Bielefeld Vision for Roadmap

Britta Wrede, Katharina Rohlfing, Gerhard Sagerer

1. Background of current research within ITALK

The Bielefeld perspective on developmental learning is based on the idea that learning is driven through interaction. This idea is supported by Gergeley (2003) and Csibra & Gergeley (2006) who state that learning through imitation is limited because the observed action not always reveals its meaning. In order for a learner to understand an action, s/he needs to be provided with additional information given by a teacher who demonstrates what the goal, the means and – most importantly – the constraints of a task are („pedagogy").
On the other hand, in order for the teacher and learner to engage in a meaningful interaction, they have to follow certain interactive rules. Such interactive rules have been described in terms of „grounding" (e.g. by Clark (1992)) otherwise more generally with the term "turn-taking" or „contingency". Clark (1992) provided one of the first grounding models with the claim that every individual contribution to a discourse has to be registered by the listener, that is the listener has to provide a signal of understanding in order for both participants to add the content to their pool of commonly shared information and beliefs ("common ground"). More generally, the term contingency refers to a temporal sequence of behavior and reaction and it has been shown that it plays an important role in the process of developmental learning (e.g. Kindermann, 1993). Mechanisms that detect (and produce) contingency could be a pre-cursor of later dialogical competencies as described in the framework of grounding. While contingency mainly describes a temporal pattern, where one event occurs as an answer to a previous one, grounding relies on semantic information in the sense that one event (or speech act) needs to be grounded by an interaction partner through a signal of understanding.

In our research within ITALK, we are focusing on the first aspect (Tutoring) by analyzing and modeling tutoring behavior and more specifically the mechanism of „synchrony". In experimental settings infants have been shown to learn object names more easily when the verbal referent was uttered in synchrony with a movement of the named object. In contrast, the name of an object being moved out of sync was not learned as well (Zukow-Goldring, 1997; Gogate et al., 2003). While synchrony has been described as a means to provide „invariance" we are at the same time analyzing the variability of the tutor behavior in order to better understand how tutors structure their actions towards infants. Here we follow the idea of "acoustic packaging" that has been pushed forward in experimental work by Brand (2007). Her results indicate that apparently infants bind sequences of (sub)actions together based on their co-occurrence with speech. That is, given an action sequence and a verbal utterance overlapping with only part of this sequence, infants are likely to interprete only those action sequences as belonging together that fall within the range of the verbal utterance.
The overall goal of our activities in ITALK is to build a theoretically and technically sound framework for modeling learning through interaction by making use of the phenomenon of synchrony as a means to anchor actions in speech.

Current results – based on automatically derived measurements directly from the audio and video signal (Rohlfing & Fritsch, 2006) suggest that we can indeed identify specially structured actions when demonstrated towards children through analyzing synchrony of speech and action (Wrede et al, 2005; Rohlfing & Jungmann, 2005). The results indicated that the speech on- and offsets as well as the intonation contour in infant directed speech are segmenting actions in smaller sub-units than in comparable adult directed speech.

2. Open questions from a theoretical perspective

While there is a growing body of research on the phenomenon of synchrony there exists up to now no definition that would allow to model synchrony on an artificial system. Based on current results reported in literature the following questions arise:

- What is synchrony (in terms of termporal structure as well as correlation measure)? (Definition)
- What are the entities that synchrony works on? (Segmentation)
- How can it be detected in the signal? (Recognition)
- What functions does it serve? (Model)
- What is the role of the different modalities (e.g. does vision provide rather spatial information whereas auditive synchrony is more related to temporal structure?) and how do they play together?

The scientific debate (especially reflected in the ZiF Workshop on “Action Structuring”, Bielefeld, 3rd-5th July 2008) currently seems to converge towards a consensus that the important criteria for synchrony are (1) temporal co-occurrence of an event in different modalities and (2) a correlation between the characteristics of these events. In contrast, “inverse synchrony” where events in two modalities show a temporally exactly disjunct distribution – such as a sequence of speech being followed by a speech pause with a sound of noise that is deliberately being framed by the tutor’s utterance – does not belong to the term of synchrony but rather describes the characteristics of causality or – within the context of interaction - contingency.

A further important issue is the question of contingency. As outlined above, contingency as a temporal pattern may be interpreted as a precursor of turn-taking and grounding mechanisms and is a powerful mechanism in the context of developmental learning. While there is agreement that contingency is an important factor in the cognitive development of infants – as researched within the still face paradigm (e.g. Tronick et al., 1978; Muir & Lee, 2003) - there is only very little work on the exact definition and how it might work in children. There is evidence that mothers decrease their level of contingency with their infant’s increase of development for a certain task (Kindermann, 1993) indicating that contingency plays an important role in teaching. Also infants have been shown to not only detect contingency but also to expect and to try to elicit it (Okanda & Itakura, 2006). In accordance with this infants prefer persons who are and have previously been interacting contingently with them (Bigelow & Birchba, 1999). This importance of contingency has been recognized by computer scientists and there exist already some computational models for contingency (e.g. Di Paolo et al., 2007). However, these models tend to be uni-modal and very limited to a concrete application where an “event” is clearly defined (e.g. Auvray et al., 2006). In order to foster research with respect to developmental learning on robots the following questions need to be addressed in the near future:

- What is contingency (in terms of temporal structure as well as with respect to semantic content, if any)? (Definition)
- What are the entities that contingency works on? (Segmentation)
- How can contingency be detected in the signal? (Recognition)
- What functions does it serve? (Model)
- What is the role of the different modalities and how do they play together?

Based on such knowledge about the notions of synchrony and contingency in the framework of developmental robotics the question how these two phenomena play together can be tackled. Our current hypothesis is that in order for an infant to learn new actions it can rely (1) on structuring information provided by the tutor through heavy use of synchrony and (2) on grounding – and thus influencing - this information within the interaction based on a mechanism of contingency. Within this context the following questions arrise:

- What role does contingency play in the tutor's behavior with respect to synchrony? E.g. is it the case that the infant through its feedback is actually designing the way the tutor is structuring the action for the infant?
- Is the development of contingency and synchrony models interdependent? If so, how do they depend? (E.g. is synchrony a pre-cursor of contingency?)

3 Open questions from a technical perspective

Research on the above drafted more theoretically oriented questions need to be framed in the development of human-robot interaction research. More specifically we identified three main research areas from this perspective:

- Use of learning through interaction paradigm for approaches to bootstrap recognition or interaction capabilities (e.g. automatic speech recognition or dialog / contingency mechanisms)
- Analysis of more modalities (e.g. gaze, facial expressions for mor socially related functions and hand movements / gestures for more task-oriented functions)
- Development and strengthening of existing and new methodologies
    - Evaluation through human-robot interactions
    - Strongly intertwined development-evaluation cycles

In parallel the question if there is "cognition beyond human cognition" should be targeted, based on the following considerations:

- Humans are able to adapt to changing environments very quickly. Is it thinkable and feasible that we can observe this adaptation process by analyzing the development of human behavior when confronted with new modalities that are provided by new techniques?
- When building artificial cognitive systems we will be confronted with the fact that artificial systems will have artificial states that are completely different from human ones (e.g. „system crash" or „reboot") due to their specific hardware and their rather modular software.
    - Do we think that these states are ok for modeling cognition?
    - Is it possible for humans to learn such states (this is more relevant for HRI in general)

References

Auvray, M., Lenay, C. and Stewar, J. (2006). The attribution of intentionality in a simulated environment: the case of minimalist devices. In: Proc. Of 10th Meeting of the Association for the Scientific Study of Consciousness, UK, June, 2006.

Bahrick, L. E., Lickliter, R. and Flom, R. (2004). Intersensory Redundancy Guides the Development of Selective Attention, Perception, and Cognition in Infancy. Current Directions in Psychological Science 13(3), 99-102.

Bigelow A.E.; Birch S.A.J. (1999). The effects of contingency in previous interactions on infants' preference for social partners.In: Infant Behavior and Development, Volume 22, Number 3, pp. 367-382(16)

Brand, R. J. and Tapscott, S. (2007). Acoustic Packaging of Action Sequences by Infants. Infancy 11 (3), 321 - 332.

Csibra, G. & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In: Processes of Change in Brain and Cognitive Development, Attention and Performance, XXI. Pp. 249-274.

Di Paolo, E., Rohde, M. and Itzuka, H. (2007). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. In: New Ideas in Psychology.

Gergely, G. (2003). What should a robot learn from an infant? Mechanisms of action Interpretation and observational learning in infancy. Connection Science, 13(4), 191-209.

Gergely, G. (2003). What should a robot learn from an infant? Mechanisms of action Interpretation and observational learning in infancy. Connection Science, 13(4), 191-209.

Gogate, L., Bahrick, L.E. & Watson, J.D. (2003). A Study of Multimodal Motherese: The Role of Temporal Synchrony between Verbal Labels and Gestures. In: Child Development 71(4), 878–894.

Kindermann, Thomas A. (1993). Fostering Independence in Mother-Child Interactions: Longitudinal Changes in Contingency Patterns as Children Grow Competent in Developmental Tasks. International Journal of Behavioral Development, Vol. 16, No. 4, 513-535

Matatyaho, D. J., Mason, Z. and Gogate, L. J. (2007). Word learning by Eight-Month-Old Infants: The Role of Object Motion and Synchrony. Epigenetic Robotics, 201-202.

Muir, D.; Lee, K. (2003). The Still-Face Effect: Methodological Issues and New Applications. In: J. Infancy, Volume 4, Issue 4, January 2003, pp. 483-491.

Okanda, Mako; Itakura, Shoji (2006). Development of contingency: How infants become sensitive to contingency? In: Proc. of the XVth Biennial International Conference on Infant Studies, Kyoto, Japan, 2006

Schwade, J.A., Goldstein, M.H., Stone, J.S., Wachterhauser, A.V.Z. (2004). Children's use of speech and motion cues when learning novel words. Poster at ICIS 2004.

Tronick,E., Als,H., Adamson,L., Wise,S. & Brazelton, T.B. (1978). The infants' response to entrapment between contradictory messages in face-to-face interactions. Journal of the American Academy of Child Psychiatry, 17, 1–13.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (Eds.), Evolving explanations of development: Ecological approaches to organismenvironment systems (pp. 199–252). Washington, DC: American Psychological Association.