult	But instead of trying to **push it upwards**	Right hand palm up, moves up	PUSH UP	Yes	Yes
Child	And then he finally got up and **started going** faster	Right hand point makes circles	SPIN	No	No
Child	The um elephant **was too heavy** and the bar made a u-shape	Both hands palms up, wrists touching	BENT BAR	No	Yes
Child	And he put it up with **the top of his hat**	Right hand touches head	HAT	Yes	Yes

Note. Words in bold represent the words being spoken at the moment the gesture was produced.

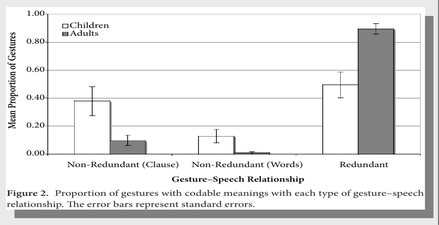
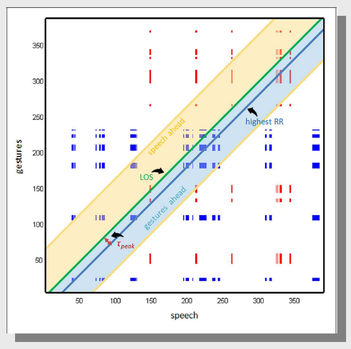


Figure 2. Proportion of gestures with codable meanings with each type of gesture–speech relationship. The error bars represent standard errors.

(-) However: speech has no separate semantics

Jonge-Hoekstra et al. (2016)
- 12 Dutch children, ~39,1 months old
- Natural teaching-learning interaction

→ increasing age: speech and gesture more synchronized



(+)Time series of gesture and speech
(+) measuring speech-gesture Interaction
(-) no semantics

Bergmann & Kopp (2006)
- 28 adults in a direction giving experiment
- semantic features of gestures based on their lexical affiliate

(+) taxonomy of semantic features
(-) the taxonomy is mainly based on objects and their properties, the specifics of actions are not included

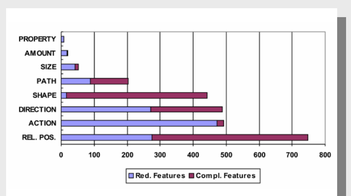


Figure 4: Distribution of the different kinds of redundant/complementary SFs.

Data

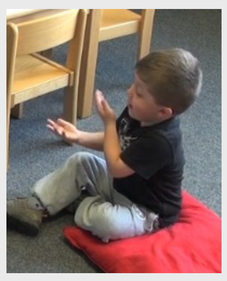
- 40 children were recorded at the age of 4 in different experimental situations
- Context 5: retelling the mother how the dog-puppet did everything wrongly
- Children: Intonation phrases: 1299, iconic gestures (hand and body): 255

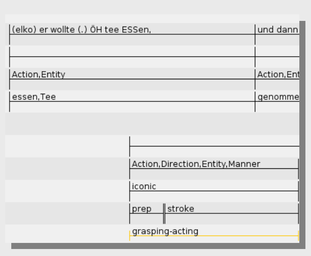
Context 1: playing a game with the experimenter

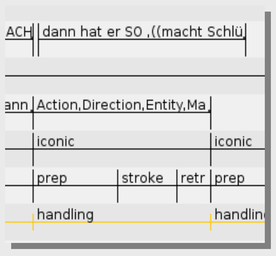
Contexts 2,3,5: retelling to the mother

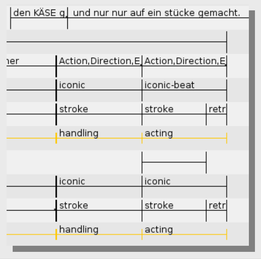
Context 4: a puppet-dog does everything wrong

Example Gestures



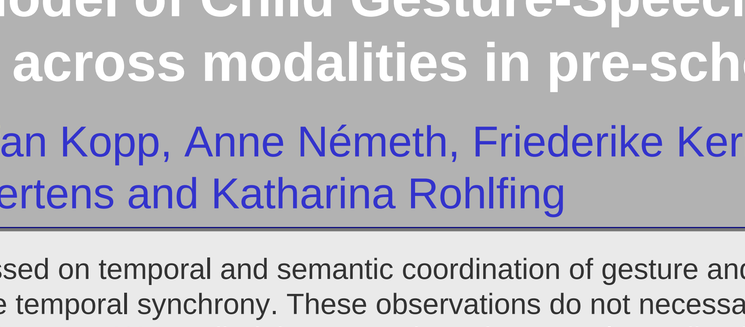






Annotation

- Semantic features were annotated separately on speech and gesture
- Inter-rater reliability: 20% of the data by 2 independent coders
- Kappa: speech (0.89), gesture (0.76)



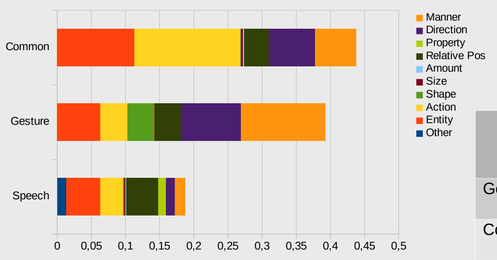
Semantic Features of actions

Semantic features present in gesture and speech, related to shape, postion of objects and actions (manner, path, action with objects etc.) (cf. Bergmann & Kopp 2006)

Question: how are semantic features distributed across modalities?

Results

Global corpus statistics

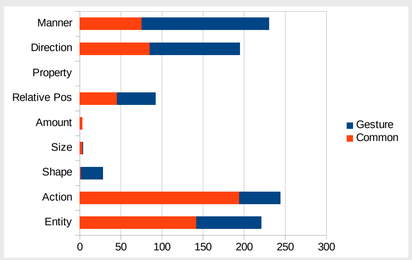


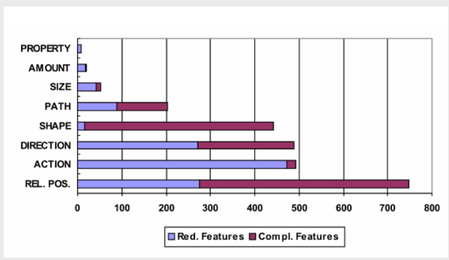
Corr S+G:	0.33
Corr G+C:	0.55
Corr S+C:	0.68
H(SF):	0.54
H(Mod):	0.31
MI:	0.04

- 46 % of SF are complementary(gesture), 54% are common (cf. Cassell et al. 1996, 2000)
- some features appear mostly in speech (property, amount, relative pos)
- others in gesture (manner, shape, direction)

	Entity	Action	Shape	Size	Amount	Relative Pos	Property	Direction	Manner	Total
Gest%	36	21	96	25	0	51	0	56	67	46
Com%	64	79	4	75	100	49	0	44	33	54

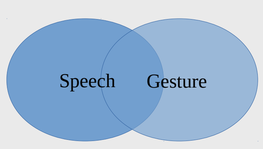
Child action reports vs. direction giving dialogues





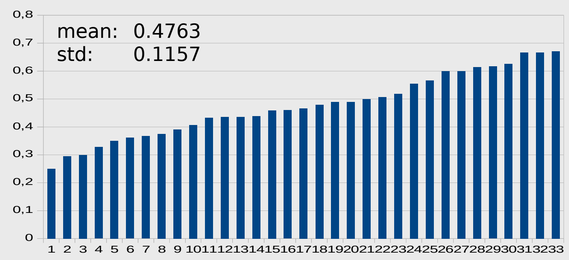
- **manner** is in 67.4 % complementary to speech
- **shape** (compared to adults) is used mostly in gesture (96,4 %)
- **property** is used only in Speech (not displayed here)
- **direction** and **relative pos**, though slightly larger in gesture (56.4 % and 51.1 %) are equally balanced in gesture and common
- **amount** and **size** are distributed similar to adults

Jaccard Similarity as a measure of gesture-speech integration

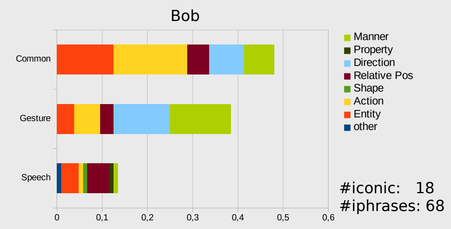


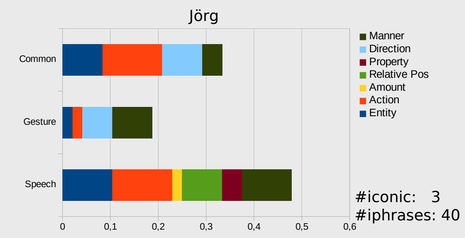
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

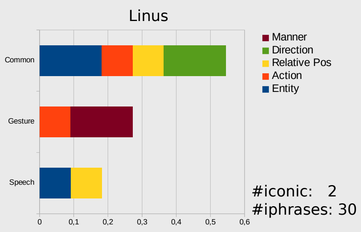
Average Jaccard Index per child (33 children)

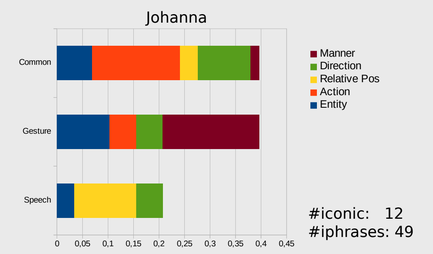


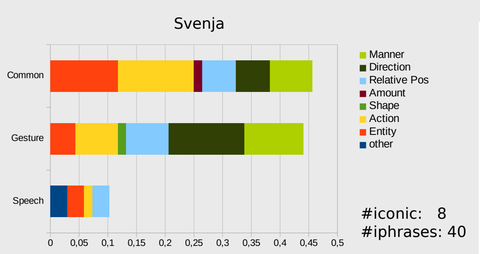
Individual child profiles







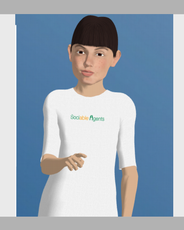




- Children with similar numbers of iconic gestures and intonation phrases can highly differ in the use of semantic features
- Bob's distribution is close to the average and also resembles the one of the adults (50:50)
- Jörg has a verbal preference
- Linus and Johanna have a specific language impairment (SLI)

Discussion / Outlook

- The overall distribution of semantic features is similar to the distribution of the adults
- The overall rate of overlap (Jaccard index) is ~ 48% (+- 12%)
- However, individual child profiles reveal substantial differences among the children
- Gesture speech integration seems to compensate problems in speech development
- This study paves the way for the cognitively plausible model of a 4-year old where different parameters observed emperically will be varied



References

Alibali et al. (2009). Gesture-speech integration in narrative: Are children less redundant than adults?. Gesture. 9. 290-311.

Justine Cassell and Scott Prevost (1996). Distribution of semantic features across speech and gesture by humans and computers. In Proc.: Workshop on integration of gesture in language and speech.

Justine Cassell et al. (2000). Coordination and Context-dependence in the generation of embodied conversation. In Proc.: 1. International conference on natural language generation.

de Jonge-Hoekstra, Lisette. (2016). Asymmetric dynamic attunement of speech and gestures in the construction of children's understanding.

Bergmann, Kirsten & Kopp, Stefan. (2006). Verbal or visual? How information is distributed across speech and gesture in spatial dialog.

Dr. Olga Abramov, oabramov@techfak.uni-bielefeld.de, URL: http://wwwhomes.uni-bielefeld.de/opustylnikov/pustylnikov/