Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 32 (2012) 213 – 223

Procedia
Social and Behavioral Sciences

4th International Conference of Cognitive Science (ICCS 2011)

# Gesture processing as grounded motor cognition: Towards a computational model

Amir Sadeghipour[a,*] , Stefan Kopp[a]

[a]Cognitive Interaction Technology (CITEC), Bielefeld University, Tehran, Iran

## Abstract

In this paper, we present an approach to treat and model the processing (i.e. recognition and production) of communicative gestures as grounded motor cognition. We first review cognitive theories and neuropsychological studies on human motor cognition. On this basis, we propose a computational framework that connects the sensorimotor processing of hand gestures in representational structures of meaning (visuospatial imagery), other modalities (language), and communicative intentions. We present an implementation that enables an embodied virtual agent to engage in gesture-based interaction with a human user.

*Keywords*: Motor Cognition; embodiment; grounded cognition; gestures; social interaction; computational model; embodied conversational agents

## 1. Introduction

The interaction between artificial systems and humans is often envisioned to be carried out in the same way as humans communicate and interact. In this context, *embodied conversational agents* (in short, ECAs) are applied more and more as user interfaces. To this end, their underlying cognitive models need to support a wide range of social cognitive abilities like perceiving, recognizing, generating and learning of social signals. Brain-imaging studies show that humans interact with artificial agents (with sufficiently natural appearance and movements) in the same way and use the same bodily-grounded cognitive mechanisms as with humans (Oztop, Franklin, Chaminade, & Cheng, 2005). This entails *motor cognition,* defined to encompasse all processes involved in the production and comprehension of one's own and others' actions (Sommerville & Decety, 2006): processes such as planning, generating, performing one's own actions, as well as perceiving, recognizing, anticipating and understanding actions of others.

In this work, we aim for ECAs to be engaged in social interaction with humans and *like* humans. For this aim, we focus on hand gestures as a nonverbal communication means and propose a computational framework that realizes and implements the relevant mechanisms of *motor cognition* for communicative gestures. This requires to ground

[*] Corresponding author. Tel.: +49-521-10612141; fax: +0-000-000-0000
*E-mail address:* asadeghi@techfak.uni-bielefeld.de

doi:10.1016/j.sbspro.2012.01.032

these processes in structures of meaning and intentions as they are brought to bear in communicative gesturing. In the next section, we review psychological and neurobiological evidence as well as cognitive theories on motor cognition. Section 5 presents the computational framework in its current state of implementation. Results section presents some findings briefly, which are obtained with this implementation in a live interaction between an ECA and a human user.

## 2. Embodied Grounded Cognition

The traditional views of cognition assume a divergence between perception and cognition, so that cognitive representations are non-perceptual and separate from the brain's modal system. In contrast, *grounded cognition* claims that modal *simulation* underlies cognition (Barsalou, 2008). In this context, simulation is the reenactment of perceptual, motor and introspective states in the neural representation networks, which arises – and is acquired – during perceiving and acting (Goldman, 2006). Accordingly, *embodied cognition* refers to the assumption that cognition is grounded in the same neural representations that underlie perception, action and imagination (Hommel, Müsseler, Aschersleben, & Prinz, 2001). In this regard, embodied grounded cognition claims that cognition is grounded on such a shared sensory-motor representation. The assumption of shared representation is supported by many psychological and neurobiological studies. Next, we review evidence for this claim in three relevant modalities for gesture processing, namely: motor knowledge, visual knowledge and the language system.

Concerning motor knowledge, many behavioral social characteristics and capabilities of humans (such as alignment, priming, mimicry, imitation and emulation) are assumed to rest upon such a shared neural representation (e.g., Iacoboni, 2009). Loula, Prasad, Harber, and Shiffrar (2005) showed that one is more sensitive to one's own movements than to visually familiar movements produced by friends. Thus, perception of an action is constrained by one's knowledge of producing that action. The other way around, motor execution is affected by sensory signals. For example, the execution of finger movements is influenced by observing congruent or incongruent movements (Brass, Bekkering, & Prinz, 2001). The *mirror neuron system* is the strongest neurobiological evidence for such a shared representation. Mirror neurons respond to execution, observation and imagination of movements (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010; Montgomery, Isenberg, & Haxby, 2007). Such a shared neural network for perception and action has been also evidenced in auditory processing. Fadiga, Craighero, Buccino and Rizzolatti (2002) showed that listening to phonemes increases the excitability of the motor cortical area corresponding to the relevant tongue muscles.

In the context of embodied cognition, the activation of the motor system to the sight of an action is called *motor resonance* (or *motor simulation*, see Jeannerod, 2006). During perception, motor simulation is assumed to be available prior to the action event and thus it can anticipate the intended action effects (Hommel et al., 2001). This capability together with context information can allow humans to infer the intention behind a movement. Indeed, many studies claim for expanding motor simulation to the *simulation theory of mind* (e.g., Gallese & Goldman, 1998). Moreover, Gallese (2003) extends the concept of simulation from motor system to a basic functional mechanism used by vast parts of the brain.

Shared visual representations between perception and imagination also allow for a simulation capability in the visual modality. During visual perception, the outputs of sensory receptors activate primary visual areas. In higher levels, the observed object are recognized, categorized and stored in memory. Similarly, humans can rehearse this perceptual representation consciously in the form of visual mental images, which involves activation of the visual system, including low-level areas of processing (Kosslyn et al., 1993).

All in all, there is abundant evidence that the production, observation and processing of motor, visual, or verbal stimuli rests on shared mental structures.

## 3. Hierarchical knowledge representation

Neurobiological evidence suggests hierarchical neural coding in different modalities. For example, motor knowledge is assumed to be organized hierarchically, from representation of motor commands and low level kinematics, to more complex and conceptual representations, to goals and intentions behind movements (e.g., Hamilton & Grafton, 2007). Evidence for such a separation is provided by studies on apraxia patients, who have no problem in executing simple actions but fail in generating, imagining and recognizing actions involving more complex and conceptual representations (e.g., Buxbaum, Johnson-Frey, & Bartlett-Williams, 2005). Some other studies claim for a hierarchical representation of visual knowledge (e.g., Hochstein & Ahissar, 2002). In this context, when an entity is observed, it activates detectors in relevant feature maps. This ends in a representation as a global pattern of activation across a hierarchically organized structure.

## 4. Intra-modal and cross-modal associations

Concerning hierarchical sensory-motor representations, the question arises how these levels are interconnected and how they interact with each other. It is commonly assumed that a continuum of bottom-up and top-down processes utilizes these sensory-motor representations. It is the combination of both processes that coordinates the perception process (Barsalou, 1999). Many studies claim for such dual processing in motor knowledge (Csibra, 2007), visual knowledge (Gilden, Blake, & Hurst, 1995) and speech (Marslen-Wilson & Tyler, 1980).

Apart from the associations within a modality, evidence suggests that modalities are interconnected as well. With a focus on motor knowledge, we review some studies that suggest associations between this modality and the visual knowledge and language system. Associated representations of action and visual knowledge is best known through the mirror neuron system, which processes the visual stimuli of biologically relevant and plausible motions for humans, in addition to the perceptual visual system. Further evidence for such association can be found in studies dealing with *affordances* (Gibson, 1977). Further, strong associations between motor knowledge and the language system figure in the abundant gesturing that humans produce along with their speech (even when talking on the phone). A close integration of both modalities has been suggested for a long time in speech-gesture research (McNeill, 1992). Neurobiologically, Pulvermüller, Shtyrov, and Ilmoniemi (2005) explain such associations as a result of simultaneous activation of neural representations of motor programs and verbal constructions.

In this context, the *binding problem* deals with the question, how different representational elements are integrated into conceptual wholes through intra-modal and cross-modal associations. Damasio (1989) proposed so-called *convergence zones,* which can be thought of as neural regions that encode coincidental activities within or across modalities. McNorgan, Reid, and McRae (2011) recently provided evidence for a deep hierarchy of such convergence zones in a multimodal distributed *semantic* memory system.

## 5. The computational framework

On the basis of the aforementioned cognitive theories and evidence, we devised a computational framework for cognitively plausible gesture processing. We adopted a probabilistic approach for two main reasons. On the one hand, probabilistic approaches are more flexible for exploring human cognition (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010), reaching into causal and structured models. On the other hand, a growing body of evidence shows that human perceptual computations are Bayes' optimal. That is, the brain can be seen as functionally representing sensory information in the form of probabilities to deal with uncertainty in the form of degrees of beliefs (Oaksford & Chater, 2009).

## 6. The overall structure

Following theories of embodied grounded cognition, we associate two further modalities with motor knowledge that are relevant to communicative gestural movements: visual knowledge and the language system. Since the focus of this work is on motor knowledge, it supports the main required cognitive processes for gesture processing in social interaction: perception, recognition, generation and imitation learning. These processes were presented in previous work (Sadeghipour & Kopp, 2010). Here, we extend them along a continuum of motor cognition, from gestural motor capabilities to their meaning and intentions behind gestures. Since learning those modalities of visual knowledge and the language system is out of the focus of this work, we use predefined representations (i.e., not emergent or acquired) suitable for the interaction scenario in which our model is being tested and developed.

Figure 1 illustrates the conceptual design of the framework. It consists of four modules: perception, shared representation, generation, and deliberation. In the remainder of this section, we describe each module separately.

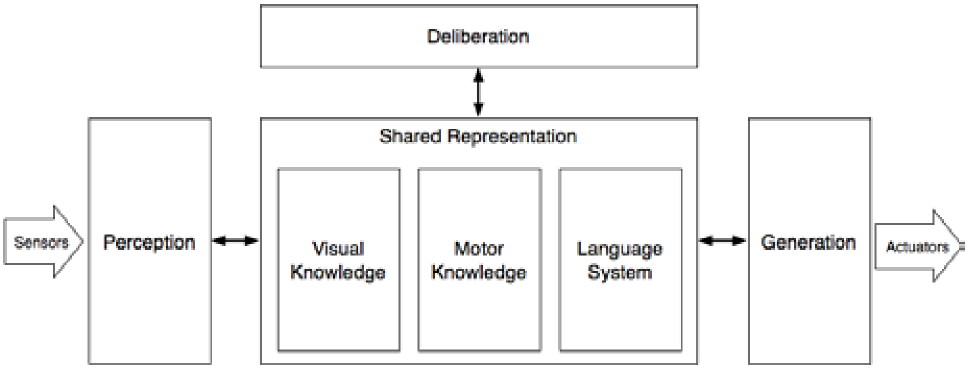


Figure 1. The overall structure of the computational framework

## 7. Shared representation

As previously mentioned, evidence suggests hierarchical and multimodal representation of knowledge, which is shared between processes of perception, generation and imagination. In our framework, the shared representation module comprises three modal knowledge representations, which are shared between – and interact with – all those processes: motor knowledge, visual knowledge and language system.

## 8. Motor knowledge

The hierarchical motor knowledge comprises three different levels of abstraction (see Figure 2). At the lowest level the *motor commands* (in short, MC) represent simple movement segments, which are specified by their spatiotemporal features (cf. *motor primitives* in the brain, e.g. see Mussa-Ivaldi & Solla, 2004). These segments are arranged as edges in a graph-like structure for each wrist. A path in this graph (i.e. a sequence of motor commands) composes at the next level a *motor program* (MP), which represents a particular performance of a gesture. At the most abstract level, each *motor schema* (MS) clusters different performances of a gesture (i.e. different MPs) together. A motor schema is a sparser representation of gestures, which discards potentially irrelevant features of a gesture, such as handedness or sub-movement repetition.

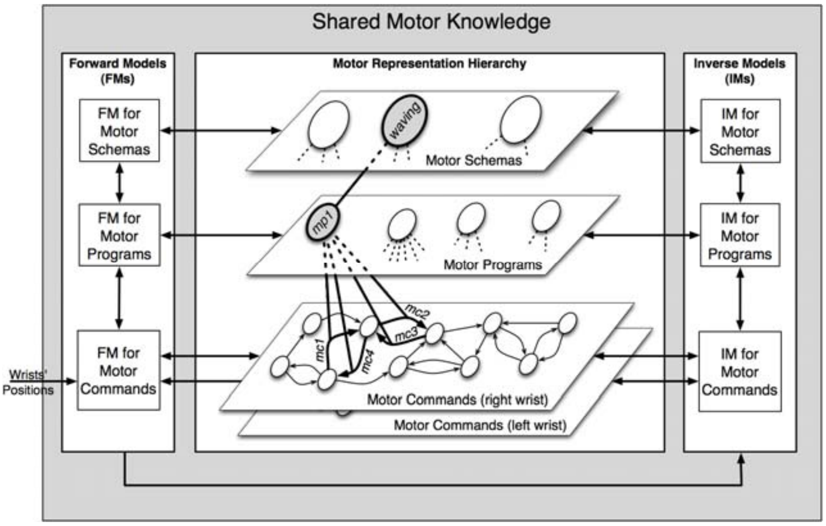


Figure 2. The hierarchical representation of motor knowledge. The representation of waving gesture is highlighted as an example

Apart from the knowledge representation, the motor knowledge module comprises two further submodules, which are used for recognition (forward models) and learning (inverse models). During observing hand movements, the forward models turn the represented motor structures at all levels into anticipation hypotheses, and compute a probabilistic confidence value for recognition (degree of belief) by comparing them against the current observations. If none of the represented motor structures at one level gains a sufficient confidence value, the perception process switches to the inverse models, which extend or adjust the motor knowledge to the newly observed movement (see Wolpert, Miall, & Kawato (1998) for neurobiological evidence on such processing).

## 9. Visual knowledge

The visual knowledge is added to ground the iconic gestures (i.e. shape-related) and emblems (i.e. conventionalized gestures) in a hierarchical visual representation of objects and events. Figure 3 illustrates the predefined representation for our interaction scenario. At the lowest level *entity schemas* represent humanoid agents and objects as compositions of their relevant parts. Iconic gestures can thus refer to an object, by referring to its part. A humanoid agent can be a virtual one or a human user. At the next level, entity schemas are generalized in terms of visual concepts and events. In this way, gestures can refer to events or even a category of entities. *Concepts* are clusters of entity schemas, which group different exemplars of an entity, with respect to their forms. *Events* consider the spatiotemporal variability of entity schemas. Finally, the highest level clusters related events and concepts into *event schemas*.

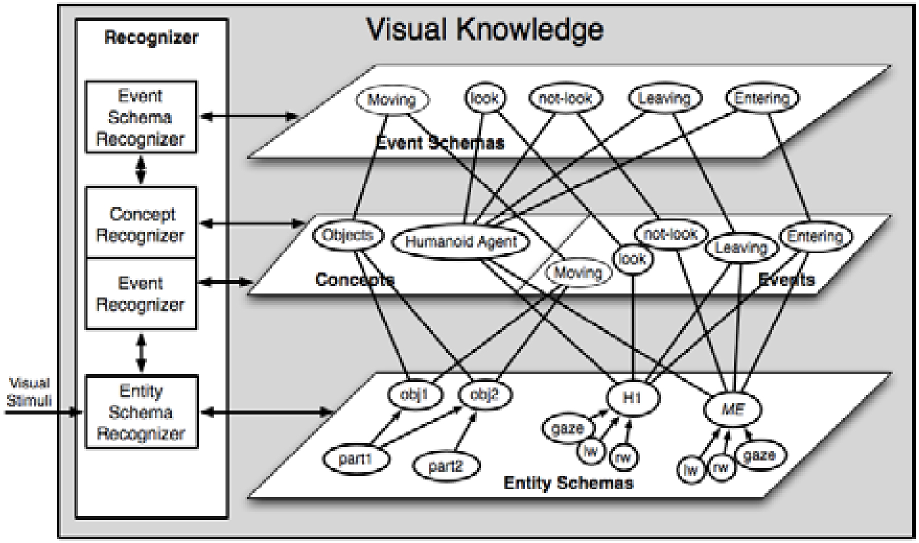


Figure 3. The hierarchical visual knowledge, including predefined representations for our interaction scenario

In our implementation, the representations of objects are specified with the aid of *Imaginary Descriptive Tree* (*IDT*) model (Sowa, 2006). This model is suggested for a unified semantic representation of shape, conveyed by speech and co-verbal gestures. A humanoid agent is represented as a composition of the relevant body parts for our scenario, namely spatial positions of wrists and gaze direction. Humanoid agents are associated with four events, which consider the variability of gaze direction and position of an agent. Objects are associated with the "Moving" event. Similar to forward models in the motor knowledge module, here, the recognizer submodule receives visual signals and computes a probabilistic confidence value of recognition for each of the representational elements in the hierarchy.

## 10. Language System

Gestures and speech are considered as two communication channels, which form an integrated cognitive system to utter intentions (McNeill, 1992). Therefore, the motor knowledge of gestures requires to be grounded additionally in language, and vise versa. Through such mutual associations, both communication means use the same knowledge, represented across the whole multimodal network.

The representation of speech can be a classical hierarchy from phonemes, to phrases, to grammatical constructions. However, as illustrated in Figure 4, in our interaction scenario we consider only the representation of speech at the level of phrases (see Kopp, Bergmann, & Wachsmuth, 2008) for a more complex language structure in association with IDT model). Similar to motor knowledge, this sparse representation of speech interacts with two submodules, which are responsible for recognition and acquisition of phrases (using a speech recognizer software).

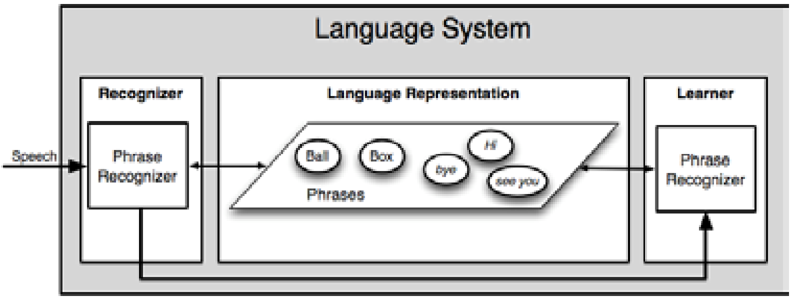


Figure 4. An example of a sparse representation within the language system

## 11. Intra-modal and cross-modal associations

In the overall framework, the associations between representational elements act as convergence zones, which exist both across and within different modalities. Associations are weighted connections, which allow the activation of each representational element to propagate. The cross-modal associations arise from coincidental activation of two representational elements in different modalities (see Figure 5). Following standard approaches, these associations change their weights proportional to the activation values of the connected representational elements (cf. *Hebbian learning theory*, Hebb, 1949).

## 12. Deliberation

Apart from motor cognition, there are many other cognitive capabilities that are required for a sociable agent. Most importantly, the agent has to be aware of social and communication norms (Tomasello, 2008), and the dialog course. In the current implementation of this framework, we realize this module as a state-machine dialog manager. Hence, this module can be replaced with any implementation of a dialog manager, which can be mathematically mapped onto a nondeterministic finite-state machine. It must be pointed out that in this way the Dialog Manager is a technically motivated module, and thus, it represents knowledge apart from the cognitively motivated shared representation module. However, as soon as the framework in endowed with the higher cognitive capabilities for participating in a dialog, this module should be modeled as a coherent extension of the shared knowledge.

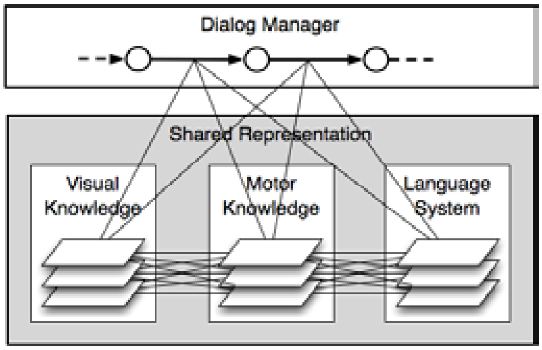


Figure 5. Cross-modal associations act as convergence zones that propagate neural activations across modalities

Each state in the dialog manager specifies the relevant aspects of the mental states of both the ECA and its interaction partner. Additionally, each state specifies the current state of the environment. All transitions represent actions or events which can change the state of the dialog. As illustrated in Figure 5, each transition is interpreted as a representational component, which can be associated with all other modality-specific components via cross-modal associations.

In this framework, an *interactive intention* is defined as the volition to change the mental state of the interlocutor, and/or informing the interlocutor about one's own change of mental state. Hence, transitions in the dialog manager specify intentions, which are associated with – and thus realizable through – the network of multimodal hierarchical representations.

## 13. Processes of perception, generation and imagery

Based on embodied cognition, the neural activation of representational elements can be the result of perception, generation and/or imagination. In this framework we apply probabilistic approaches, which compute the neural activations as probabilistic values. At each time step, we apply the Bayesian rule to update the current activation value (the *a priori* term) to a new activation value (*a posteriori*) with respect to the current evidence or belief (*likelihood*). If a representational element in any modality is not updated at one time step (e.g. because of irrelevant evidence), its activation decreases. Next, we describe how neural activations arise through those three processes.

## 14. Perception process

The perception module transmits the relevant stimuli to the corresponding modalities, after performing the required preprocessing steps. Using the received stimuli as evidence we apply the Bayesian rule, while all representational elements are considered as hypotheses. The computed posterior probabilities indicate the corresponding neural activations. Such perceptual activations are updated through three sequential steps: (1) bottom-up belief propagation, (2) cross-modal activation propagation, and (3) top-down belief guidance. Here, we focus on the motor knowledge as an example, and describe how these processes are carried out in our implementation.

Different levels of abstraction in the hierarchical motor knowledge can be represented as a Bayesian network (see Figure 6). While perceiving the wrists' movement observations, the forward models first activate (resonate) each motor command at the lowest level. Their activations are computed following the Bayesian rule. Accordingly, the likelihood term refers to the spatiotemporal divergence between the observation, and the motor commands that indicate how the ECA would perform that movement by itself. The probabilistic activation can be computed for each higher level, following the Bayesian inference. This bottom-up propagation of activations is equivalent to motor resonances.

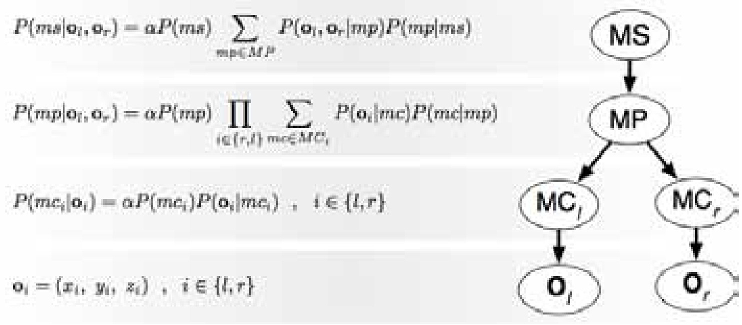

$P(ms|\mathbf{o}_l, \mathbf{o}_r) = \alpha P(ms) \sum_{mp \in MP} P(\mathbf{o}_l, \mathbf{o}_r|mp) P(mp|ms)$
$P(mp|\mathbf{o}_l, \mathbf{o}_r) = \alpha P(mp) \prod_{i \in \{r,l\}} \sum_{mc \in MC_i} P(\mathbf{o}_i|mc) P(mc|mp)$
$P(mc_i|\mathbf{o}_i) = \alpha P(mc_i) P(\mathbf{o}_i|mc_i) \ , \ i \in \{l, r\}$
$\mathbf{o}_i = (x_i, y_i, z_i) \ , \ i \in \{l, r\}$
MS
MP
$MC_l$
$MC_r$
$O_l$
$O_r$

Figure 6. The Bayesian network representing probabilistic dependencies along the hierarchy of motor knowledge

Similarly, during top-down belief guidance, the activation at each level is updated through the Bayesian rule, given the activation of the associated higher-level motor components as the likelihood term. Finally, the cross-modal activation propagation is carried out by an additional Bayesian belief update, which takes the associations' weights into account to compute the likelihood term. These cross-modal associations, ground perceiving of a gestural behavior on visual knowledge, language system and the dialog manager. In the latter case, the propagated activations resonate the associated transitions in the dialog manager, which consequently changes the active dialog state. In this way, the ECA uses these cross-modal associations to infer the interactive intention of an observed behavior, performed by an interlocutor. Behaviors can also be learned to be associated with specific interactive intentions. For instance, in our scenario, the ECA learns to wave when an interlocutor ends a dialog and leave the scene, since this behavior is perceived coincidental with the visual representation of "leaving" event, which is in turn associated with the transition of "ending a dialog". Additionally, hearing the phrases of "bye" or "see you" can contribute to the learning process.

Sequential performance of these three aforementioned perceptual steps updates the activation of representational elements at each time step. The frequency of each update step indicates the temporal synchrony between firing rates of neurons during perception. The more frequent one of these steps is performed, the more dominant is that process relative to the other ones.

## 15. Generation and imagery processes

The generation module uses the shared motor knowledge and language system to generate a behavior, which utters an intention in social context. When a transition in the dialog manager is assigned to a behavior to be performed by the ECA, the corresponding transition gets activated deliberately. Consequently, the activation propagates into the multimodal representation through the cross-modal associations. In this case, the ECA does not inhibit the resulted motor resonance (as yet set to do so by the dialog manager), and thus the most activated behaviors will be generated. For instance, in the case of generating a represented gesture in the motor knowledge, the activations flow top-down, from a motor schema to the level of executable motor commands. Similarly, activations can flow in the language system and activate some language structures that can be uttered by the ECA.

In cognitive studies, imagery is considered as voluntary activation of representational neural networks. Similarly, in this framework, an ECA can activate any representational element intentionally, but up to a certain degree. As a result, this activation propagates automatically via its associations through the shared representational network.

## 16. Simulation theory of mind for gestures

Comprehension of an action is one of the most challenging aspects of motor cognition. In the case of understanding communicative gestures, their meanings are contingent upon the intention behind them. The proposed implementation of the framework supports iconic gestures and emblems, which can refer to the represented events and objects, on the one hand. On the other hand, they can refer to intentions specified in the dialog manager. Furthermore, since the framework is based on the theories of grounded cognition, complex concepts can be comprehended through their representation within and across modalities. As illustrated in Figure 7, abstract concepts such as "valediction" can be – and have to be – grounded in several modalities.

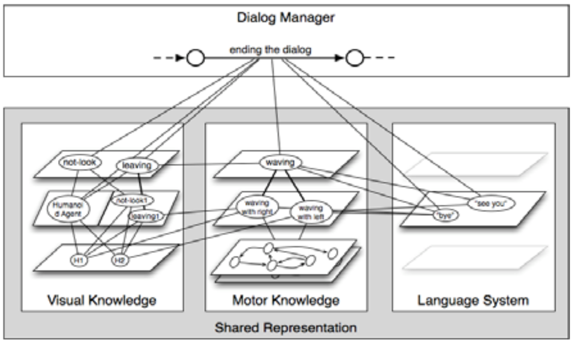


Figure 7. An exemplary representation of the concept of "valediction", which is grounded in several modalities

The shared knowledge representation allows for motor simulations that process actions irrespective of their actors. This support, together with the hierarchical knowledge representation, allows for simulations from the low level of motor commands up to high level of intentions and mental states. In this way, the framework supports a single form of a simulation theory of mind, by applying own knowledge, bodily and mental states as cognitive resources to simulate and understand the behavior of others and their intentions by grounding.

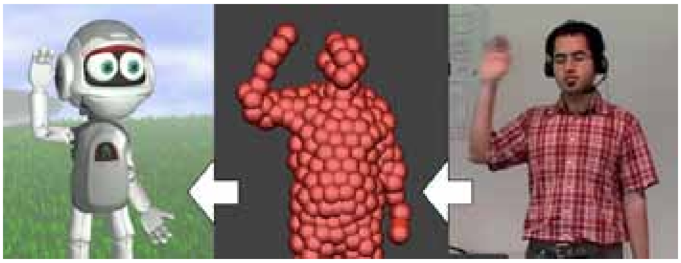
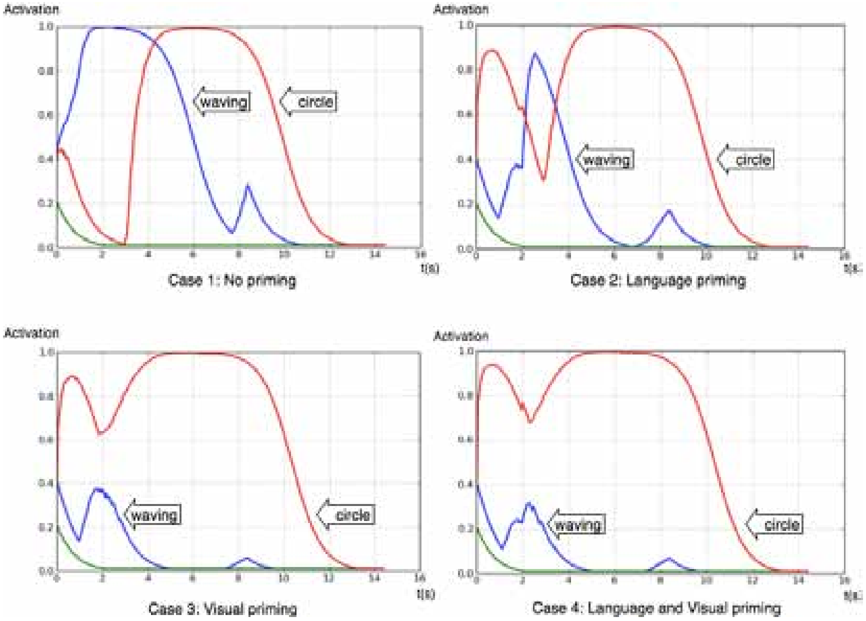


Figure 8. Capturing 3D spatial positions of wrists (among other relevant information) allows for direct imitation by the applied ECA

## 17. Results

We apply the proposed computational model, which is an adjusted implementation of this framework, in an interaction scenario between an ECA and a human. We applied a 3D time-of-flight camera (SwissRanger™ SR3000) and the marker-free tracking software Softkinetic iisu 2.0 to capture and recognize the relevant visual stimuli and events (see Figure 8). The head orientation is interpreted as gaze direction, leaving and entering the view field of the camera indicate the corresponding events. Additionally, humans are identified and the spatial positions of both wrists are captured.



Case 1: No priming

Case 2: Language priming

Case 3: Visual priming

Case 4: Language and Visual priming

Figure 9. The effects of perceiving the word "round" and/or observing the object "ball" on the recognition of a circle gesture

Here, we show the support for *priming* as a result of shared representation of sensory-motor knowledge. Priming is defined as the effect of a stimulus on the consequent stimuli processing. Based on shared knowledge representation between perception, imagery and generation; priming refers to a wide range of the mutual effects between these processes in different modalities. We consider a scene in our interaction scenario, in which the ECA is familiar with three different gestures: waving, drawing a circle and swinging the hand to refer to a horizontal surface.

Figure 9 shows how the verbal cue "round" and visual observation of a ball, which are associated with the circle gesture schema, facilitate (or prime) recognizing a particular performance of the circle gesture, which starts similar to a known waving gesture.

For more detailed results concerning the capabilities of the shared motor knowledge (such as recognition, imitation learning, and behavior coordination), please refer to our previously published results of this modality (Sadeghipour & Kopp, 2010).

Conclusion

In this paper, we have proposed a computational framework to model the use of communicative gestures as a form of grounded motor cognition, and presented an implementation for our ECA-human interaction scenario. The model is based upon principles of embodied and grounded cognition. It supports a wide range of social behavioral characteristics and capabilities with a focus on intransitive hand movements. In this context, social characteristics refer to unconscious phenomena, such as priming, alignment and mimicry, while social capabilities refer to more conscious cognitive processes, such as recognition and inferring intention, deliberation, emulation and imitation learning. Future work will explore more specifically algorithms and approaches to implement the presented framework.

Acknowledgement

This research is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence ‘‘Cognitive Interaction Technology’’.

## References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*, 617-645.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*, 3-22.

Buxbaum, L. J., Johnson-Frey, S. H., & Bartlett-Williams, M. (2005). Deficient internal models for planning hand-object interactions in apraxia. *Neuropsychologia*, *43*, 917-929.

Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In P. Haggard, Y. Rosetti & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*, 123-132.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*, 399-402.

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 517-528.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*, 493-501.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale. NJ: Lawrence Erlbaum Association.

Gilden, D., Blake, R., & Hurst, G. (1995). Neural adaptation of imaginary visual motion. *Cognitive Psychology*, *28*, 1-16.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York: Oxford University Press.

Griffiths, T., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*, 357-364.

Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: From kinematics to goals and intentions. In P. Haggard, Y. Rosetti & M. Kawato (Eds.), *Sensorimotor foundations if higher cognition* (pp. 381-407). Oxford: Oxford University Press.

Hebb, D. (1949). *The organization of behavior.* New York: Wiley & Sons.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791-804.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849-878.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, *60*, 653-670.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self.* Oxford: Oxford University Press.

Kopp, S., Bergmann, K., & Wachsmuth, I. (2008). Multimodal communication from multimodal thinking - towards an integrated model of speech and gesture production. *International Journal* of *Semantic Computing*, *2*, 115-136.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, *5*, 263-287.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 210-220.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1-71.

McNeill, D. (1992). *Hand and mind: What gestures.* Chicago: University of Chicago Press.

McNorgan, C., Reid, J., & McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition*, *118*, 211-233.

Montgomery, K. J., Isenberg, N., & Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Social Cognitive and Affective Neuroscience*, *2*, 114-122.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, *20*, 750-756.

Mussa-Ivaldi, F., & Solla, S. (2004). Neural primitives for motion control. *IEEE Journal of Oceanic Engineering*, *29*, 640-650.

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences*, *32*, 69-84.

Oztop, E., Franklin, D. W., Chaminade, T., & Cheng, G. (2005). Human-humanoid interaction: Is a humanoid robot perceived as a human? *International Journal of Humanoid Robotics, 2*, 537-559.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*, 884-892.

Sadeghipour, A., & Kopp, S. (2010). Embodied gesture processing: Motor-based integration of perception and action in social artificial agents. *Cognitive Computation*, 1-17.

Sommerville, J. A., & Decety, J. (2006). Weaving the fabric of social interaction: Articulating developmental psychology and cognitive neuroscience in the domain of motor cognition. *Psychonomic Bulletin and Review*, *13*, 179-200.

Sowa, T. (2006). Towards the integration of shape-related information in 3-D gestures and speech. *Proceedings of the 8th international conference on Multimodal interfaces, USA,* 92-99.

Tomasello, M. (2008). *Origins of human communication.* Cambridge: MIT Press.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2*, 338-347.