Learning to Manipulate Objects: A Quantitative Evaluation of Motionese

Katharina Rohlfing
Communication Science and Disorders
Northwestern University
k-rohlfing@northwestern.edu

Jannik Fritsch and Britta Wrede
Applied Computer Science
Technical Faculty, Bielefeld University
{jannik,bwrede}@techfak.uni-bielefeld.de

It is a well-known phenomenon that adults are speaking differently when addressing children. The purpose and the perceptual parameters of the child directed speech or Motherese have been discussed over the last three decades. Recently, this particular behavior has been characterized as directing the child's attention to relevant aspects of the speech signal [2]. What if - not limited to the speech signal - children are learning from input that is specifically designed for them? This is an idea that has barely been considered in the classical view on learning in pattern recognition but is gaining popularity in developmental psychology. Brand et al. [1] characterize a modification in mothers' infant-directed action as Motionese. In their studies, they observed that mothers' infant-directed actions reveal distinctive characteristics that amplify or exaggerate meaning and structure within their bodily motions. These characteristics were identified using 8 intuitive categories: range of motion, rate, repetitiveness, proximity to partner, enthusiasm, interactiveness, punctuation and simplification. Human coders gave each object demonstration a rating (0-4) for each category. However, the definition of the categories was tailored for human coders and often a single feature like, e.g., velocity was contained in several categories.

We present an algorithmic solution to identify visually observable modifications in actions in a quantitative manner requiring only the video stream. Our approach is based on a system for recognizing manipulative gestures [3]. The human hands are tracked based on skin color and the acting hand is related to objects in its vicinity based on their distance. Based on these processing steps, a variety of parameters like, e.g., hand velocity can be measured objectively. While our findings are in accordance with Brand's observations, we also provide insights into which specific aspects of the demonstration are relevant for motionese.

In our case study, we looked at an adult addressing either a child or another adult in his actions. Our approach allows us to identify the differences with respect to infant-directed action. In addition to the features extracted from visual data we performed an analysis of the sound signal to achieve a quantification of the multi-modal information directed towards the child. We hypothesize that cues in the acoustic signal can help to guide attention to relevant parts of actions. For example, the end of many actions like placing a cup on a table is signaled by the noise of the cup hitting the table top thus indicating the goal position of both the hand and the object. Analyses of the acoustic signal allow to highlight the action structure by (1) aligning noise from actions to the trajectories of gestures and (2) identifying especially emphasized regions in the speech signal.

Our findings suggest that techniques such as tracking manipulative gestures are an important step towards the analysis of qualitative and quantitative differences in the multi-modal input. They also indicate that multi-modal cues can help in understanding (and learning) of actions and their meaning. This will help research in robotics to develop ways of how to reduce the complexity of the acquisition problem, i.e. to reduce the requirements on built-in architectural constraints of learning mechanisms [4]. We will also discuss the potential application possibilities (such as plug-ins to transcription programs) to make these techniques available to developmental psychologists.

[1]This work has been supported by the German Research Foundation within the Emmy Noether-Program, the Collaborative Research Center 'Situated Artificial Communicators' as well as the Graduate Program 'Task Oriented Communication'.

References

[1] R. J. Brand, D. A. Baldwin, and L. A. Ashburn. Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*, 5(1):72–83, 2002.

[2] P. Dominey and C. Dodane. Indeterminacy in language acquisition: the role of child directed speech and joint attention. *Journal of Neurolinguistics*, 17:121–145, 2004.

[3] J. Fritsch. *Vision-based Recognition of Gestures with Context*. Dissertation, Bielefeld University, Technical Faculty, 2003.

[4] G. Gergely. What should a robot learn from an infant? Mechanisms of action interpretation and observational learning in infancy. In *Proc. Third Int. Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 13–24, Boston, MA, USA, 2003.