A Probabilistic Model of Motor Resonance for Embodied Gesture Perception

Amir Sadeghipour and Stefan Kopp

Sociable Agents Group, Cognitive Interaction Technology (CITEC), Bielefeld University
P.O. 100131, D-33501 Bielefeld, Germany
{asadeghi,skopp}@techfak.uni-bielefeld.de

Abstract. Basic communication and coordination mechanisms of human social interaction are assumed to be mediated by perception-action links. These links ground the observation and understanding of others in one's own action generation system, as evidenced by immediate motor resonances to perceived behavior. We present a model to endow virtual embodied agents with similar properties of embodied perception. With a focus of hand-arm gesture, the model comprises hierarchical levels of motor representation (commands, programs, schemas) that are employed and start to resonate probabilistically to visual stimuli of a demonstrated movement. The model is described and evaluation results are provided.

1 Introduction & Background

In social interactions, we are continuously confronted with a variety of nonverbal behaviors, like hand-arm or facial gestures. The same holds true for intelligent virtual agents that are increasingly employed in interfaces where they are to engage in similar face-to-face interactions. Consequently, they are ultimately required to perceive and produce nonverbal behavior in a fast, robust, and "socially resonant" manner, i.e. based on an understanding of and entrainment with what the other intends, means, and how she behaves. In humans, this capability is supposed to be rooted in an embodied basis of communication and intersubjectivity. Many studies (e.g. [2,15,3]) have demonstrated that the motor and action (premotor) system become activated during the observation of bodily behavior. The resulting *motor resonance* is assumed to be due to perception-action links and to emerge at various levels of the hierarchical human perceptuomotor system, from kinematic features to motor commands to goals [11]. These resonances allow for imitating or mimicking the observed behavior, either overtly or covertly, and thus form a basis for understanding other embodied agents [22]. In addition they foster coordinating with others, e.g., in mimicry or alignment, in order to establish social resonance and rapport (see Fig. 1 for illustration).

As evidenced by brain imaging studies [18,16], an animated interlocutor with sufficiently natural appearance and motion can – to a certain extent – evoke in humans similar motor resonances. However, behavior perception and understanding on the part of the artificial interlocutor is usually treated as pattern

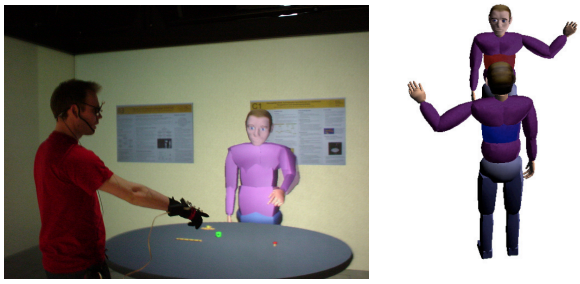

Fig. 1. Interacting agents engaging in embodied perception and behavior matching.

classification focused on trajectory recognition rather than intention recognition. Many approaches employ probabilistic methods with convenient properties like graceful degradation, processing of uncertainty, or learning schemes. Calinon and Billard [6] apply Hidden Markov Models to recognize gestures after applying PCA and ICA in order to decorrelate, denoise and reduce the dimensionality of data. Further work [5] applies Gaussian mixture models to provide a more accurate modeling of uncertainty. However, the classification of movements is based on spatio-temporal feature correlations and does not aim at the abstraction into the intention or meaning of a gesture. Some recent approaches [21,20]) apply Bayesian inference to derive the goal of a movement, defined as a spatial configuration. Hierarchical probabilistic models were proposed [1] for temporally grouping motor primitives into sequences. However, co-speech gestures are meant to transfer information to the addressee and different, spatio-temporally uncorrelated movements can be employed for this inter-changeably. Thus more abstract levels are eventually necessary for capturing a gesture's intention. None of the techniques applied so far has attempted to tightly link perception and action in motor resonances, which should enable fast and incremental embodied gesture perception, across different levels of abstraction. In the effort to endow IVAs with increasing capabilities of social interactivity, we present a probabilistic approach to model the automatic emergence of motor resonances when embodied agents come to observe another agent's hand-arm gestures. In the following Section 2 we introduce the overall computational model, and we present an probabilistic approach to simulating motor resonance in Section 3. In Section 4 we present results of applying this model to real-world gesture data.

2 A Model for Embodied Gesture Perception

In previous work, we proposed an approach to learning motor acts of hand-arm gestures by imitation, built atop a model for procedural gesture animation [12,14]. It has been developed in a scenario with two virtual humans of identical embodiment, one demonstrator and one learner and imitator. In the present work, we extend this model in two ways to allow for resonance-based gesture per-

ception. First, as motivated above, we add more abstract and less contextualized motor levels in order to work hierarchically from reception toward understanding. Second, we add a probabilistic method for how these hierarchical structures can be utilized for behavior perception by starting concurrently and incrementally to resonate when observing a gesture. Finally, the framework is connected to a marker-free 3D camera to enable embodied human-agent interaction (Sect.4). Overall, the model consists of four modules (see Fig. 2): preprocessing, motor knowledge, forward models, and inverse models. We will describe them here briefly; details can be found in [12,14].

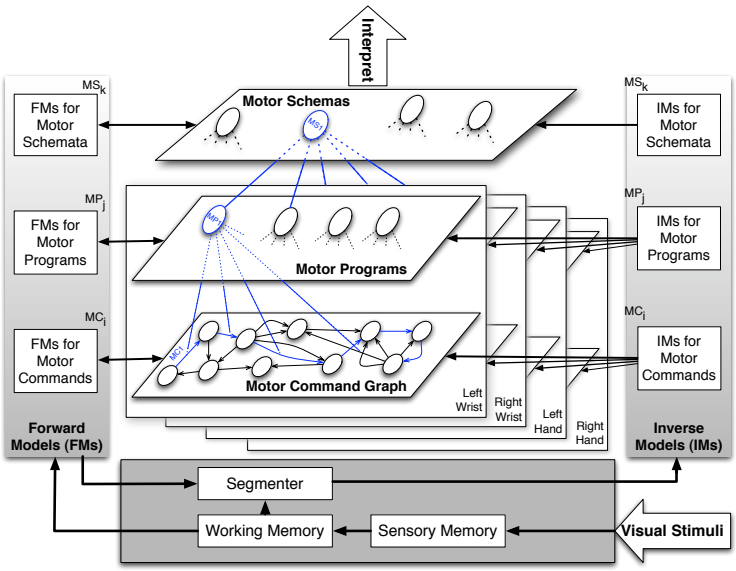

Fig. 2. Overall model for resonance-based embodied gesture perception.

2.1 Preprocessing

The preprocessing module receives visual stimuli about the movement of relevant body parts of a demonstrator (positions and orientations of the wrist, fingers of both hands) and buffer them in chronological order in a *working memory*. A *segmenter* then decomposes the received movement of each body part into sub-movements, based on its kinematic features (velocity profile, direction changes). For example, the movements of a wrist in space are decomposed into spatio-temporal segments, called guiding strokes [13]. The movements of the fingers are decomposed into key postures of the hand. A guiding stroke represents a spatial movement segment in 3D space and, suitably parametrized, describes the movement trajectory and the performance speed. Since the focus in this paper

is on intransitive actions, all parameters attributed to the segments refer to the morphological features of the movement, and they are not defined relative to an object. Such parametrized segments are atomic movement components (called submovements) of each body part and a sequence of them represents a gesture.

2.2 Motor knowledge

A directed graph is used to store the motor knowledge about gestural movements as a sequence of its submovements. Edges of the graph stand for movement segments; nodes represent the intermediate states of the corresponding body part. That is, edges for the wrists' spatio-temporal movements are assigned the proper guiding strokes and each node represents a spatial position. In the case of hand and finger configuration, each node represents a keyframe (hand posture) and the edges indicate the transition (parametrized with a velocity profile) to reach the next hand posture. In that way, a movement becomes a sequence of edges, i.e. a path in the graph (see [4] for similar modeling approach). A novel gesture can be added as new path in the graphs and, if necessary, it may add new edges and nodes to the graph. When performing a gesture, the agent should follow the corresponding path in the graph and perform its edges sequentially. Therefore, each edge in the graph, independent from the related body part, represents a motor command (MC), and a path in the graph stands for a motor program (MP). Neurobiological studies showed that the human brain uses a similar principle of decomposing complex movements into simpler elements, called motor primitives, and generates them (in performing phase) in parallel and sequence [9,10]. Modeling internal motor representations of each body part separately is also consistent with the somatotopic organization found in motor cortex.

Due to the fact that MCs for each body part have their own features and parameterization, they are stored separately in specialized knowledge submodules, called motor command graph (MCG) and motor program graph (MPG), respectively. The MPG is a more compact representation of the MCG and clusters each motor sequence as a single node. In this way, the agent has an exact representation of the individual gestures in its own repertoire. However, in general, gestures are not limited to a specific performance but have some variable features. These are the parameters of the performance which, when varied, do not change the meaning and intention of the gesture but the way of performing it. Consequently, understanding a gesture can *not only* involve an exact motor simulation (direct matching), but also infernal of communicative meaning. For example, seeing a demonstrator waving should be recognized by an imitator as the act of waving, independent of the absolute spatial position of the wrist joint, the swinging frequency or to some degree the speed of the movement. Although different persons have different styles of waving, all those performances can be classified by an observer to the same meaning. And, when reciprocating, the observer likewise performs an individual way of doing it. Thus, embodied agents must be able to cluster numerous forms of a gestural movement into one schema, which ignores the variable features of the gesture, e.g. spatial position, number of repetitions, etc. Therefore, we define motor schemas (MS) as a generalized

representation that groups different allowed performances (motor programs) of a gesture, possibly performed through many body parts, into a single cluster. Analogously, a motor schema graph (MSG) consist of motor schemas as nodes. Such a generalization process is an important capability and can foster the understanding and imitation of behavior in two ways. First, it forwards the problem of inferring the goal of a gesture from the motor level to a more abstract, yet less complex level, namely schema interpretation. Second, an imitator can retain his own personal form of performing a gesture, while being able to relate other performances of the same gesture to the same schema.

2.3 Forward and inverse models

An agent may follow two routes to imitate an observed gesture. On the one hand, it can recognize a movement as familiar, i.e. approximately similar to an act in the own repertoire. In this case, the agent can perform an *active imitation*, activating the motor system during the perception process. One the other hand, when the model is not in the observer's repertoire, the agent perceives the new movement and analyze it afterwards, drawing upon the motor knowledge it has acquired before. In result, new motor knowledge about the novel movement is created, inserted into the internal motor representation, and can then be executed for imitation. This process is called *passive imitation* [7].

In our model, active and passive imitation are modeled with forward and inverse models, respectively. Forward models are predictors derived from the agent's motor knowledge in order to predict the continuation of a familiar gesture at each motor level. By comparing this prediction with the actual percepts at each time step, these models are to find the motor command, program or schema that most likely correspond to the observed behavior. If there is no sufficiently similar representation, the analyses switch from the forward models to inverse models, which receive their input from the segmenter and turn it into parametrized submovements. These submovements are used to augment the MCG, MPG or MSG, if necessary, with new nodes and edges. Performing the newly acquired act, then, accomplishes the modeling of *true imitation.*

3 Probabilistic Motor Resonances

The basic mechanism of perceiving a gesture (either for imitation or understanding) is to compare the predictions of the forward models, derived for possible candidate motor structures, with the observed movement of the other. This basic mechanism is employed at all three levels (by different kinds of forward models) and results in motor resonances that represent the agent's confidence about the correspondence between what it sees the other doing and what it "knows" from itself. Given the visual stimuli about moving relevant body parts of the other as the only evidence, we define this confidence in recognizing a certain motor candidate as the mean over time of its a posteriori probability given the evidences at each time step (cf. eq. 1). This can also be considered as a kind of *expectation*

value of the respective motor candidate. We apply the prior feedback approach (see [17]) to accumulate the expectation up to each time step. This also enables the use of Bayesian networks to model how the motor levels interact in order to allow resonances to percolate bottom-up and top-down in between them, to find (possibly a variant of) a known gesture fast, effectively and robustly. Furthermore, in this way, the probabilities of motor candidates of different lengths are comparable. In the following we focus on the perception of hand position or trajectory; finger movements can be modeled analogously.

3.1 Level 1: Resonating motor commands

At this level, the spatial positions of a wrist at each time step t are our evidences and the motor commands in the MCG are the hypotheses. Since our approach should work incrementally and in real time, the more evidences we have the higher the recognition confidence should be. The probability of a hypothesis equals the resonance or expectation of the corresponding motor command c on basis of all perceived evidences ($\mathbf{o} = \{\mathbf{o}_{t_1}, \mathbf{o}_{t_2}, ...\}$) up to the current time step, T. Employing the Bayesian law, we have:

$$P_T(c \in H_c) = P_T(c|\mathbf{o}) := \frac{1}{T}\sum_{t=t_1}^{T} P(c|\mathbf{o}_t) = \frac{1}{T}\sum_{t=t_1}^{T} \alpha_c P_{T-1}(c) P(\mathbf{o}_t|c) \quad (1)$$

The term $P_{T-1}(c)$ is the a priori of the hypothesis c and indicates the previous knowledge about the probability of motor command c, which is equal to the expectation of c at the previous time step, $T-1$. In the case of $T = t_1$, the a priori will by default be the uniform distribution across all alternative motor commands outgoing from the same parent node. The likelihood term $P(\mathbf{o}_t|c)$ refers to the probability of passing the coordinate $\mathbf{o}_t = \{x_t, y_t, z_t\}$ with motor command c and, now, represents a probabilistic prediction of the forward model. In other words, it represents the probability of where the hand would be if the agent now performed the motor command c. We model this as a four dimensional Gaussian probability density function of $\{x, y, z, t\}$ (PDF, in short), which is formed for each possible next motor command, i.e., each possible continuing submovement of the wrist in space (see Fig. 3). Each likelihood reaches its maximum value if the observed performance exactly matches the own motor execution in both spatial and velocity features.

Let H_c be the set of currently active ("resonating") motor command hypotheses. The criterion to add a motor command into this set is as follows. As soon as the first evidence, $\mathbf{o}_{t_1}$, is perceived, its probability to represent a node of the MCG is computed with the aid of Gaussian densities centered at the 3D position of each node. Comparing with a predefined threshold yields the most likely candidate nodes for the starting point of a gesture (or not). All outgoing motor commands from these nodes are added to H_c. At the next time steps, the probability of each of these hypothesis is computed from the next evidence (eq. (1)). If the probability of a hypothesis is smaller than a predefined threshold, it will be omitted from H_c. Note that the resonance of each motor command varies

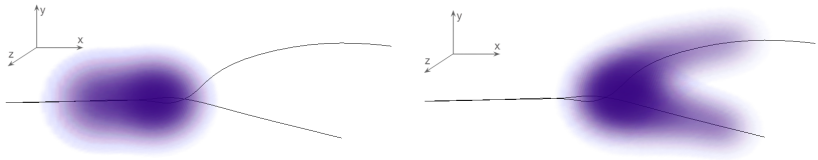

Fig. 3. Visualization of the likelihood of motor command hypotheses, models as (4D) Gaussian density functions that change over time in accord with the motor command and its corresponding velocity.

with the duration of its execution: the longer the performance takes, the more evidences are used to update the expectation of that motor command. That is to say, the confidence of the imitator in the computed probability of each motor command increases.

3.2 Level 2: Resonating motor programs

The probability (or resonance) of a motor program p, which is represented as a path in the MCG and as a node in the MPG, depends on the probabilities of its components (motor commands) and thus, indirectly, on the evidences $\mathbf{o}_t$. We compute this probability, similar to motor commands, as an expectation of p considering all evidences until the current time step, T.

$$P_T(p \in H_p) = P_T(p|C, \mathbf{o}) := \frac{1}{T} \sum_{t=t_1}^{T} P(p|C, \mathbf{o}_t) = \\ \frac{1}{T} \sum_{t=t_1}^{T} \alpha_p P_{T-1}(p) \sum_{c \in H_c} P(\mathbf{o}_t|c) P_t(c|p) \tag{2}$$

The a priori term is equal to the expectation at the previous time step. The term $P_t(c|p)$ indicates the probability of performing the command c at time t, if the demonstrator were to perform the program p. This probability is time-dependent and is modeled using a PDF as a function of t and the motor commands c. The mean of the Gaussian moves through the motor commands of a motor program, as fast as the velocity of each motor command. Thus, this term along with $P(\mathbf{o}_t|c)$ together yield high resonance to the observation $\mathbf{o}_t$ of the right position at the right time step with respect to p.

The set of motor programs considered as hypotheses H_p is defined to contain all programs with at least on active motor command in H_c. Motor programs with too small expectations will be removed from the set. At each point in time, the computed expectation for each motor program refers to the confidence of the agent in recognizing that gesture for which, in contrast to the MCG, not only a submovement but the morphological properties of the whole gesture performance are considered. Note, however, that these probabilities are incrementally computed and adjusted from the evidence at hand, also during the perception while

only parts of the gesture have been observed yet. That is, the agent does not need to specify the start and end point of gestures, but can even recognize gestures that are started at a later point of a trajectory, e.g., in the case of performing several gestures successively without moving the hand to rest position.

3.3 Level 3: Resonating motor schemata

The top level of motor knowledge consists of motor schemas, represented in MSG, which group different motor programs for different body parts into a single node. The expectancy (resonance) of each motor schema depends on the expectation values of active motor programs in all body part modules, and indirectly on the related motor commands and evidences about each body part. Figure 4 illustrates these causal influences between the graph nodes in a hierarchical Bayesian network.

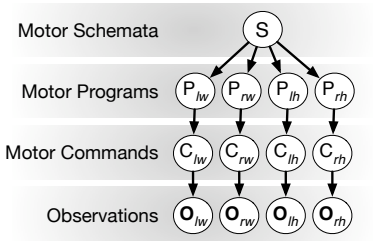

Fig. 4. Bayesian network of the relations between different levels of the motor hierarchy.

The expectation of each schema s is computed as follows:

$$P_T(s \in H_s) = P_T(s|\mathbf{C}, \mathbf{P}, \mathbf{o}_{lw}, \mathbf{o}_{rw}, \mathbf{o}_{lf}, \mathbf{o}_{rf}) := \frac{1}{T} \sum_{t=t_1}^{T} P(s|\mathbf{C}, \mathbf{P}, \mathbf{o}_{lw}, \mathbf{o}_{rw}, \mathbf{o}_{lf}, \mathbf{o}_{rf})$$

$$= \frac{1}{T} \sum_{t=t_1}^{T} \alpha_s P_{T-1}(s) \prod_{i \in \{rw,lw,rh,lh\}} \sum_{p_i \in H_{p,i}} P(p_i|s) \sum_{c_i \in H_{s,i}} P(\mathbf{o}_{i,t}|c_i) P_t(c_i|p_i) \tag{3}$$

The likelihood $P(p_i|s)$ is uniformly distributed among the $p_i \in s$, and 0 otherwise. Because of the OR relation among the associated motor programs ($p_i \in H_{p,i}$), the probability of a motor schema s is the sum of the probabilities of its possible performances. However, we may also consider schemes that are more tolerant to velocity, position or direction deviations than discrete paths in the MCG. For one thing, such allowed deviations should be set by the motor schema that has a view to the goal of the gesture and can differentiate between acceptable and unacceptable deviations context-sensitively. In addition, in this way we avoid rapid extension of the MCG and MCP, which are brought about

by the inverse model analysis when expectations run too low.

But how to define what's a waving gesture and what's not? Or, in other words, how to define the invariants and variants in a motor schema? We allow four different possible variations in performing gestures, which we can map onto the model's structure in order to define parameters for the related motor schema: (i) velocity variability, (ii) position variability, (iii) repetition of a submovement, and (iv) left and right hand performance.

Velocity variability: Many hand-arm gestures can, within certain limits, be performed with different speed without altering the intention being the gesture, e.g., showing the victory sign or pointing somewhere. In order to recognize such variants as instances of the same gesture, the motor commands should deliver same expectations in all cases. One argument of the Gaussian likelihood $P(\mathbf{o}|c)$ for each motor command is time. Hence, its variance σ_t defines the tolerance of the motor command c to variations in performance speed. By increasing the value of this parameter of the likelihood model through the motor schema, we decrease the tolerance of the corresponding schema regarding the performance velocity during the perception process.

Position variability: The spatial position of a gesture often does not decisively affect its meaning. In order to avoid creating too many motor commands and programs for different performances of the same gesture, special prototype nodes in the MCG and MPG are created as the invariant structures of a motor schema, while leaving the variant features open. Each evidence position $\mathbf{o}$ is normalized to the corresponding position in this prototype as given by the distance between the start positions: $\Delta\mathbf{o} = \mathbf{o}_{template} - \mathbf{o}_{perceived}$. That is, when starting to perceive a gesture, all prototype nodes as well as the matching normal nodes in the MCG are considered as start point candidates.

Repetition of a submovement: Some gestures comprise repetitive parts, like waving or beat gestures, and the number of repetitions is often subject to considerable variation. Such repetitions correspond to cycles in the MCG that start and end at nodes which represent branching points for such a schema (one more cycle or continue otherwise). This can be handled straightforward, by splitting the PDFs that model the likelihoods $P(\mathbf{o}_t|c)$ and $P_t(c|p)$ into distributions that covers both expectations. The expectation of the corresponding motor schema then equals performing one of the alternatives, i.e., the sum of the expectations.

Left and right hand performance: A gesture should be recognizable as the same schema, regardless of the hand it is performed with. Since a motor schema comprises motor programs for both hands, it can specify how their probabilities affect the expectation of the schema. In the normal case (3), all body parts are assumed to have their own task during performance. Nevertheless, the way of combining different body parts is not always an AND relation ($\prod_{i\in\{rw,lw,rh,lh\}}$) but sometimes an OR relation, like in this case. Therefore, each motor schema specifies the way of combining the body parts depending on the gesture.

In order to be able to cover all aforementioned variations, each motor schema has following parameters: (1) the means and variances of the Gaussian PDFs of

all comprised motor commands; (2) a flag to specify if the schema is a template for position variable gestures; (3) a set of all cycles in the graph; (4) flags indicating the causal relation (AND or OR) between body parts.

3.4 Horizontal and vertical integration

The described probabilistic model simulates the bottom-up emergence of motor resonances, where the expectations at each level induce expectations at higher levels. The other way around, higher levels should also affect and guide the perception process at lower levels. For instance, after recognizing a motor schema the agent should expect to perceive the remaining movements over the next time steps. That is, the expectation of a motor command should increase the expectation of subsequent motor commands from the same gesture. In our framework, this capability can be mediated via the higher levels: The computed expectation of a motor program determines the a priori knowledge in computing the expectation of next motor commands. To this end, we update the a priori of the future motor commands, $c \in p$, using the Bayesian rule $P(c|p) = \alpha P(p|c)P(c)$, where $P(c)$ indicates the *previous* a priori of c. Likewise, a "resonating" motor schema affects the expectation of its comprised motor programs. Overall, every time new evidence arrives, we not only percolate expectations about active hypotheses up, but also adjust the prior probabilities of current or future hypotheses top-down in a context-dependent way. To this end, the a priori probabilities are calculated both from default priors and expectations during the last time step, as well as new a priori knowledge coming from higher levels. This vertical interaction of motor levels occurs continuously; see Sect. 4 for a simulation of this.

Horizontal integration refers to how forward model-based perception and inverse model-based learning interact. Switching from the former to the latter is controlled by continuously comparing the current likelihoods with predefined rejection thresholds. That is, as long as the MCG can predict the observed movement, and as long as the MPG can predict the resonating motor commands, the agent remains in perception mode. Beyond the scope of this paper, we briefly mention that the other mode, i.e. acquiring motor structures that then can resonate to observed behavior, is a crucial problem for embodied agents. Our model comprises the inverse models (Fig. 2, right-hand side) to analyze a demonstrated behavior for new motor commands and programs, which are then inserted into the graphs and can be tested and refined in subsequent imitation games [12,13]. The learning of motor schemas likewise can only succeed in social contexts, where repeated demonstration-imitation interactions with informative feedback guide the learner in finding the schema boundaries. While these acquisition processes are subject of ongoing work, we note that the model presented here directly enables behavior generation (internally or overtly) and, thus, imitation. This can be mediated by each of the three levels. For example, when starting from the highest level, the agent chooses a motor schema and then selects those comprised programs or commands with the highest priors, which encode how often the agent has observed the corresponding performance for that motor schema. In other words, the imitator tends to act in the way observed (and imitated) most

often. The schema-specific parameters for the motor commands indicate the velocity and position changes for movements. The agent takes the mean values of these parameters and simply executes the correspondingly set motor commands.

Horizontal integration also refers to how behavior perception and generation in an embodied agent come to interact because they both employ identical motor structures. One direct consequence is that the behavioral tendencies of the agent are affected by its perceptions. In our model, the a priori for each motor representation in MCG, MPG or MSG is defined by default, depending on the number of alternative hypotheses. However, during observing and perceiving a gesture as described above, the a priori probabilities that match the observation are increased as an effect of the top-down propagations. We do not reset these priors directly after perception to their default values, but let them decline following a sigmoidal descent towards the default values. As a result, when producing gestures the agent tends to favor those schemas, programs, and motor commands that have been perceived last (cf. [8]).

The other way around, our model also allows to simulate so-called "perceptual resonance" ([19]), which refers to the opposite effect of action on perception. Since we use the same a priori probabilities for both generation and perception processes, we can model this phenomenon by simply increasing the a priori probabilities of *generated* motor commands, programs and schemas temporarily. This will bias the agent's gesture perception toward the self-generated behavior, which has been suggested to be another mechanism for coordination in social interaction.

4 Results

We have implemented the proposed model for resonance-based gesture perception and evaluated it with real-world gesture data. In a setup with a 3D time-of-flight camera (SwissRanger^{TM}SR3000[1]) and the marker-free tracking software iisu[2], the agent observes the hand movement of a user during several performances of three different gestures: waving, pointing upwards, and drawing a circle. These gestures are familiar to the agent and we report how the present motor structures resonate, i.e., how the confidences of the alternative hypotheses evolve *during* perceiving a gesture. All gestures have been started at the same position increasing the agent's uncertainty as to which gesture is performed. Figure 5 (top-left) shows the agent's MCG, which corresponds to the spatial arrangement of the corresponding motor commands. This MCG is generated during learning by the corresponding inverse model after segmentation. The overlaid dashed-line shows the trajectory of a demonstrated waving gesture. The other subfigures show how the expectancies, i.e., resonances, of different motor commands (top-right), motor programs (bottom-left), and motor schemas (bottom-right) evolve.

[1] http://www.mesa-imaging.ch
[2] http://www.softkinetic.net

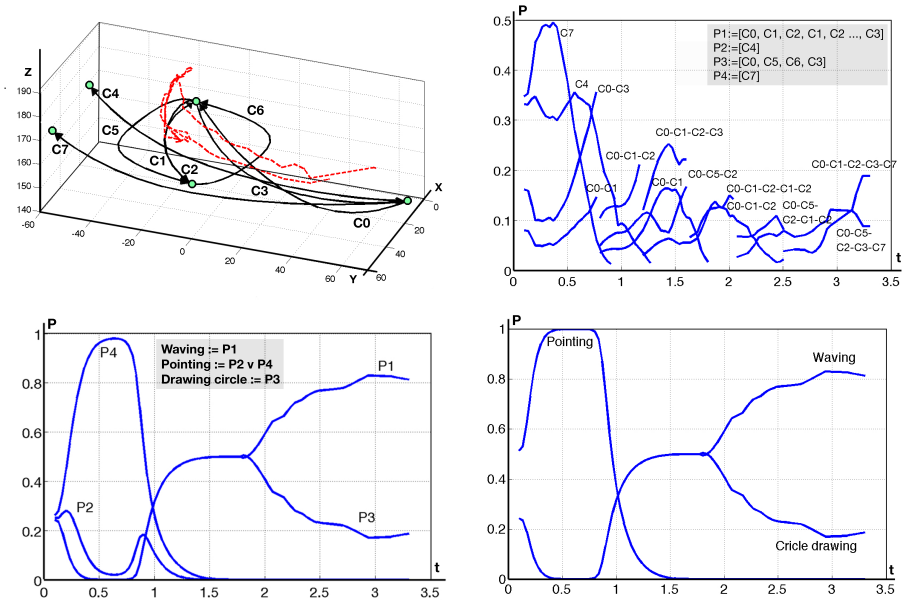

Fig. 5. Evaluation results: *top-left:* motor command graph with observed trajectory overlaid (dashed line); *top-right/bottom-left/bottom-right:* changing probabilities of the hypotheses currently entertained on the three motor levels (*commands/programs/schemas*).

The motor commands in the MCG imply hypotheses about how familiar gestures would proceed. While perceiving the demonstrated gesture, new hypotheses are generated, old hypotheses are extended, and unlikely ones are omitted. At each time step, one hypothesis corresponds to the most expected movement segment. Depending on the number of hypotheses, the maximum expectation value changes over time and the winner threshold is adopted respectively. Figure 5 (top-right) shows a subset of the active motor commands hypotheses. The first winning hypothesis indicates that the observed movement starts similar to a pointing gesture, c_7. Therefore, the agent thinks that the user is going to point upwards (p_4). However, after one second the user starts to turn his hand to the right and, thus, the resonance of the motor commands c_1 and c_5 increases. Consequently, the other gestures (p_1 and p_3) attain higher expectancies but the agent still cannot be sure whether the user is going to draw a circle or wave to him. After about two seconds, the agent perceives a swinging movement, which is significantly similar to the waving gestures known to the agent. In result, the agent associate the whole movement to waving schema and, e.g., could start to perform a simultaneous imitation.

5 Conclusion

In this paper we have described a probabilistic model for simulating motor resonances and, thus, perception-action links in the processing of non-verbal behavior. Based on a hierarchy of graph-based representations of motor knowledge, our model enables an embodied agent to immediately start to "resonate" to familiar aspects of gestural behavior, from kinematic features of movement segments (modeled through motor commands) to complete movements (motor programs) to more general prototype representations (motor schemas) that cover possible variants in a gesture's performance. The hierarchical motor structures of the agent are employed to realize two proposed key components of embodied gesture perception, horizontal and vertical processing. The former refers to prediction-evaluation schemes to figure out on each level which command, program, or schema matches best an observed behavior; the latter refers to the bottom-up and top-down flow of activation, which affords concurrent and incremental abstraction and recognition of the perceived stimulus. The probabilistic model proposed here implements these fundamental processes in an integrated way. In this view, resonance of a particular motor unit is broken down to the expectancy of its effects (if it were executed) given the evidence at hand (what has been observed so far), given current activations of the connected motor structures. Resonance thereby results from a Bayesian inference, in which we take not only the conditional probabilities to change depending on what arrives bottom-up, but also adjust the priors continuously in accordance to predictions and biases that flow top-down. Evaluations with simulated and real-word data (gesture trajectories) showed this approach's potential for fast and incremental perception–two properties indispensable for smooth social interaction. Building upon the perceptual and motor representations employed in the agent architecture thus paves the way for engaging in social behavior in a more human-like way, including automatic coordination effects like motor mimicry, imitation, or alignment.

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

References

1. R. Amit and M. Mataric. Learning movement sequences from demonstration. In *ICDL '02: Proceedings of the 2nd International Conference on Development and Learning*, pages 203–208, 2002.
2. M. Brass, H. Bekkering, and W. Prinz. Movement observation affects movement execution in a simple response task. *Acta Psychologica*, 106(1–2):3–22, 2001.
3. G. Buccino, F. Binkofski, G. R. Fink, L. Fadiga, L. Fogassi, V. Gallese, R. J. Seitz, K. Zilles, G. Rizzolatti, and H.-J. Freund. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13:400–404, 2001.

4. D. Buchsbaum and B. Blumberg. Imitation as a first step to social learning in synthetic characters: a graph-based approach. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 9–18, New York, NY, 2005.
5. S. Calinon and A. Billard. Incremental learning of gestures by imitation in a humanoid robot. In *HRI '07: Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 255–262, New York, NY, USA, 2007. ACM.
6. S. Calinon and A. Billard. *Learning of Gestures by Imitation in a Humanoid Robot*, pages 153–177. Cambridge University Press, 2007.
7. J. Demiris and G. R. Hayes. Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model. In *Imitation in animals and artifacts*, pages 327–261. MIT Press, Cambridge, MA, 2002.
8. 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.
9. T. Flash and B. Hochner. Motor primitives in vertebrates and invertebrates. *Journal of Current Opinion in Neurobiololgy*, 15:660–666, 2005.
10. G.-F. Gutemberg and A. Yiannis. A language for human action. *Computer*, 40(5):42–51, 2007.
11. A. Hamilton and S. Grafton. The motor hierarchy: From kinematics to goals and intentions. In *Attention and Performance 22*. Oxford University Press, 2007.
12. S. Kopp and O. Graeser. Imitation learning and response facilitation in embodied agents. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents*, pages 28–41, Marina Del Rey, CA, 2006.
13. S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents. *Journal of Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.
14. S. Kopp, I. Wachsmuth, J. Bonaiuto, and M. Arbib. Imitation in embodied communication – from monkey mirror neurons to artificial humans. In I. Wachsmuth, M. Lenzen, and G. Knoblich, editors, *Embodied Communication in Humans and Machines*, pages 357–390. Oxford University Press, Oxford, 2008.
15. G. P. L. Fadiga, L. Fogassi and G. Rizzolatti. Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73(6):2608–2611, 1995.
16. E. Oztop, T. Chaminade, and D. Franklin. Human-humanoid interaction: is a humanoid robot perceived as a human? *Humanoid Robots, 2004 4th IEEE/RAS International Conference on*, 2:830–841, 2004.
17. C. P. Robert. Prior feedback: A Bayesian approach to maximum likelihood estimation. In *Technical Report 91-49C*, 1991.
18. L. Schilbach, A. M. Wohlschlaeger, N. C. Kraemer, A. Newen, N. J. Shah, G. R. Fink, and K. Vogeley. Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, 44(5):718–730, 2006.
19. S. Schutz-Bosbach and W. Prinz. Perceptual resonance: action-induced modulation of perception. *Journal of Trends in Cognitive Sciences*, 11(8):349–355, 2007.
20. A. Shon, J. Storz, and R. Rao. Towards a real-time bayesian imitation system for a humanoid robot. *Robotics and Automation, 2007 IEEE International Conference on*, pages 2847–2852, 2007.
21. R. Verma, D. Rao. Goal-based imitation as probabilistic inference over graphical models. *Advances in neural information processing systems*, (18):1393–1400, 2006.
22. M. Wilson and G. Knoblich. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3):460–473, 2005.