Embodied Behavior Processing in ECAs by Perception-Action Integration

Amir Sadeghipour
CITEC, Bielefeld University
P.O. 100 131
33501 Bielefeld, Germany
asadeghi@techfak.uni-bielefeld.de

Stefan Kopp
CITEC, Bielefeld University
P.O. 100 131
33501 Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
Perception and generation of verbal and nonverbal behavior is one of the main foundations of human social interaction. We model these abilities for embodied conversational agents (ECAs) on the basis of perception-action links as in humans. With a focus on gesture processing, we propose a computational model which enables ECAs to interact with humans in an embodied manner and supports many aspects of social interaction. The model performance is briefly illustrated on the basis of an interaction scene.

Keywords
Perception-action Links, Gestures, Cognitive Model, Social Embodiment

ACM Classification Keywords
I.2.10 Vision and Scene Understanding - Motion [Cognitive Model, Interactive System]

Introduction

An increasing number of findings and theoretical considerations in Cognitive Science suggest that human interaction and intersubjectivity is grounded in embodied processes. According to this view, perceiving and generating behavior are not separate processes but are both grounded in the perceiver's own motor repertoire (cf. *mirror neurons*). Moreover, such couplings between perception and action can be considered as a basis for creating common ground, mutual coordination, and *social resonance* [3]. Some of these processes apply also to the interaction of humans with artificial anthropomorphic agents [4]. The development of embodied conversational agents (ECAs), however, has so far neglected embodied processing of social behavior. Although coupling of perception and action has been touched upon by work on computational models of mirror neurons and in particular imitation learning [5], these approaches do not focus on social behavior, which requires fast and concurrent processing based on motor resonances during observation. In this paper we propose a computational model for the processing of communicative hand gestures in ECAs when interacting with humans. In general, this model has to account for a number of behaviorally and neurobiologically suggested requirements: (1) Hierarchical structure: Perception-action links [1] are assumed to be effective at various levels of a hierarchical sensorimotor system, from kinematic features to motor commands to goals and intentions [2]. (2) Motor resonance: Motor representations are shared between processes of perception and generation, and this accounts for motor resonances and covert imitation during embodied perception [7]. (3) Top-down and bottom-up processing: The levels of the action representation hierarchy in the model must be able to interact bidirectionally with each other [8] during both perception and generation. (4) Fast and incremental processing: With incoming stimuli, resonances and activation of sensorimotor structures must arise in a fast, robust, concurrent and incremental manner. (5) Imitation learning: The integration of perception and generation abilities must support the social learning of behaviors. (6) Interpersonal coordination: Perception-action links provide a likely basis for the fast and often non-conscious interpersonal coordinations (e.g., alignment, mimicry, interactional synchrony) that lead to rapport and social resonance [3] between interactants.

In the remainder of this paper, we describe our computational model and show how these requirements are met.

The Computational Model

Our computational model provides an ECA with motor knowledge that is shared between – and interacts with – perception and generation processes (see Figure 1). On the one hand, the perception module receives wrists' spatial positions of a human interlocutor at each time step, preprocesses them, and tries to recognize or learn gestures based on the shared motor knowledge. On the other hand, the generation module employs the represented motor knowledge to control the wrists movements of the ECA for gesture generation.

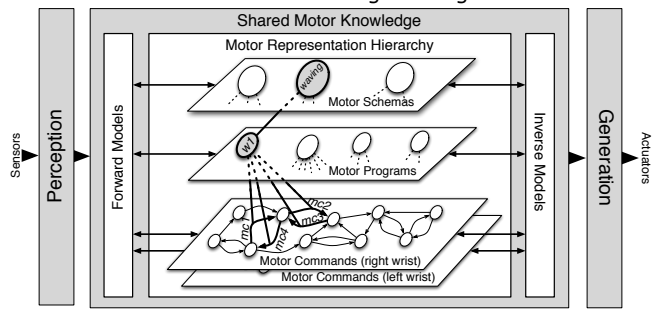

Figure 1: Overall model for embodied gesture perception and generation, integrated via shared motor knowledge. An example representation of a "waving" gesture is highlighted through bold lines and nodes.

Shared Motor Knowledge

The shared motor knowledge module consists of a pair of generic forward and inverse models and a hierarchical motor representation. At the lowest level, *motor commands* (in short, MC) represent spatiotemporal features of simple movement segments, arranged in a graph-like structure for each wrist (cf. *motor primitives*). At the next level, *motor programs* (MP) represent particular performances of a gesture as sequence(s) of motor commands. At the highest level, *motor schemas* (MS) cluster different performances of a gesture (i.e. MPs) and separate between invariant and variant features such as handedness or sub-movement repetition.

When the ECA observes a hand movement, all these represented motor components (MCs, MPs and MSs) serve as recognition hypotheses. At each time step, forward models evaluate those hypotheses against the observed movements, which results in a recognition confidence for each motor component. If the agent is not confident enough about observing any of the known motor component at one level, the perception process switches to inverse models that extend or adjust the motor knowledge to the newly observed movement. The same motor repertoire is, in turn, also used to perform gestures through a generation process in which probabilistic activation flows top-down to the level of executable MCs. That is, the ECA perceives hand movements in an embodied manner as he recognizes a movement by continuously comparing it with a motor repertoire that represents how the agent itself would perform that gesture.

Embodied Motor Resonances

The previously described perception process is realized in a probabilistic Bayesian framework. Following the Bayesian inference (see Figure 2), forward models resonate each motor component probabilistically w.r.t. the current observation. In order to make this embodied perception process robust, fast and incremental, we take three methodological steps: First, the motor resonance of each component ($m \in \{mc, mp, ms\}$) at each time step T is defined as its average activation over time: $P_T(m) := \frac{1}{T}\sum_{t=t_1}^{T} P_t(m)$. This step makes motor resonances incremental and robust against sensory noise. Second, at each time step t, the a priori of each motor component as a hypothesis is set to the previous a posteriori at time $t-1$, which supports incremental processing. Third, motor resonances are updated at each time step by two processes: (1) bottom-up belief propagation computes the a posteriori at each level given wrists observations and the a posteriori of the associated components at lower levels; (2) top-down belief guidance updates these probabilistic motor resonances by setting priors according to their dependence on higher level components. This cognitively plausible processing makes online recognition faster and more robust.

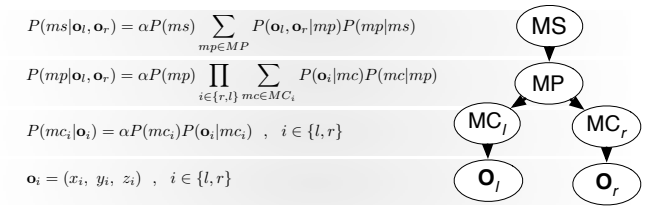

Figure 2: Probabilistic dependencies of motor resonances at different levels, given observations of left and right wrists ($\mathbf{o}_l$ and $\mathbf{o}_r$); α indicates the Bayesian normalizing constant.

Perception-Action Integration

To support the mutual effects between perception and generation processes, we define a *neural activation* for each motor component which is updated *and* used by both processes at each time step. On the one hand, the perception process sets the activations equal to the corresponding recognition probabilities. The generation process activates all generating motor components and the activations of all not-updated

components decrease. On the other hand, these neural activations are considered as prior probabilities while recognizing or generating gestures. In this way, we create perception-action links which account for different social capabilities and characteristics. For instance, this coupling enables direct, simultaneous imitation when the agent is set to perform motor resonances overtly. Furthermore, alignment and behavior coordination becomes possible because the ECA automatically tends to perform gestures which have been observed or generated last. Likewise, the ECA will tend to recognize previously self-generated or observed gestures.

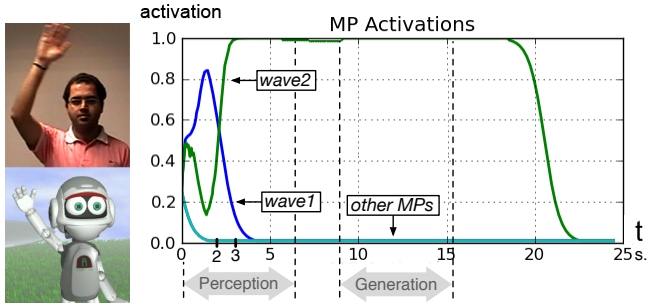

Figure 3: *Left*: Human user interacting with an ECA. *Right*: Evolving motor activations at the motor program level while, first, observing a known "waving" gesture ($wave2$), and then performing it in return (see [6] for more detailed results).

Conclusion

We have argued that embodied interactive agents like ECAs should be based more on principles of embodied cognitive processing to support many aspects of social interaction, from microscopic effects of behavior coordination to macroscopic abilities of imitation learning. We have presented a model that assumes a common sensorimotor structure and provides an embodied account of how to perceive, recognize, learn and generate hand gestures at the motor level. In this context, extending this model to higher representational levels that capture referential, communicative, and social intentions will be an important step for future work.

References

[1] A. Dijksterhuis and J. Bargh. The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, 33:1–40, 2001.

[2] A. Hamilton and S. Grafton. The motor hierarchy: From kinematics to goals and intentions. In R. Y., K. M., and H. P., editors, *Attention and Performance*. Oxford University Press, 2007.

[3] S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication, special issue on Speech and Face-to-Face Communication*, 52(6):587–597, 2010.

[4] E. Oztop, D. W. Franklin, T. Chaminade, and G. Cheng. Human-humanoid interaction: is a humanoid robot perceived as a human? *Humanoid Robotics*, 2(4):537–559, 2005.

[5] E. Oztop, M. Kawato, and M. Arbib. Mirror neurons and imitation: a computationally guided review. *Neural Networks*, 19(3):254–271, 2006.

[6] A. Sadeghipour and S. Kopp. Embodied gesture processing: Motor-based integration of perception and action in social artificial agents. *Cognitive Computation*, pages 1–17, 2010.

[7] M. Wilson and G. Knoblich. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3):460–473, 2005.

[8] J. M. Zacks. Using movement and intentions to understand simple events. *Cognitive Science*, 28(6):979–1008, 2004.