A Neurocomputational Model of an Imitation Deficit following Brain Lesion

Biljana Petreska and Aude G. Billard

Learning Algorithms and Systems Laboratory (LASA)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Station 9, CH-1015 Lausanne, Switzerland
biljana.petreska@epfl.ch, aude.billard@epfl.ch

Abstract. This paper investigates the neural mechanisms of *visuo-motor imitation* in humans through convergent evidence from neuroscience. In particular, we consider a deficit in imitation following *callosal brain lesion*, based on the rational that looking at how imitation is impaired can unveil its underlying neural principles. We ground the functional architecture and information flow of our model in brain imaging studies and use findings from monkey brain neurophysiological studies to drive the choice of implementation of our processing modules. Our neural model of visuo-motor imitation is based on self-organizing maps with associated activities. Patterns of impairment of the model, realized by adding uncertainty in the transfer of information between the networks, account for the scores found in a clinical examination of imitation [1]. The model also allows several interesting predictions.

1   Introduction

Apraxia is generally defined as the inability to perform voluntary movements that cannot be explained by elementary motor, sensory or cognitive deficits (not caused by weakness, ataxia, akinesia, deafferentation, inattention to commands or poor comprehension). A standard test for clinical examinations of apraxia is *imitation of meaningless gestures* which is believed to test the integrity of a direct route from visual perception to motor control, not mediated by semantic representations or verbal concepts [2]. Goldenberg has shown that knowledge about body parts is also relevant, as apraxic patients were unable to map body configurations to their own body nor to a mannikin [3]. Kinematic studies of apraxia show spatial parapraxias (i.e., normal kinematic profiles with abnormal final positions) that seem to arise from a basic deficit that concerns the mental representation of the target position [4].

Goldenberg's study. A seminal study of imitation of meaningless gestures examines a patient with callosal brain lesion (disconnected hemispheres) [1]. The patient was asked to imitate a set of visual stimuli that present different positions of the hand relative to the head (see Fig. 1). To disentangle the contribution of each hemisphere the patient was tested tachistoscopically (i. e., the stimulus was

presented either to the left or right visual field) in a left- or right-hand imitation condition. As shown on the figure (upper right) the pattern of errors varies as a function of the visual field to which the stimuli were displayed and the hand used to execute the imitative movement. The schema in Figure 1 shows the hypothesized non-uniform information flow across the two hemispheres in the different conditions, related to regions in the brain based on brain imaging and lesion studies [5–7]. The stimulus is visually processed in the hemisphere contralateral to the visual field (due to optic chiasm) and the motor command is prepared in the hemisphere contralateral to the hand. The arrows show the necessary transfer of information between the two hemispheres, thus a possible source of spatial errors in the imitation (as the patient suffers from disconnected hemispheres). Imitation was perfect only in the right visual field - right hand condition, indicating a lateralization of the processing to the left hemisphere and a necessary computational process in the brain area shown in dark grey.
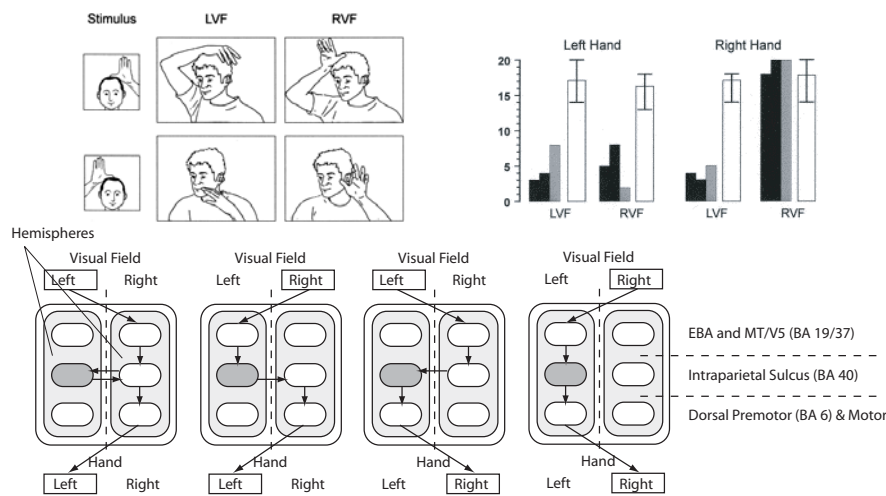

Fig. 1. Upper left *Goldenberg's experiment* of imitation of meaningless gestures, an example of a visual stimulus to imitate and the errors made by the patient. Upper right the patient's score of success in the four conditions (several trials, in white control data), taken from [1]. In the lower part, *schema of information flow* through the left and right hemispheres of the brain in the four conditions, see the text for explanation.

2 Neurocomputational Model of Imitation

In this paper, we investigate impaired imitation of meaningless gestures, namely hand postures relative to the head as the one shown in Figure 1. This work follows from a general effort in our group to decipher the neural mechanisms of

visuo-motor imitation [8, 9]. In order to model the behavioral data reported in Goldenberg's study, we developed a neural network architecture that accounts for the transformations required to translate the observation of the visual stimulus to imitate to the corresponding tactile and proprioceptive information that will guide the imitative gesture. We simulate a callosal lesion by impairing the transfer of information between the networks and observe the occurrence of spatial parapraxias. Next, we describe the model.

2.1 Description of the model

The model is composed of three neural networks, see Fig. 2: a *face visual network* in Brodmann Area BA 19/37 at the level of the occipito-temporal junction, a *face somatic network* in area BA 40 in the parietal cortex and a *hand position network* probably in dorsal premotor area BA 6. As it is the case in imitation of meaningless gestures we have implemented a visuo-motor route mediated by somatic knowledge of body parts. The face visual network receives geometrical properties of the visual stimulus to imitate (such as the position and angle of the hand relative to the nose, see Fig. 2. The face somatic network receives input from the face visual network and somatic input from tactile sensors of the face. The hand position network receives visuo-somatic input from the face somatic network and proprioceptive input from the arm. The neurons in our model are leaky integrator neurons in order to account for variations of the membrane potential in time and to have integrating properties.

Face visual network The face visual network encodes geometrical properties of the stimulus to imitate. The network receives the two-dimensional input $\boldsymbol{x^H}$ composed of the distance $d^H \in \mathbb{R}[0,9]$ and angle $\phi^H \in \mathbb{R}[0,2\pi]$ of the hand relative the nose (shown on Fig. 2). We decided to use these two properties as they univocally define the stimulus to imitate and are quantities easy to process visually. It is certain that the brain uses also other quantities when imitating a hand posture relative to the head (position relative to the eye may be more appropriate in some cases), however we decided to limit the number of visual properties for simplicity. It was important that the visual and somatic networks rely on completely different representations.

The membrane potential m_i of the visual neuron i is governed by a first order differential equation modulated by a gaussian input:

$$\tau^V \frac{d}{dt} m_i^V = -m_i^V + \mathrm{e}^{-\left(\frac{|\boldsymbol{w_i^H} - \boldsymbol{x^H}|}{2\sigma_V^2}\right)} \tag{1}$$

where τ^V is a time constant, $\boldsymbol{w_i^H}$ are the synaptic weights that connect the neuron i to the input $\boldsymbol{x^H} = \{d^H, \phi^H\}$ and σ_V corresponds to the "sensitivity" of the neuron to the input (a neuron with a large σ_V responds to a larger interval of inputs values).

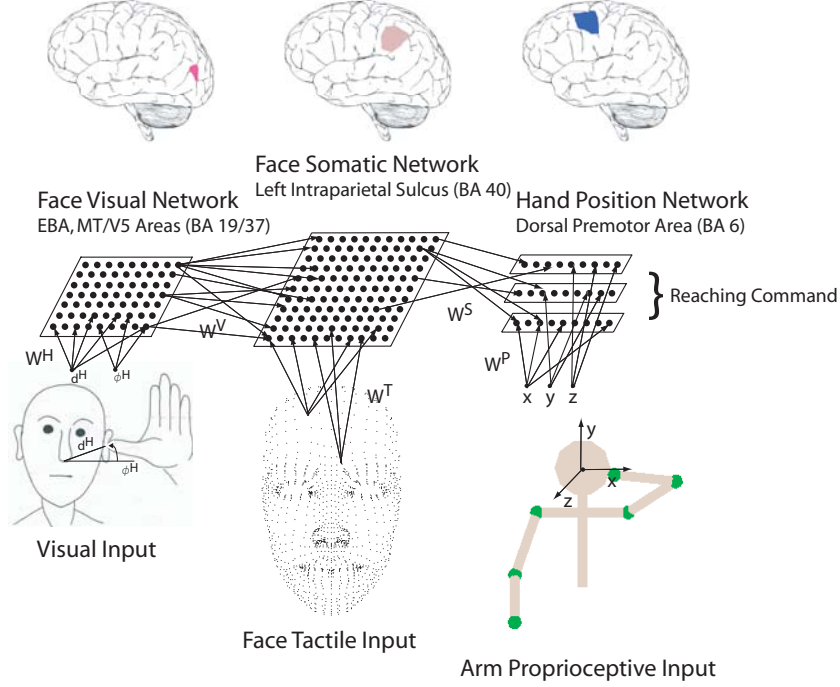The firing rate is a sigmoid function of the membrane potential with slope a and offset b:

Fig. 2. Schema of the neurocomputational model. The model is composed of three neural networks that receive visual, tactile and proprioceptive input: a *face visual network* that corresponds to Brodmann Area BA 19/37 at the level of the occipito-temporal junction, a *face somatic network* that corresponds to area BA 40 in the parietal cortex and a *hand position network* in dorsal premotor area BA 6.

$$g(m_i^V) = \frac{1}{(1 + e^{a(-m_i^V + b)})} \tag{2}$$

Face somatic Network The face somatic network is a somatotopically organized network principally processing tactile information from the face. It receives input $\boldsymbol{x^T} \in \mathbb{R}^{N^T}[0, 1]$ from $N^T = 1500$ tactile sensors non-uniformly distributed on the face (with a preponderant number of sensors around the eyes, nose and mouth). It also receives visual input from the face visual network described previously. The membrane potential m_j^S of a somatic neuron with index j is equal to:

$$\tau^S \frac{d}{dt} m_j^S = -m_j^S + \sum_{k=1}^{N^T} w_{jk}^T e^{\left(-\frac{|\boldsymbol{x^P} - \boldsymbol{r_k^T}|}{2\sigma_T^2}\right)} + \sum_{i=1}^{N^V} w_{ji}^V g(m_i^V) \tag{3}$$

where τ^S is a time constant, w_{ik}^T is the synaptic weight of the neuron to the tactile sensor with index k and w_{ij}^V is the synaptic weight to the visual neuron

with index i, N^T and N^V are the numbers of tactile sensors and visual neurons respectively, $r_k^T \in \mathbb{R}^3$ is the position of the tactile sensor k in space, $x^P \in \mathbb{R}^3$ is the center position of the hand-face contact and σ_T is the width of the contact. Note that the face somatic network integrates inputs of different types, namely somatic input from the tactile sensors and visual input preprocessed by the face visual network.

Three layers of the *hand position network* encode proprioceptive information from the arm. Each layer encodes a different coordinate of the position of contact $x^P \in \mathbb{R}^3$ of the hand and the face, expressed in head-centered cartesian coordinates. Our motivations were the following: there is no "real" proprioceptive information from the face and we hypothesized that this information could be learned from correlations between the face tactile sensory activity and arm proprioceptive activity during reaching movements toward the face. A "positional code" may well be used in the brain where different coordinates are processed in segregated neural substrates, possibly in Cartesian coordinates [10]. The frame of reference is centered in the head to maximize the invariance of the positions of the tactile sensors (which would not be the case in a body-centered frame of reference because of the rotation of the head).

The neurons in the hand position network each have a preferred coordinate value c_k, preferred values were uniformly distributed in a volume that contains the head $\mathbb{R}^3$[-8,8]. The membrane potential m_k^P integrates over the proprioceptive input x^P and the visuo-somatic input $g(m_j^S)$ from the face somatic network (the vectorial notation expresses the three layers of the hand position network):

$$\tau^P \frac{d}{dt} m_k^P = -m_k^P + \mathrm{e}^{-\left(\frac{(x^P - c_k)^2}{2\sigma_P^2}\right)} + \sum_{j=1}^{N^S} w_{kj}^S g(m_j^S) \qquad (4)$$

where w_{kj}^S are the weights between a somatic neuron j and a position neuron k, σ_P is the width of the receptive field of the position neuron and N^S is the number of neurons in the face somatic network. The activation function is the same as in equation 2. The output of the hand position network is decoded using a weighted average of N^P (number of position neurons) firing rates which corresponds to the position p on the face:

$$p = \frac{\sum_{k=1}^{N^P} c_k g(m_k^P)}{\sum_{k=1}^{N^P} g(m_k^P)} \qquad (5)$$

The decoded activity of the hand position network is used as a target for the imitation of a visual stimulus.

2.2 Training the weights

The synaptic weights between the networks and their sensory inputs (i. e., weights W^H between the face visual network and the extracted visual parameters and weights W^T between the face somatic network and the face tactile

input) have been trained with Kohonen's algorithm [11]. Thus our networks are self-organizing maps (SOM) whose weights preserve the topology of the input. The unsupervised learning algorithm consists of randomly choosing a sensory input $\boldsymbol{x}$ and determining the "winning neuron" with index j^* whose weights are closest to the input. It then updates the synaptic weights of the "winning" neuron and neurons in its neighborhood by the following rule:

$$\Delta \boldsymbol{w_i}(j^*) = \epsilon \cdot \mathrm{e}^{-\frac{|i-j^*|}{2\sigma_K^2}} \left[\boldsymbol{x} - \boldsymbol{w_i}\right] \tag{6}$$

where ϵ is the learning rate, $\boldsymbol{w_i}$ are the synaptic weights of the neuron with index i and σ_K corresponds to the size of the neighborhood[1]. In the end stimuli close in the input space are also close in the 2D neural space and more frequent inputs yield larger neural activities.

The synaptic weights between the networks (i. e., weights W^V between the visual and the face somatic network and weights W^S between the face somatic and hand position network) were trained with a presynaptic gating anti-hebbian learning rule:

$$\Delta w_{i,j} = \eta \cdot x_j [2 \sum w_{i,k} x_k - m_i] \tag{7}$$

where $w_{i,j}$ is the synaptic weight between a presynaptic neuron x_j and a postsynaptic neuron with membrane potential m_i and η is the learning rate. The learning process associates correlated activities of two networks. The connecting weights learn a mapping between the neural activity of one input and one output network for a given stimulus. In other words the weights organize in order to have the sensory activity in the input network represent the sensory activity in the output network. Both W^V and W^S were trained during the same process of self-observation, which simulates sensory input during reaching movements toward the face in front of a mirror. For example, the activity in the face visual network is associated to the somatic activity due to touching the face and is associated with a position in space through proprioceptive information from the arm. In the end presentation of the visual stimulus to imitate alone yields the corresponding neural activities in the face somatic and position networks thus guiding a correct imitative action. The values used for the parameters of the model were selected by trial and error and are shown in Table 1[2].

2.3 Simulation of the lesion

For simulating the lesion of the corpus callosum (i. e., impaired transfer of information across the two hemispheres) we have taken into account two obser-

---

[1] The weights were initialized with random values between 0 and 1 and the parameters ϵ and σ_K were decreased at each step according to the functions in Table 1.

[2] The inputs selected in the learning processes form a random uniform distribution in the input space. For a faster convergence all the time constants were set to 1. The Kohonen algorithm was run 100 times for the face visual network, 9000 times for the face somatic network and the anti-hebbian learning process was iterated 5000 times.

Table 1. Parameter values

$N^V = 400$	$\tau^V = 35ms$	$\sigma_V = 0.6$	$\sigma_{VK} = 8$	$l^V = 0.97$	$\epsilon^V = 1$	$n^V = 0.98$
$N^S = 1225$	$\tau^S = 35ms$	$\sigma_S = 0.3$	$\sigma_{SK} = 22$	$l^S = 0.9996$	$\epsilon^S = 1$	$n^S = 0.99999$
$N^P = 3\text{x}100$	$\tau^P = 35ms$	$\sigma_P = 0.3$				
$a = 15$	$b = 0.5$		$\sigma = l\sigma$		$\epsilon = n\epsilon$	$\eta = -0.02$

vations. First, some of the visual information must cross the callosum since the patient succeeds to imitate some hand positions when he/she visually processes the stimulus in one hemisphere and prepares the motor command in the other hemisphere. Second, interestingly enough, time is a very important variable. If the patient was given "unlimited time" he/she imitated correctly [12]. To model the observation that some of the information crosses, we introduce a probability of information transfer ρ . The impairing function is either applied at the level of the connection (model 1) or at the level of the input of the neuron (model 2). To model the improvement of the patient's performance with time we hypothesized an integrating factor greater than the decay factor. We added a constant $\lambda \in \mathbb{R}[0,1]$, which slows down the membrane decay. The dynamics of the membrane potential m of one neuron for the two models is then expressed by:

$$1) \ \tau \frac{d}{dt}m = -\lambda m + Wf(I) \quad 2) \ \tau \frac{d}{dt}m = -\lambda m + f(WI) \quad m < f(WI) \Rightarrow \lambda = 0$$

(8)

where W is the weights matrix, I is the membrane input and f is the impairment function such as $f(x) = x$ with probability ρ and $f(x) = 0$ otherwise, see Figure 3. Therefore even if the neuron receives bits of information from time to time, the membrane potential is no more precisely tuned to the input but continues to integrate. As the face somatotopic network is situated in the left parietal cortex,we impair the connecting weights W^V in condition "left visual field" and the weights W^S in condition "left hand".

3   Results

To analyze the performance of our impairment models we have trained the weights once, then quantified the spatial parapraxias as the distance E between the desired end-target position $\boldsymbol{r}$ and the position $\boldsymbol{p}$ computed from the hand position network under different patterns of impairment[3].

A property of the model is to always converge to the right response given unlimited time no matter how impaired the transfer of information is, as long as

---

[3] For a simpler analysis of the results we have impaired all the connections equally, but our implementation allows variations of the percentage of impairment or location and size of the lesion.
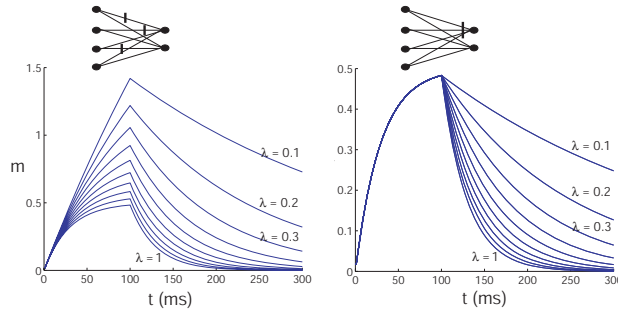
Fig. 3. The dynamics of the membrane potential of one neuron in model 1 which impairs the connection (on the left) and in model 2 where the impairment occurs at the neuron's input (on the right). Input $I = 0.5$ was applied during 100ms and τ was set to 30ms.

some information does transfer ($\rho > 0$) and λ is small. As you can see in Figure 4 even for a probability of information transfer as small as $\rho = 0.1$ at the level of a single connection, the model converges to the correct position over time given a sufficiently small λ (0.1 in model 1 and 0.03 in model 2). The presence of λ deteriorates the performance in the unimpaired situation ($\rho = 1$) in model 1 (see Fig. 4) as the neuron membrane "overintegrates" in the first model, as shown on Figure 3. Another drawback is that small values of λ render adaptation to a novel stimulus slower. However a longer decay time presents the advantage of having a "fading memory" of the stimulus, the stimulus remains represented in the brain after the presentation time, which is compatible with the occurrence of perseveration errors observed in experimental studies. Several predictions can be made on the basis of these models. With severe lesions, the patient needs more time to do a correct imitation, shown in Figure 4. It suggests that it is possible to obtain a measure of severity of the lesion based on the time needed by the patient to do the imitation. Small λ values would enable a correct processing even at very high impairment rates, but would depreciate the reaction time.

We compared the results of the simulations to the scores in Goldenberg's study with some adaptation. As we consider only the end-target spatial errors and not errors in the hand posture (such as orientation of the hand or finger configuration), we took the upper bound of the score used in the study (2 points for a correct imitation). We replicated the same experimental conditions (i. e., same visual stimuli, 180ms of stimulus presentation and weights impairment coherent with the four conditions as described in the study). A set of values could explain the scores in the Goldenberg study, as shown in Figure 5. The second model shows slightly better results, however this was not significant.

The representation of parts of the face in the "face somatic" network is non-uniform, some face parts such as the eyes or the mouth are overrepresented in contrast with the cheek or the chin. This is due to the non-uniform distribution of the tactile sensors. Therefore we observe inhomogeneities in the precision of
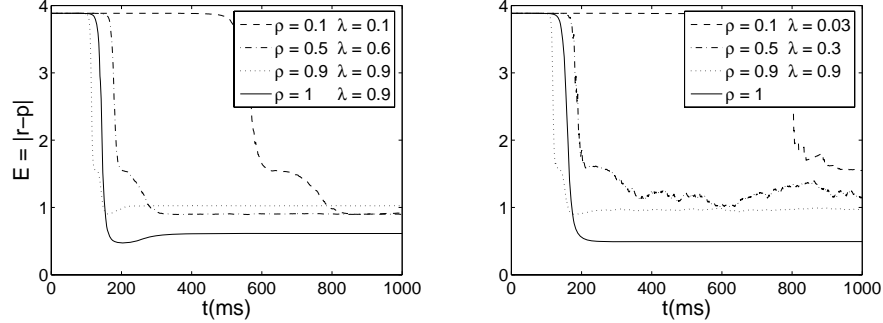
Fig. 4. The error in imitation computed as the distance between the desired and simulated end-target position for different values of ρ and λ according to model 1 (on the left) and model 2 (on the right). The observation that the patient required 180ms of visual stimulus presentation time to be able to imitate motivated the choice for the parameters of the activation function and of τ ($\tau = 35$). The starting position is the same throughout the trials and corresponds to the origin of the head-centered coordinate axis. We observe that more severe lesions necessitate longer processing time.

the imitation task and in the processing times (shown in Fig. 5). Interesting predictions can be made from focal rather than diffuse lesions (i. e., stroke vs degenerative lesions). If only one part of the information transfer in weights W^V connecting the visual and somatic networks is impaired, then one should observe deficits in imitation only in some parts of the face and not in others. Specific local impairment of the weights W^S connecting the somatic and position networks could provoke errors in only one coordinate. For example, if the brain really uses a Cartesian representation in a head-centered frame of reference, then the position of the hand when reaching for the final target would be shifted only along one coordinate axis around the head. Spatial errors made by stroke patients should be used to test the plausibility of the model. However, because of brain reorganization, one should look at the impairment in imitation immediately after the lesion. As our model has learning properties, the model could possibly account for some of the effects of brain organization.

3.1 Conclusion

We presented a neural network architecture that could reproduce the deficits in visuo-motor imitation of meaningless gestures, reported in Goldenberg's seminal study [1]. We modelled two types of lesions that would affect either the integrative computation of the neuron or the connectivity across the neurons, leading to different predictions. Further, the model makes hypotheses on the type of representation used for the stimuli, for which there is as yet no neurological evidence. Further behavioral studies will be required to validate or invalidate the model's hypotheses and predictions.
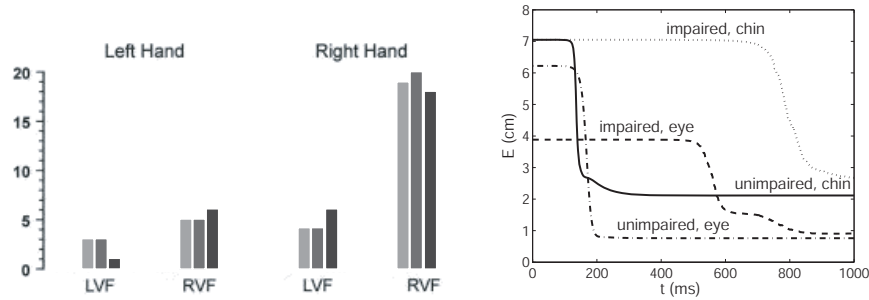
Fig. 5. Left, comparison of the results of the Goldenberg's study (light grey histograms) to the results of the simulations using the impairment model 2 (dark grey, $\tau = 30ms$, $\rho = 0.5$, and $\lambda = 0.3$) and model 1 (black, $\tau = 30ms$, $\rho = 0.4$, and $\lambda = 0.4$) respectively. The imitation was considered correct if the error distance was lower than $2.5/1.3$. Right, inhomogeneities in the precision and processing time of imitation gestures toward different parts of the face, dependent on how well represented they are in the face somatic network (in our case the eye has a larger representation than the chin).

References

1. Goldenberg, G., Laimgruber, K., Hermsdörfer, J.: Imitation of gestures by disconnected hemispheres. Neuropsychologia **39** (2001) 1432–43
2. Poeck, K., Kerschensteiner, M.: Ideomotor apraxia following right-sided cerebral lesion in a left-handed subject. Neuropsychologia **9** (1971) 359–361
3. Goldenberg, G.: Imitating gestures and manipulating a mannikin – the representation of the human body in ideomotor apraxia. Neuropsychologia **33**(1) (1995) 63–72
4. Hermsdörfer, J., Mai, N., Spatt, J., Marquardt, C., Veltkamp, R., Goldenberg, G.: Kinematic analysis of movement imitation in apraxia. Brain **119** (1996) 1575–1586
5. Decety, J., Grèzes, J., Costes, N., Jeannerod, M., Procyk, E., Grassi, E., Fazio, F.: Brain activity during observation of actions. Brain **120** (1997) 1763–1777
6. Mühlau, M., Hermsdörfer, J., Goldenberg, G., Wohlschläger, A.M., Castrop, F., Stahl, R., Röttinger, M., Erhard, P., Haslinger, B., Ceballos-Baumann, A.O., Conrad, B., Boecker, H.: Left inferior parietal dominance in gesture imitation: an fMRI study. Neuropsychologia **43** (2005) 1086–1098
7. Haaland, K.Y., Harrington, D.L., Knight, R.T.: Spatial deficits in ideomotor limb apraxia. A kinematic analysis of aiming movements. Brain **122** (1999) 1169–1182
8. Sauser, E., Billard, A.: Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. Neural Networks, Special Issue on The Brain Mechanisms of Imitation Learning **19(3)** (2006)
9. Billard, A.: Imitation: In M. A. Arbib (ed.). Handbook of Brain Theory and Neural Networks (2002) 566–569
10. Lacquaniti, F., Guignon, E., Bianchi, L., Ferraina, S., Caminiti, R.: Representing spatial information for limb movement: role of area 5 in the monkey. Cerebral Cortex **5** (1995) 391–409
11. Kohonen, T.: Self-Organizing Maps. 3. ed., Springer-Verlag (2001)
12. Zaidel, D., Sperry, R.W.: Some long term motor effects of cerebral commissurotomy in man. Neuropsychologia **15** (1977) 193–204