

The Mirror-Neurons System: data and models

Laila Craighero[°], Giorgio Metta^{#,*}, Giulio Sandini ^{#,*} and Luciano Fadiga ^{§^{#,°}}

[°] D.S.B.T.A., Section of Human Physiology, University of Ferrara, Ferrara, Italy.

^{*} D.I.S.T., LiraLab, University of Genova, Genova, Italy.

[#] The Italian Institute of Technology, Morego, Genova, Italy

Corresponding Author§: Luciano Fadiga, Department of Biomedical Sciences and Advanced Therapies, section of Human Physiology, University of Ferrara, via Fossato di Mortara, 17/19, 44100-Ferrara, Italy. E-mail: fdl@unife.it.

Color Figures: 4

Word Count: 12,800

Abstract

In this chapter we discuss the mirror-neurons system, a cortical network of areas that enables individuals to understand the meaning of actions performed by others through the activation of internal representations which motorically code for the observed actions. We review evidence indicating that this capability does not depend on the amount of visual stimulation relative to the observed action, or on the sensory modality specifically addressed (visual, acoustical). Any sensorial cue that can evoke the 'idea' of a meaningful action activates the vocabulary of motor representations stored in the ventral premotor cortex and, in humans, especially in Broca's area. This is true also for phonoarticulatory actions which determine speech production. We present also a model of the mirror-neurons system and its partial implementation in a set of two experiments. The results, according to our model, show that motor information play a significant role in the interpretation of actions and that a mirror-like representation can be developed autonomously as a result of the interaction between the individual and the environment.

Introduction

Since our discovery of mirror neurons we suggested that they might have a role in action recognition and understanding (Di Pellegrino et al, 1992; Gallese et al., 1996; see Rizzolatti and Craighero, 2004). The core of this proposal is the following: when an individual acts, the motor consequences of her action are known by her brain. Mirror neurons allow this knowledge to be extended to actions performed by others. Every time an individual observes an action performed by another individual, neurons that represent that action are activated in the premotor cortex. The observer "understands" someone else's actions because the evoked motor representation corresponds to that generated internally during action execution (see Rizzolatti et al., 2001).

In order to better understand how this mechanism works it is necessary to clarify the functional properties of the monkey's cortical region where mirror neurons have firstly been recorded (named F5 after Matelli et al, 1985).

Premotor mirror neurons: functional properties of monkey area F5

The monkey's area F5 is a premotor area cytoarchitecturally non homogeneous. Indeed, its part lying on the cortical convexity, that located in the the caudal bank of the arcuate sulcus and that in the fundus of the arcuate sulcus - which defines the anterior border of area F5 - differ as far as cytoarchitectonics is concerned (Nelissen et al., 2005; Petrides, 2006). While the convexity and the caudal bank are mainly agranular, the fundus is dysgranular. The caudal bank and the convexity of area F5 differ also for their connections with the parietal lobe. While the bank is mainly connected with parietal area AIP (buried inside the intraparietal sulcus), the convexity is mainly connected with the exposed part of the inferior parietal lobule (areas PF and PFG of Barbas and Pandya, 1987; see Figure 1).

Insert Figure 1 approximately here

Area F5 contains three types of neurons: motor neurons, "object observation-related" visuomotor neurons (also called canonical neurons), and "action-observation-related" visuomotor neurons (mirror neurons). While motor neurons are distributed in the whole area, mirror neurons are mainly located in F5 convexity, and canonical neurons are mainly located in its bank.

Motor neurons

Motor neurons selectively discharge during execution of goal-directed hand/mouth actions (Rizzolatti et al., 1988). The specificity of the goal seems to be an essential

prerequisite in activating these neurons. The same neurons that discharge during grasping, holding, tearing, and manipulating are silent when the monkey performs actions that involve a similar muscular pattern but a different goal (e.g., grasping to put away, scratching, grooming). Further evidence in favor of such a goal-directed representation is given by F5 neurons that discharge when the monkey grasps an object with either the right, the left hand or with the mouth. This observation suggests that some F5 premotor neurons can generalize the goal of the action, independently from the effector. F5 neurons can be subdivided into several classes on the basis of the action that triggers the neural discharge. The most common types are “grasping,” “holding,” “tearing,” and “manipulating” neurons. Grasping neurons form the most represented class in area F5. Many of them are selective for a particular type of prehension such as precision grip, finger prehension, or whole-hand prehension. In addition, some neurons show specificity for different finger configurations, even within the same grip type. Thus, grasping a large spherical object (whole-hand prehension, requiring the opposition of all fingers) is coded by neurons different from those coding the prehension of a cylinder (also a type of whole-hand prehension but performed with the opposition of the four last fingers and the palm of the hand). Typically, F5 premotor neurons begin to discharge before the contact between the hand and the object. Some of them stop firing immediately after contact, whereas others keep firing for a while after the contact. The temporal relation between grasping movement and neuron discharge varies from neuron to neuron. A group of neurons become active during the initial phase of the movement (opening of the hand), some discharge during hand closure, and others discharge during the entire movement from the opening of the hand until their contact with the object. Taken together, the functional properties of motor F5 neurons suggest that this area stores a set of motor schemata (Arbib, 1997) or, as proposed earlier (Rizzolatti and Gentilucci, 1988), contains a “vocabulary” of motor acts. The “words” of this vocabulary are constituted by populations of neurons. Some of them indicate the general category of an action (hold, grasp, tear, manipulate), yet others specify the effectors that are appropriate for that action. Finally, a third group is concerned with the temporal segmentation of the actions.

The motor vocabulary of actions of area F5 can also be addressed without explicit action execution. Recent experiments have shown that several F5 neurons discharge at the mere presentation of objects whose shape and size is congruent with the type of grip that is coded motorically by the same neurons (object-observation visuomotor neurons) (Murata et al, 1997) or during observation of another monkey or the experimenter making a goal-directed action similar to that coded by the same neurons (action-observation visuomotor neurons) (di Pellegrino et al, 1992; Gallese et al, 1996).

Object observation-related visuomotor neurons (canonical neurons)

Object observation visuomotor neurons are active when manipulating an object and when fixating the same object. These neurons discharge at the mere presentation of objects whose shape and size is congruent with the type of grasp coded motorically by the same neurons.

The visual responses of object observation-related F5 neurons have been formally studied by Murata and colleagues (Murata et al, 1997) using a behavioral paradigm, which allowed to test separately the neurons response to object observation, to the waiting phase between object presentation and movements onset, and during movement execution. The results showed that the majority of these canonical visuomotor neurons are selective to one or at most a few specific objects. Moreover, there is a strict congruence between their visual and motor properties: Neurons that become active when the monkey observes small objects discharge also during precision grip. On the contrary, neurons selectively active when the monkey looks at a large object discharge also during actions directed toward large objects (e.g., whole-hand prehension). The most likely interpretation for the visual discharge of these visuomotor neurons is that there is a close link between the most common three-dimensional stimuli and the actions required to interact with them. Thus, every time a graspable object is presented visually, the corresponding F5 neurons are activated and the action is “automatically” evoked. Under certain circumstances this neural activity guides the execution of the movement directly; under others, it remains an unexecuted representation of the action that might be used for semantic knowledge.

Action observation-related visuomotor neurons (mirror neurons)

Action observation-related visuomotor neurons are active when manipulating an object and when watching someone else performing the same action on the same object.

To be triggered by visual stimuli, action observation visuomotor neurons require an interaction between a biologic effector (hand or mouth) and an object. The sight of the object alone, that of an agent mimicking an action, or an individual making intransitive (non-object-directed) gestures are all ineffective. The object significance for the monkey has no obvious influence on the mirror neuron response: Grasping a piece of food or a geometric solid produces responses of the same intensity.

Mirror neurons show a large degree of generalization. Very different visual stimuli, but representing the same action, are equally effective. For example, the same mirror neuron that responds to a human hand grasping an object responds also when the grasping hand is that of a monkey. Similarly, the response is, typically, not affected if the action is done near or far from the monkey, despite the fact that the size of the observed hand is obviously different in the two conditions. It is also of little importance for neuron activation if the observed action is eventually rewarded. The discharge is of the same intensity if the experimenter grasps the food and gives it to the recorded monkey or to another monkey introduced in the experimental room.

Typically, mirror neurons show congruence between the observed and executed action. This congruence can be extremely faithful, i.e., the effective motor action (e.g., precision grip) coincides with the action that, when seen, triggers the neurons (e.g., precision grip). For other neurons, the congruence is broader and the motor requirements (e.g., precision grip) are usually stricter than the visual ones (any type of hand grasping).

The most likely interpretation for visual discharge in mirror neurons is that it evokes an internal representation of the observed action. In other terms, the observed action selects, in the F5's motor vocabulary, a congruent "motor word," a potential action.

Actions do not generate only visual consequences and in fact action-generated sound and noise are also very common in nature. One could expect, therefore, that also

this sensory information, related to a particular action, can determine a motor activation specific for that same action. Kohler and colleagues (2002) addressed this point by investigating F5 neurons that discharge when the monkey makes a specific hand action but also when it *hears* the corresponding action-related sounds. Auditory properties of the neurons were studied by using sounds produced by the experimenter's actions and non-action-related sounds. Neurons were not activated by non-action related sounds, while they responded specifically to the sound of an object breaking and of paper ripping, which are the hand actions more frequently executed by the monkey. Neurons were studied in an experimental design in which two hand actions were randomly presented in vision-and-sound, sound-only, vision-only, and motor conditions (monkeys performing object-directed actions). The authors (Kohler et al, 2002) found that 13% of the investigated neurons discharge both when the monkey performs a hand action and when it hears the action-related sound. Moreover, most of these neurons discharge also when the monkey observed the same action, demonstrating that these "audiovisual mirror neurons" represent actions independently of whether they are *performed, heard, or seen*.

A typical property of F5 mirror neurons is that their response is quite independent from the observer's point of sight. In other words, the same grasping action activates a given mirror neuron also if it is observed from different points of view (see Figure 2).

Insert Figure 2 approximately here

How can the brain achieve an invariant description of a given action by using so different visual information? One possibility is that the system recognizes others' actions by using the same mechanisms it uses to visually control the execution of its own actions. In other terms, the point-of-view dependent visual information could be generalized by the invariance of the motor command driving action execution. To test this hypothesis, we manipulated the amount of visual information on the monkey own acting hand during grasping execution. Results showed that a significant percentage of F5 purely motor neurons are modulated by the vision of the own hand in action and that this modulation is

mainly negative (less discharge) when the hand is not visible. These F5 (visuo)-motor neurons may have formed the original nucleus from which mirror neurons may have developed, possibly during the ontogenesis (Gesierich et al, in preparation). Figure 3 depicts a simplified schema of this model which will be more formally described in the last section of this chapter.

Insert Figure 3 approximately here

The mirror-neuron system and action recognition: prospective or reactive mechanisms?

The functional properties of F5 neurons indicate that in primates the action representations are addressed not only for motor execution, but also during observation of graspable objects, and perception (visual, acoustical, other?) of actions performed by others. The presence of such a “vocabulary” of actions has important functional implications. Firstly, the execution of motor commands is strongly facilitated. The existence of preformed motor schemata, which are anatomically linked (hard-wired) with cortical (primary motor cortex) and subcortical motor centers, facilitates the selection of the most appropriate combination of movements simply by addressing the general idea of an action. Thus, the number of variables that the motor system (at the premotor level) has to control to achieve the action goal is reduced. Secondly, it simplifies the association between a given stimulus (i.e., a visually presented object) and the appropriate motor response. This is the case of object observation–related visuomotor responses. Thirdly, it gives the brain a store of “ideas of action” that could be activated whenever visual or acoustic stimuli suggesting that another person is executing an action are perceived. This is the case of action perception-related sensorimotor responses.

On the basis of these functional properties that characterize not only mirror neurons but all the neurons in F5 region, it can be hypothesized that mirror neurons are at the basis

of action recognition/understanding, and that this capability is not strictly dependent from the amount of stimulation perceived by the individual.

Evidence in monkeys

The hypothesis that complete visual information about the perceived action is not necessary to determine mirror neurons activation was directly tested by Umiltà and colleagues (Umiltà et al, 2001). The experimental paradigm consisted of two sessions. In the first session, the monkey was shown with a fully visible action directed toward an object or with the mimicry of the same action in the absence of the object. From previous studies it was known that mirror neurons do not discharge when the object is absent. In the second session, the monkey saw exactly the same experimental conditions but with the final part of the action hidden by a screen. Before each trial the experimenter could choose whether to place a piece of food behind the screen so that also the monkey knew whether a target for the action was present. The main result of the experiment was that several neurons discharged in the 'hidden' condition, but only when the animal knew that the food was present. This evidence was interpreted as a good demonstration that mirror neurons fire also during the reaching/grasping of an object placed out of sight, as long as the intention and the plausibility of the reaching/grasping action are clear.

The conclusion is that understanding of the action is not fully based on the visual description of the scene but it refers also to the motor representation of the action goal, shared by both the agent and the observer, and triggered by the context in which the action is performed (i.e. the presence or the absence of the food on the table behind the screen).

Evidence in humans

In recent years, a series of brain imaging studies demonstrated that a mirror-neuron system is also present in the human brain. When an individual observes an action, or executes it, a network of cortical areas is activated, including the ventral premotor cortex, the inferior frontal gyrus, the inferior parietal lobule and the superior temporal cortex (see

for review Rizzolatti and Craighero, 2004). Furthermore, transcranial magnetic stimulation (TMS) was used to directly investigate the involvement of the motor system in humans during observation of others' actions. TMS is an alternative technique to the single neuron recordings that can be used in humans to obtain good temporal resolution. Single or paired-pulse TMS allows to measure cortical excitability during different phases of an observed action. Moreover, this technique can help to verify the specific involvement of the motor system by discriminating the muscles that are involved in the motor replica. A series of TMS experiments showed that also in humans the mirror system is not strictly dependent on the visual stimulation but it is active whenever a motor representation is addressed (Gangitano et al, 2004; Borroni et al, 2005).

Gangitano and colleagues (Gangitano et al. 2001) in a TMS experiment evoked MEPs in the first dorsal interosseus muscle at different time intervals, while subjects were watching a video clip of a hand approaching and grasping a ball and demonstrated that the specific activation of the observer's muscles is temporally coupled to the dynamics of the observed action. In a further experiment Gangitano et al. (2004) investigated whether this pattern of modulation was the consequence of a "resonant plan" evoked at the beginning of the observation phase or whether the plan was fractioned in different phases sequentially recruited during the course of the ongoing action. The authors therefore used the same procedure as in Gangitano et al. (2001) with the following exception: Subjects were shown video clips representing an unnatural movement, in which the temporal coupling between reaching and grasping components was disrupted, either by changing the time of appearance of maximal finger aperture, or by substituting it with an unpredictable closure. In the first case, the observation of the uncommon movements did not exert any modulation in motor excitability. In the second case, the modulation was limited to the first time of stimulation. Modulation of motor excitability was clearly suppressed by the appearance of the sudden finger closure and was not substituted by any other pattern of modulation. This finding suggests that a motor plan, which includes the temporal features of the natural movement, is activated immediately after the observed movement-onset and is discarded when these features cease to match the visual properties

of the observed movement. Thus, the human mirror system seems to be able to infer the goal and the probability of an action during the development of its ongoing features.

Recently, Borroni et al. (2005) aimed at verifying the degree of correspondence, especially with respect to a fine temporal resolution, between the observation of prolonged movements and its modulatory effects in the observer. For this purpose the authors asked subjects to watch a cyclic flexion-extension movement of the wrist. The same sinusoidal function was used to fit both observed wrist oscillation and motor resonance effects on the observer's wrist motor circuits. In this way the authors could describe a continuous time course of the two events and precisely determine their phase relation. MEPs were elicited in the right forearm muscle of subjects who were observing a 1-Hz cyclic oscillation of the right hand executed by another person. The results indicated that movement observation elicited a parallel cyclic excitability modulation of the observer's MEP responses following the same period as the observed movement. Interestingly, the MEP modulation preceded the observed movement, being related to time course of muscular activation of the demonstrator and not to the visually perceived movement. This finding indicates that the mirror-neuron system anticipates the movement execution, rather than simply reacting to it.

Thus, the involvement of observer's motor system is not necessarily consequent to the explicit visual description of the complete action but, rather, it may intervene in filling gaps because it gives to the observer an implicit motor knowledge about the observed action. In other words, the mirror system seems to possess the capability to predict the action outcome.

In this line are the data by Kilner and colleagues (Kilner et al. 2004). In an event-related potentials experiment, these authors showed that the readiness potential (RP), a well known electrophysiological marker of motor preparation, is detectable also during action observation. Furthermore, when the upcoming action is predictable, the rise of the RP precedes the observed movement onset. They recorded electroencephalograms from subjects while they watched a series of short video clips showing an actor's right hand and a colored object. In half of the videos the hand moved, while in the other half it remained

stationary. At the beginning of each video the color of the object indicated whether the hand would subsequently move or not. Thus, the observed movements were entirely predictable from the color of object in the video. The results revealed a significant negative gradient that started 500 ms before the onset of the observed hand movement. This activity was comparable with the onset of the movement-related RP produced when subjects actually executed a movement. These results suggest an active role of the mirror system in setting up an anticipatory model of another person's action, endowing our brain with the ability to predict his or her intentions ahead of their realization.

All the reported experiments, however, investigated the involvement of the mirror system during observation of real hands executing goal directed actions. In a very recent experiment our group (Fadiga et al, 2006) wanted to verify if the vision of a real hand is a necessary prerequisite to activate the mirror system, or if any cue suggesting the presence of a hand performing meaningful movements is a sufficient stimulus. To this purpose we submitted human subjects to an fMRI scanning while they were observing a particular category of hand gestures: Hand shadows representing animals opening their mouths. Hand shadows only implicitly 'contain' the hand creating them (i.e. hands are not visible but subjects are aware of the fact that the presented animals are produced by the hands). Therefore, they are interesting stimuli that might be used to answer the question of how much and what details of a hand gesture activate the mirror-neuron system. During the fMRI scan, healthy volunteers (n=10) observed videos representing (i) the shadows of human hands depicting animals opening and closing their mouths, (ii) human hands executing sequences of meaningless finger movements, or (iii) real animals opening their mouths. Each condition was contrasted with a 'static' condition, where the same stimuli presented in the movie were shown as static pictures (e.g. stills of animals presented for the same amount of time as the corresponding videos). In addition, to emphasize the action component of the gesture, brain activations were further compared between pairs of conditions in a block design.

Insert Figure 4 approximately here

Figure 4 shows, superimposed, the results of the moving vs. static contrasts for animal hand shadows and real animals conditions (red and green spots, respectively). In addition to largely overlapping occipito-parietal activations, a specific differential activation emerged in the anterior part of the brain. Animal hand shadows strongly activated left parietal cortex, pre- and post-central gyri (bilaterally), and bilateral inferior frontal gyrus (BA 44 and 45). Conversely, the only frontal activation reaching significance in the moving vs. static contrast for real animals was located in bilateral BA 6, close to the premotor activation shown in an fMRI experiment by Buccino and colleagues (Buccino et al, 2004) where subjects observed mouth actions performed by monkeys and dogs. This location may therefore correspond to a premotor region where a mirror-neuron system for mouth actions is present in humans. The results shown in Figure 4 indicate that the shadows of animals opening their mouths, although clearly depicting animals and not hands, convey implicit information about the human being moving her hand in creating them. Indeed, they evoke an activation pattern which can be superimposed to that evoked by hand action observation (Grafton et al, 1996; Buccino et al, 2001; Grèzes et al, 2003). Thus, the results demonstrate that the mirror-neuron system becomes active even if the pictorial details of the moving hand are not explicitly visible: In the case of our stimuli, the brain ‘sees’ the performing hand also behind the visual appearance. Consequently, the human mirror system (or at least part of it) can be seen more as an active interpreter than as a passive perceiver (or resonator).

The possibility to be an active interpreter is based on the knowledge of the context in which the observed action is performed. Different cues coming from the environment activate the representation of the most probable action: in the Umiltà experiment (Umiltà et al, 2001) the knowledge of the presence of the food behind the screen gives plausibility to the hidden action; in the Kilner experiment (Kilner et al, 2004) the color of the object at the beginning of the trial prompts the belief that a grasping towards that object is going to be executed; in the Fadiga experiment (Fadiga et al, 2006) the old memory of playing with shadows on the wall during childhood links the observed animal shadows with hand

movements. Recently, we tested the possibility that the mirror-neurons system could be modulated by the canonical neurons activation determined by the vision of the to-be-grasped object (Craighero et al, in press). To this purpose we asked subject to detect the instant at which the demonstrator's hand touched the object. Two different types of grasping on the same object were presented, differing for the type of fingers opposition space: In one case the type of grasping was the one more commonly chosen to grasp the presented object, in the other case it was a less appropriate one. This experimental manipulation created a situation of conflict in terms of motor representations determining two main conditions: A congruent one, in which the motor program evoked by object observation coincides with that executed by the experimenter, and an incongruent one, where the two motor programs differ. Our results showed that subjects' response times are well below those commonly found in simple reaction times tasks (usually around 120-150 ms), indicating that, to accomplish the task, subjects indeed use a predictive model of the seen action. Moreover, response times were shorter for suitable grasping trials than for not suitable ones. This indicates that action prediction is based on the internal motor representation of the seen action, and that whenever an incongruence is present between the action evoked in the observer by the to-be-grasped object and the observed action, actually executed on it, the ability to predict the action outcome decreases.

Considering both the functional properties of the neurons of the ventral premotor cortex of the monkey and those of the human mirror-neurons system, as described by TMS and brain imaging experiments, we can argue that the ventral premotor cortex is automatically activated whenever the 'idea' of an action is even suggested. This suggestion can derive from the sight of a graspable object, as in the case of canonical neurons (Murata et al, 1997), from the visual (di Pellegrino et al, 1992) or acoustical (Kohler et al, 2002) perception of a transitive action performed by another individual, as in the case of mirror neurons in monkeys and of mirror-neurons system in humans (see Rizzolatti and Craighero, 2004). Moreover, experimental evidence indicates that the idea of an action does not necessarily require a complete perceptual stimulation to be elicited (Umiltà et al, 2001; Fadiga et al. 2006).

In conclusion, the properties of the mirror-neurons system not only are in favor of its role in action understanding, but clearly suggest that mirror neurons are fundamental in interpreting others' intentions and in anticipating the outcome of others people's actions, providing a key mechanism to successfully interact in a social environment.

Broca's area is the core center of the human mirror-neurons system

As discussed in the previous section, experimental evidence demonstrates that a mirror-neurons system is also present in the human brain. The first evidence of the existence of a mirror-like visuomotor activation in the human brain has been provided by Fadiga, Fogassi, Pavesi, and Rizzolatti (1995) by a TMS experiment. The motor cortex of normal human participants was magnetically stimulated and motor evoked potentials (MEPs) were recorded from intrinsic and extrinsic hand muscles. It was reasoned that, if the observation of a hand movement activates the premotor cortex, this should, in turn, induce an enhancement of MEPs elicited by the magnetic stimulation of the hand representation of the motor cortex. The results confirmed this hypothesis showing a pattern of muscle facilitation during action observation that strictly resembles that occurring during the actual execution of the observed movements. In other words, looking at a hand closing onto an object evokes a facilitation of the observer's flexors muscles. Strafella and Paus (2000), by using the double stimulus TMS technique, demonstrated the cortical origin of this facilitation. They showed that the interstimulus interval between two close stimulations, which evoked the larger motor facilitation during action observation, was compatible with cortico-cortical facilitating connections.

Further evidence that cortical motor areas are activated during movement observation comes from MEG experiments. Hari and colleagues (Hari et al, 1998) recorded neuromagnetic oscillatory activity of the human precentral cortex elicited by median nerve stimulation in healthy volunteers during rest (i), manipulation of a small object with their right hand (ii), and observation of another individual performing the same task (iii). The

cortical 15–25 Hz rhythmical activity was measured. In agreement with previous data (Salmelin and Hari, 1994), this activity was suppressed during movement execution. Most interestingly, the rhythm was also significantly diminished during movement observation. Control experiments confirmed the specificity of the suppression effect. Because the recorded 15–25 Hz activity originates mostly in the anterior bank of the central sulcus, it appears that the human primary motor cortex desynchronizes (and therefore becomes more active) during movement observation in the absence of any active movement. Similar results were obtained also by Cochin and colleagues (Cochin et al, 1998), who recorded EEG from subjects observing videos where human movements were displayed. As a control, moving objects, moving animals and still objects were presented. The data showed that the observation of human movements, but not that of objects or animals, desynchronizes the EEG pattern of the precentral cortex.

A series of brain imaging experiments were carried out in order to assess which cortical area could be the homologue of the monkey F5 mirror system. Hand grasping movements (Grafton et al, 1996; Rizzolatti et al, 1996) as well as, more recently, more complex hand/arm movements were used as visual stimuli (Decety et al, 1997; Grèzes et al, 1998). The results of the first experiments showed that during the observation of hand grasping, among the activation of other areas, there was an activation of the left inferior frontal cortex, in correspondence of the Broca's region, a region historically known to be involved in language production (Broca, 1861). In studies carried out by the Lyon group (Décety et al., 1997; Grèzes et al., 1998) the involvement of Broca's area during observation of hand/arm actions was further confirmed. The authors instructed subjects to observe meaningful (with a goal) and meaningless movements. The main result when subjects observed meaningless arm movements was the bilateral activation of the parietal lobe, the activation of the left precentral gyrus and that of the right side of the cerebellum (Grèzes et al, 1998). On the contrary, the observation of meaningful hand actions, in addition to the already mentioned frontal and parietal areas, activated the left inferior frontal gyrus (Broca's region). More recently, two additional studies have shown that a meaningful hand-object interaction, more than pure movement observation, is effective in triggering

Broca's area activation (Hamzei et al, 2003; Johnson-Frey et al, 2003). Similar conclusions have been reached also for the observation of mouth movements (Campbell et al, 2001). These results, together with comparative cytoarchitectonical data (see Petrides and Pandya, 1994; Nelissen et al, 2005; Petrides, 2006), and fMRI data from Binkofsky and colleagues (Binkofsky et al, 1999) demonstrating that Broca's region becomes active also during manipulation of complex objects, suggest that Broca's region has the putative role of human homologue of area F5 in the monkey.

Mirror-neurons system and speech recognition

The presence of an audio-motor resonance (Kohler et al, 2002) in a region that, in humans, is classically considered a speech-related area, immediately evokes the Liberman's motor theory of speech perception (Liberman et al, 1967; Liberman and Mattingly, 1985; Liberman and Wahlen, 2000). This theory maintains that the ultimate constituents of speech are not sounds but articulatory gestures that have evolved exclusively at the service of language. Consequently, speech perception and speech production processes can share a common repertoire of motor primitives that, during speech production, are at the basis of the generation of articulatory gestures, and during speech perception are activated in the listener as the result of an acoustically evoked motor "resonance". According to Liberman's theory, the listener understands the speaker when his/her articulatory gestural representations are activated by listening to verbal sounds. Although this theory is not unanimously accepted, it offers a plausible model of an action/perception cycle in the frame of speech processing.

To investigate if speech listening activates listener's motor representations, our group (Fadiga et al, 2002), in a TMS experiment, tested for the presence in humans of a system that motorically "resonates" when an individual listen to verbal stimuli. Healthy subjects were requested to attend to an acoustically presented randomized sequence of disyllabic words, disyllabic pseudo-words and bitonal sounds of equivalent intensity and duration. Words and pseudo-words were selected according to a consonant-vowel-consonant-consonant-vowel (cvccv) scheme. The embedded consonants in the middle of words and

of pseudo-words were either a double 'f' (labiodental fricative consonant that, when pronounced, requires slight tongue tip mobilization) or a double 'r' (lingua-palatal fricative consonant that, when pronounced, requires strong tongue tip mobilization). Bitonal sounds, lasting about the same time as verbal stimuli and replicating their intonation pattern, were used as a control. The excitability of the motor cortex in correspondence of the representation of tongue movements was assessed by using single pulse TMS and by recording MEPs from the anterior tongue muscles. The TMS stimuli were applied synchronously with the double consonant of the presented verbal stimuli (words and pseudo-words) and in the middle of the bitonal sounds. Results showed that during speech listening there is an increase of the motor evoked potentials recorded from the listeners' tongue muscles when the word strongly involves tongue movements. This indicates that when an individual listens to verbal stimuli his/her speech related motor centers are specifically activated. Moreover, words-related facilitation was significantly larger than pseudo-words related one. These results indicate that the passive listening to words that would involve tongue mobilization (when pronounced) induces an automatic facilitation of the listener's motor cortex. Furthermore, the effect is stronger in the case of words than in the case of pseudo-words suggesting a possible unspecific facilitation of the motor speech center due to recognition that the presented material belongs to an extant word.

Similar results were obtained by Watkins and colleagues (Watkins et al, 2003). By using TMS technique they recorded MEPs from a lip (*orbicularis oris*) and a hand muscle (first dorsal *interosseus*) in four conditions: listening to continuous prose, listening to non-verbal sounds, viewing speech-related lip movements, and viewing eye and brow movements. Compared to control conditions, listening to speech enhanced the MEPs recorded from the *orbicularis oris* muscle. This increase was observed only in response to the stimulation of the left hemisphere. No changes of the MEPs in any condition were observed following the stimulation of the right hemisphere. Finally, the size of MEPs elicited in the first *interosseus* muscle did not differ in any condition.

Taken together these experiments show that when an individual listen to verbal stimuli there is an activation of the speech-related motor centers. It is however unclear if this activation could be interpreted in terms of an involvement of motor representations in speech processing and, perhaps, in perception.

In order to investigate the perceptual role of Broca's area and considering that this area has been classically considered specifically involved in phonological processing (at least in production) we decided to use a phonological paradigm in a new experiment: the "phonological priming" task. Phonological priming effect refers to the fact that a target word is recognized faster when it is preceded by a prime word, sharing with it the last syllable (rhyming effect, Emmorey, 1989). In a single pulse TMS experiment (see Fadiga et al, in press) we therefore stimulated participants' inferior frontal cortex while they were performing a phonological priming task. Subjects were instructed to carefully listen to a sequence of acoustically presented pairs of verbal stimuli (dysyllabic 'cvcv' or 'cvccv' words and pseudo-words) in which the final phonological overlap was present (rhyme prime) or, conversely, not present. The task of the subjects was to make a lexical decision on the second stimulus (target) by pressing with either the index or the middle finger one of two buttons whether the target was a word or a pseudo-word. The pairs of verbal stimuli belonged to four categories which differed for their lexical content in the prime and in the target (prime-word/target-word (W-W), prime-word/target-pseudoword (W-PW), prime-pseudoword/target-word (PW-W), prime-pseudoword/target-pseudoword (PW-PW)). Each category contained both rhyming and non-rhyming pairs. In some randomly selected trials, we administered single pulse TMS in correspondence of left BA44 (Broca's region, localized by using "Neurocompass", a frameless stereotactic system built in our laboratory) during the interval (20 ms) between prime and target stimuli.

In trials without TMS, there were three main results (Figure 5): (i) strong and statistically significant facilitation (phonological priming effect) when W-W, W-PW, PW-W pairs are presented; (ii) no phonological priming effect when the PW-PW pair is presented; (iii) faster responses when the target is a word rather than a pseudo-word (both in W-W and PW-W).

Insert Figure 5 approximately here

An interesting finding emerges from the analysis of these results: The presence or absence of lexical content modulates the phonological priming effect. When neither the target nor the prime has access to the lexicon (PW-PW pair) the presence of the rhyme does not facilitate the recognition of the target. In other words, in order to have a phonological effect it is necessary to have access to the lexicon.

In trials with TMS delivery, only W-PW pairs were affected by brain stimulation, the W-PW pair behaving exactly as the PW-PW one. This finding suggests that the stimulation of the Broca's region might have affected the lexical property of the prime (i.e. the meaningfulness of the stimulus). As consequence, the impossibility to access the lexicon determines the absence of the phonological effect. According to our interpretation, the TMS-related effect is absent in the W-W and PW-W pairs because of the presence of a meaningful (W) target. The finding that TMS administered on Broca's region during phonological priming paradigm influences the rhyming effect only in the case of W-PW pairs poses a theoretical problem. In our previous TMS experiment on motor facilitation during speech listening (Fadiga et al, 2002) we have found cortical facilitation during listening of both words and pseudo words. This discrepancy suggests that a cortical area different from Broca's one should be involved in such a "low level" motor resonance. Its localization will be the argument of our future experimental work.

By summarizing the experimental evidence we presented here, we can claim that the activation of Broca's region during speech processing, more than indicating a specific role of this area, may reflect its general involvement in meaningful action recognition. This possibility is based on the observation that, in addition to speech-related activation, this area is activated during observation of meaningful hand or mouth actions. Speech represents a particular case of this general framework: Among meaningful actions, phonoarticulatory gestures are meaningful actions conveying words. The consideration

that Broca's area is the human homologue of the monkey mirror neurons area opens the possibility that human language may have evolved from an ancient ability to recognize visually or acoustically perceived actions performed by others (Rizzolatti and Arbib, 1998).

The motor representation of hand/mouth actions present in Broca's area, which derives from the execution/observation (hearing) matching system already present in monkeys, may have given to this area the capability to deal with verbal communication because of its twofold involvement with motor goals: During execution of own actions and during perception of others' ones. Our hypothesis is that the original role played by this region in generating/extracting action meanings might have been generalized during evolution giving to this area the capability to deal with meanings (and rules) sharing with the motor system similar hierarchical and sequential structures. Recent data from our laboratory on frontal aphasic patients are in line with this idea (see Fadiga et al, in press).

A model of area F5 and of the mirror-neurons system

This section proposes a model of the mirror-neurons system, whose components are in general agreement with the functional properties of area F5 and with the knowledge on the connections that this area maintains with other cortical regions, which describes how the mirror-neurons system intervenes in action recognition (Metta et al. 2006).

It is known that F5 is part of a larger circuit comprising various areas in the parietal lobe (a large reciprocal connection with anterior intraparietal area, AIP), indirectly from superior temporal sulcus (STS), and other premotor and frontal areas. Moreover, it is strongly involved in the generation and control of action indirectly through primary motor cortex (F1), and directly by projecting to motor and medullar interneurons in the spinal cord (see Luppino and Rizzolatti, 2000).

Our model of area F5 revolves around two concepts that are certainly related to the evolution and development of this unique area of the brain. Firstly, we posit that the mirror-neurons system did not appear brand new in the brain but likely evolved from a

pre-existing structure devoted solely to the control of grasping action. The reasons for this claim are to be found in the large percentage of motor neurons in F5 (70%) compared to those that have also visual responses. Secondly, we attribute a fundamental role to canonical neurons – and in general that of contextual information specifying the action goal – in the development of the mirror neurons. Since purely motor, canonical, and mirror neurons are found together in F5, it is very plausible that local connections determine at least in part the activation of F5.

Our model follows a forward-inverse approach that has been also proposed in computational motor control theory (Kawato et al, 1987; Wolpert and Miall, 1996) and for explanatory purpose it can be divided into two parts. The first part describes what happens in the actor's brain, the second what happens in the observer's brain when watching another acting individual. As we will see the same structures are used both when acting and when observing an action.

Insert Figure 6 approximately here

The agent's point of view

We shall consider first what happens from the actor's point of view (see Figure 6). In her perspective, decision to undertake a particular action is attained by the convergence in area F5 of many factors including the contextual- (by signals from parietal and frontal areas) and object-related information (canonical neurons). Object and context bias the activation of a specific motor plan, which specifies the goal of the motor system in motor terms and, we generally suppose, it includes temporal information. Our model hypothesize that action specification is initially "described" in frontal areas in some internal reference frame and then transformed into the motor plan by an appropriate controller in F5.

The action plan unfolds mostly open loop (i.e. without employing feedback). A form of feedback (closed loop) is required though to counteract disturbances and to learn from mistakes. This is obtained by relying on a forward or direct model that predicts the

outcome of the action as it unfolds in real-time. The output of the forward model can be compared with the signals derived from sensory feedback, and differences accounted for (the cerebellum is believed to have a role in this) (Miall et al, 1993; Wolpert and Miall, 1996). A delay module is included to take into account the different propagation times of the neural pathways carrying the predicted and actual outcome of the action. Note that the forward model is relatively simple, predicting only the motor output in advance: Since motor commands are generated internally it is relatively easy to imagine a predictor for this signals.

The inverse model (Visuo-Motor Map, VMM) is much more complicated since it maps sensory feedback (vision mainly) back into motor terms. Visual feedback clearly includes both the hand-related information (STS) and the object information (AIP, IT, and canonical neurons). Finally the predicted and the sensory signals arising from action execution are compared and the feedback error sent back to the controller. The mismatch between the planned and actual action can either be used to compensate on the fly by means of a feedback controller, or to adjust over longer periods of time through learning (Kawato et al, 1987).

The output of area F5, finally activates the motor neurons in the spinal cord (directly or indirectly through motor synergies) to produce the action. This is indicated by a connection to appropriate muscular synergies representing the spinal cord circuits.

Learning of the direct and inverse models can be carried out during ontogenesis by a procedure of self-observation and exploration of the state space of the system: Grossly speaking, simply by “detecting” the sensorial consequences of motor commands.

Learning of the affordances of objects (the canonical neurons response) with respect to grasping can also be achieved autonomously by a trial and error procedure, which explores the consequences of trying many different actions of the agent's motor repertoire (different grasp types) to different objects. This includes things such as discovering that small objects are optimally grasped by a pinch or precision grip, while big and heavy objects require a power grasp.

The observer's point of view

In the observer situation motor and proprioceptive information is not directly available. The only readily available information is vision or sound. The central assumption of our model is that the structure of F5 could be co-opted in recognizing the observed actions by transforming visual cues into motor information as before. In practice the inverse model is accessed by visual information, since the observer is not acting herself, visual information directly reaches in parallel the sensorimotor primitives in F5. Only some of them are actually activated because of the “filtering” effect of the canonical neurons and of other contextual information (possibly at a higher level, knowledge of the actor, plausibility of the hand posture, etc.). This procedure could be used then to recognize the action by measuring the most active motor primitive. Thus, in our model, many factors, including the affordances of the target object, determine the recognition and interpretation of the observed action.

Ontogenesis of mirror neurons

Our model gives us also the possibility to hypothesize the ontogenesis of mirror neurons. First of all, the inverse model, the Visuo-Motor Map, can be learned through a procedure of self exploration. Motor commands and correlated visual information are readily available to the developing infant. It is easy to imagine a procedure that learns the inverse model on the basis of this information.

On the top of the Visuo-Motor Map, it is plausible that the canonical representation is acquired via the manipulation of a large set of different objects. F5 canonical neurons represent an association between objects' physical properties and the actions they afford: e.g. a small object affords a precision grip, or a coffee mug affords being grasped by the handle. The understanding of object properties and the goal of actions is subsequently fundamental for disambiguating visual information. In fact, certain actions are more likely to be applied to a particular object, and certain objects are more likely to be used in certain actions. A link between action and effects can be traced backward on the basis of our

experience: To obtain the effects we have to apply the same action that earlier led to those effects.

Bearing this in mind, when observing some other individual's actions; our understanding can be framed in the same terms: When I see someone grasping a mug full of coffee by the handle, I know that that precise affordance is the most suitable for drinking. If the same mug is empty and I see the agent grasping it by inserting the fingers into it, I can hypothesize that she is going to wash it. Consequently, I can recognize and also predict the outcome of the action on the basis of the link between the contextual information, the type of affordance, and the knowledge of object's properties. It is plausible that a mirror representation derives from the association between the visual information of others' actions and the action-effects link. To obtain this association, however, it is necessary that the observed consequences of an action are recognized as similar in the self or the other individual acting. Only if this happens, the association between the canonical response and the mirror one can then be made.

Role of motor information in the action recognition process

The simplest way of confirming the hypothesis that motor representations are the basis of action recognition is to equip a computer with means of "acting" on objects, collect visual and motor data and build a recognition system that embeds some of the principles of operation that we identified in our model (see Figure 6). In particular, the hypothesis we would like to test is whether the extra information available during learning (e.g. kinesthetic and tactile) can improve and simplify the recognition of the same actions when they are just observed: i.e. when only visual information is available. Given the current limitations of robotic systems the simplest way to provide "motor awareness" to a machine is by recording grasping actions of human subjects from multiple sources of information including joint angles, spatial position of the hand/fingers, vision, and touch.

For this purpose we assembled a computerized system composed of a data glove (CyberGlove by Immersion), a pair of CCD cameras (Watek 202D), a magnetic tracker (Flock of bird, Ascension), and two touch sensors (FSR). Data was sampled at frame rate,

synchronized, and stored to disk by a Pentium class PC. The cyber glove has 22 sensors and allows recording the kinematics of the hand at up to 112Hz. The tracker was mounted on the wrist and provides the position and the orientation of the hand in space with respect to a base frame. The two touch sensors were mounted on the thumb and index finger to detect the moment of contact with the object. Cameras were mounted at appropriate distance with respect to their focal length to acquire the execution of the whole grasping action with maximum possible resolution.

The glove was lightweight and did not limit any way the movement of the arm and hand. Data recording was carried out with the subject sitting comfortably in front of a table and performing grasping actions naturally toward objects approximately at the center of the table. Data recording and storage was carried out through a custom-designed application; Matlab was employed for post processing.

We collected a large data set and processing was then performed off-line. The selected grasping types were: power grasp-cylindrical, power grasp-spherical, and precision grasp. Since the goal was to investigate to what extent the system could learn invariances across different grasping types by employing motor information for classification, the experiment included gathering data from a multiplicity of viewpoints. The database contained objects which afford several grasp types to assure that recognition cannot simply rely on exclusively extracting object features. Rather, according to our model, this is supposed to be a confluence of object recognition (canonical neurons) with hand visual analysis (STS).

A set of three objects was employed: a small glass ball, a parallelepiped, and a large sphere. Each grasping action was recorded from six different subjects (right handed, age 23-29, male/female equally distributed), and moving the cameras to 12 different locations around the subject including two different elevations with respect to the table top which amounts to 168 sequences per subject. Each sequence contained images of the scene from the two cameras synchronized with the cyber glove and the magnetic tracker data.

The visual features were extracted from pre-processed image data. The hand was segmented from the images through a simple color segmentation algorithm. The bounding

box of the segmented region was then used as a reference frame to map the view of the hand to a standard size. The orientation of the color blob in the image was also used to rotate the hand to a standard orientation. This data set was then filtered through Principal Component Analysis (PCA) by maintaining only a limited set of eigenvectors corresponding to the first 2 to 15 largest eigenvalues.

One possibility to test the influence of motor information in learning action recognition is to contrast the situation where motor-kinesthetic information is available in addition to visual information with the control situation where only visual information is available.

In the “motor space” session we used the output of the VMM (see the model schematics) and thus employed motor features for classification. The VMM was approximated from data by using a simple feedforward neural network with sigmoidal units trained with backpropagation. The input of the VMM was the vector of the mapping of the images onto the space spanned by the first N PCA vectors; the output was the vector of joint angles acquired from the data glove. In the “visual space” session the classification was performed in visual space directly.

Classification was always performed by training a Bayesian classifier. In this formulation we identified the likelihood term with the activity of F5 motor neurons (probability of seeing certain features given the performed action and target object) and the priors with the canonical neurons (probability of using a certain action on a given object). The classifier then applies a MAP criterion (maximum a-posteriori) by computing the unnormalized posterior probability and taking the maximum over the possible actions.

The results of the experiment are reported in table 1.

Different sequences were used during the training and the testing phases. During the training phase, 24 sequences from only one point of view were used in the motor space session, while 64 sequences from all the four different points of view were used in the visual space session. Thus, the classifier was trained with the maximum available data only in the latter session. During the testing phase, 96 sequences from four points of view

were used in the motor space session, and 32 sequences from four points of view in the visual space one.

The clearest result of this experiment is that the classification in motor space is easier and thus the classifier performs better on the test set. Moreover, the distribution of the data is more “regular” (the likelihood term is simpler) in motor space than in visual space. This is to be expected since the variation of the visual appearance of the hand is larger and depends strongly on the point of view, while the sequence of joint angles tends to be the same across repetitions of the same action. It is also clear that in the experiment the classifier is much less concerned with the variation of the data since this variation has been taken out by the VMM.

Overall, our interpretation of these results is that by mapping in motor space first we are allowing the classifier to choose features that are much better suited for performing optimally, which in turn facilitates generalization. The same is not true in visual space.

The interaction with the environment and the development of mirror-like representation

In order to show that, according to our model, a mirror neuron-like representation could be acquired by simply relying on the information exchanged during the interaction with the environment, we set forth to the implementation of a complete experiment on a humanoid robot called Cog (Brooks et al, 1999). This was an upper-torso human shaped robot with 22 degrees of freedom distributed along the head, arms and torso. It lacked hands, it had instead simple flippers that could use to push and prod objects. It could not move from its stand so that the objects it interacted with had to be presented to the robot by a human experimenter. The robot was controlled by a distributed parallel control system based on a real-time operating system and running on a set of Pentium based computers. The robot was equipped with cameras (for vision), gyroscopes simulating the human vestibular system, and joint sensors providing information about the position and torque exerted at each joint.

Since the robot did not have hands, it could not really grasp objects from the table. Nonetheless there are other actions that can be employed in exploring the physical

properties of objects, such as touching, poking, prodding, and sweeping. Moreover, since the interaction of the robot's flipper with objects was limited, we employed rolling objects that show a characteristic behavior depending on how they are approached: a toy car, an orange juice bottle, a ball, and a colored toy cube. The robot's motor repertoire besides reaching consisted of four different stereotyped approach movements covering a range of directions of about 180 degrees around the object. The sequence of images acquired during reaching for the object, the moment of impact, and the effects of the action were measured following the approach in (Fitzpatrick, 2003) and (Fitzpatrick and Metta, 2003).

The experiment consisted in presenting repetitively each of the four objects mentioned above to the robot. During this stage also other objects were presented at random; the experiment ran for several days and sometimes people walked by the robot and managed to make it poke the most disparate objects. For each successful trial, the robot "stored" the result of the segmentation of the object from the background, the object's principal axis which was selected as representative shape parameter, the action – initially selected randomly from the set of four approach directions –, and the movement of the center of mass of the object for some hundreds milliseconds after the impact was detected. We grouped data belonging to the same object by employing a color based clustering technique. In fact in our experiments the toy car was mostly yellow in color, the ball violet, the bottle orange, etc.

It is possible to describe object behavior in visual terms by estimating the probability of observing object motion relative to the object's own principal axis. Intuitively, this gives information about the rolling properties of the different objects: e.g. the car tends to roll along its principal axis, the bottle at right angle with respect to the axis. For the purpose of generating actions a description of the geometry of poking is required and has to go with the description of the object rolling behavior. This can be easily obtained by collecting many samples of generic poking actions and estimating the average direction of displacement of the object.

Having a visual and a "pragmatic" description of objects, it is now possible to test whether this information can be re-used to make the robot "optimally" poke (i.e. selecting

an action that causes maximum displacement) a known object. In practice the same color clustering procedure is used for localizing and recognizing the object, to determine its orientation on the table, its affordance, and finally to select the action that it is most likely to elicit the principal affordance (roll).

A simple qualitative test of the performance determined that out of 100 trials the robot made 15 mistakes. Further analysis showed that 12 of the 15 mistakes were due to poor control of reaching (e.g. the flipper touched the object too early bringing it outside the field of view), and only three to a wrong estimate of the orientation.

Although crude, this implementation shows that with little pre-existing structure the robot could acquire the crucial elements for building knowledge of objects in terms of their affordances. Given a sufficient level of abstraction, our implementation is close to the response of canonical neurons in F5 and their interaction with neurons observed in AIP that respond to object orientation (Sakata et al, 1997).

At this point, we can test whether knowledge about object directed actions can be reused in interpreting observed actions performed perhaps by a human experimenter. According to our model, whereas the robot identified the motion of the object because of a certain action applied to it, during action observation it could backtrack and derive the type of action from the observed motion of the object. In fact, the same segmentation procedure could visually interpret poking actions generated by a human as well as those generated by the robot.

Thus, observations can be converted into interpreted actions. The action whose effects are closest to the observed consequences on the object (which we might translate into the goal of the action) is selected as the most plausible interpretation given the observation. Most importantly, the interpretation reduces to the interpretation of the “simple” kinematics of the goal and consequences of the action rather than to understanding the “complex” kinematics of the human manipulator. The robot understands only to the extent it has learned to act.

In order to test this possibility we verified whether the robot can imitate the “goal” of a poking action. The step is indeed small since most of the work is actually in

interpreting observations. Imitation was generated in the following by replicating the latest observed human movement with respect to the object and irrespective of its orientation. For example, in case the experimenter poked the toy car sideways, the robot imitated him/her by pushing the car sideways (for further details, see Metta & Fitzpatrick, 2003).

In summary, the results from our experiments seem to confirm two facts of the proposed model: first, that motor information plays a role in the recognition process – as would be following the hypothesis of the implication of feedback signals into recognition – and, second, that a mirror-like representation can be developed autonomously on the basis of the interaction between an individual and the environment.

Conclusions

In this paper, starting from known experiments in the monkey, we reviewed the evidence for the existence of a mirror-neurons system in humans. We highlighted the fact that the mirror system is not a passive observer that only “resonates” with the incoming sensory stimulation but rather it works in predicting the future course of action, in filling gaps, and in merging the available evidence for the plausibility of the ongoing observed action. This last aspect includes contextual information which in turn represents the goal of the action and eventually its meaning. We also reviewed the link between the mirror system and speech drawing a parallel between vision and sound but ultimately showing that both impinge on the motor system. Finally, we covered, although briefly, some computation modeling of the mirror system which can be used to clarify certain aspects of the functioning of the biological counterpart. Although still partial, this implementation shows that, in principle, the acquisition of the mirror neurons structure is the almost natural outcome of the development of a control system for grasping. Also, we have put forward a plausible sequence of learning phases involving the interaction between canonical and mirror neurons. This, we believe, is well in accordance with the evidence gathered by neurophysiology. In conclusion, we have embarked in an investigation that is

somewhat similar to Liberman's artificial speech recognition attempts. Perhaps, also this time, the mutual rapprochement of neural and engineering sciences might lead to a better understanding of brain functions.

Acknowledgements

This work has been supported by European Commission grants ROBOT-CUB, NEUROBOTICS and CONTACT, by Italian Ministry of Education grants and by Fondazione Cassa di Risparmio of Ferrara.

References

- Arbib, M.A. (1997) From visual affordances in monkey parietal cortex to hippocampo-parietal interactions underlying rat navigation. *Philos Trans R Soc Lond B Biol Sci*, 352, 1429-1436.
- Barbas, H. and Pandya, D.N. (1987) Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J. Comp. Neurol.*, 256, 211–228.
- Binkofski, F., Buccino, G., Posse, S., Seitz, R.J., Rizzolatti, G. and Freund, H. (1999) A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. *Eur J Neurosci*, 11, 3276-3286.
- Borroni, P., Montagna, M., Cerri, G. and Baldissera, F. (2005) Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement. *Brain Res*, 1065, 115-124.
- Broca P (1861). Remarques sur le Siége de la Faculté du Langage Articulé, Suivies d'une Observation d'aphemie (Perte de la Parole). *Bulletin de la Société Anatomique de Paris*, 6, 330-57.
- Brooks, R.A., Breazeal, C., Marjanovi, M., Scassellati, B. and Williamson, M.M. (1999) The Cog project: Building a humanoid robot. In Nehaniv, C.L. (ed), *Computation for Metaphors, Analogy and Agents*. Springer-Verlag, 52-87.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G. and Freund, H.J. (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci*, 13, 400-404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A. and Rizzolatti, G. (2004) Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. *J Cogn Neurosci*, 16, 114-126.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M.J. and David, A.S. (2001) Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res Cogn Brain Res*, 12, 233-243.

Cochin, S., Barthelemy, C., Lejeune, B., Roux, S. and Martineau, J. (1998) Perception of motion and qEEG activity in human adults. *Electroencephalogr Clin Neurophysiol*, 107, 287-295.

Craighero, L., Bonetti, F., Massarenti, L., Canto, R., Fabbri Destro M. and Fadiga L. (in press) Temporal prediction of touch instant during observation of human and robot grasping. *Brain Research Bulletin*.

Decety, J., Grezes, J., Costes, N., Perani, D., Procyk, E., Grassi, F., Jeannerod, M. and Fazio, F. (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, 120, 1763-1777.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (1992) Understanding motor events: a neurophysiological study. *Exp Brain Res*, 91, 176-180.

Emmorey, K.D. (1989) Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4, 73-92.

Fadiga, L., Craighero, L., Buccino, G. and Rizzolatti, G. (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci*, 15, 399-402.

Fadiga, L., Craighero, L., Fabbri Destro, M., Finos, L., Cotillon-Williams, N., Smith, A.T. and Castiello, U. (2006) Language in Shadow. *Social Neuroscience* 1, 77-89.

Fadiga, L., Fogassi, L., Pavesi, G. and Rizzolatti, G. (1995) Motor facilitation during action observation: a magnetic stimulation study. *J Neurophysiol*, 73, 2608-2611.

Fadiga, L., Roy A.C., Fazio P. and Craighero L. (in press) From hand actions to speech: Evidence and speculations. In P. Haggard, Y. Rossetti, and M. Kawato (Eds.) *Sensorimotor Foundations of Higher Cognition. Attention and Performance*. Oxford: Oxford University Press.

Fitzpatrick, P. (2003) *From First Contact to Close Encounters: A developmentally deep perceptual system for a humanoid robot*. Unpublished PhD thesis, MIT, Cambridge, MA.

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain*, 119, 593-609.

Gangitano, M., Mottaghy, F.M. and Pascual-Leone, A. (2001) Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, 12, 1489-1492.

Gangitano, M., Mottaghy, F.M. and Pascual-Leone, A. (2004) Modulation of premotor mirror neuron activity during observation of unpredictable grasping movements. *Eur J Neurosci*, 20, 2193-2202.

Grafton, S.T., Arbib, M.A., Fadiga, L. and Rizzolatti, G. (1996) Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp Brain Res*, 112, 103-111.

Grezes, J., Armony, J.L., Rowe, J. and Passingham, R.E. (2003) Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage*, 18, 928-937.

Grezes, J., Costes, N. and Decety, J. (1998) Top down effect of the strategy on the perception of human biological motion: a PET investigation. *Cognitive Neuropsychology*, 15, 553-582.

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C. and Buchel, C. (2003) The human action recognition system and its relationship to Broca's area: an fMRI study. *Neuroimage*, 19, 637-644.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S. and Rizzolatti, G. (1998) Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proc Natl Acad Sci U S A*, 95, 15061-15065.

Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S. and Grafton, S.T. (2003) Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron*, 39, 1053-1058.

Kawato, M., Furukawa, K. and Suzuki, R. (1987) A hierarchical neural-network model for control and learning of voluntary movement. *Biol Cybern*, 57, 169-185.

Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.J. and Sirigu, A. (2004) Motor activation prior to observation of a predicted movement. *Nat Neurosci*, 7, 1299-1301.

Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. (2002) Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297, 846-848.

- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lieberman, A.M. and Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Lieberman, A.M. and Whalen, D.H. (2000) On the relation of speech to language. *Trends Cogn Sci*, 4, 187-196.
- Luppino, G. and Rizzolatti, G. (2000) The Organization of the Frontal Motor Cortex. *News Physiol Sci*, 15, 219-224.
- Matelli, M., Luppino, G. and Rizzolatti, G. (1985) Patterns of cytochrome oxidase activity in the frontal agranular cortex of the macaque monkey. *Behav Brain Res*, 18, 125-136.
- Metta, G., and Fitzpatrick, P. (2003) Early Integration of Vision and Manipulation. *Adaptive Behavior*, 11, 109-128.
- Metta, G., Sandini, G., Natale, L., Craighero, L. and Fadiga L. (2006) Understanding mirror neurons: A bio-robotic approach. *Interaction Studies*, 7:197-232.
- Miall, R.C., Weir, D.J., Wolpert, D.M. and Stein, J.F. (1993) Is the Cerebellum a Smith Predictor? *Journal of Motor Behavior*, 25, 203-216.
- Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V. and Rizzolatti, G. (1997) Object representation in the ventral premotor cortex (area F5) of the monkey. *J Neurophysiol*, 78, 2226-2230.

Nelissen K, Luppino G, Vanduffel W, Rizzolatti G, Orban GA (2005). Observing others: multiple action representation in the frontal lobe. *Science*, 310, 332-336.

Petrides, M. and Pandya, D.N. (1994) Comparative architectonic analysis of the human and the macaque frontal cortex. In Boller, F., Spinnler, H. and Hendler, J.A. (eds), *Handbook of neuropsychology*. Elsevier, Amsterdam.

Petrides, M. (2006) Broca's area in the human and the nonhuman primate brain. In Y. Grodzinsky and K. Amunts (Eds.), *Broca's Region*, Oxford University Press, New York, pp. 31-46.

Rizzolatti, G. and Arbib, M.A. (1998) Language within our grasp. *Trends Neurosci*, 21, 188-194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. and Matelli, M. (1988) Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp Brain Res*, 71, 491-507.

Rizzolatti, G. and Craighero, L. (2004) The mirror-neuron system. *Annu Rev Neurosci*, 27, 169-192.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D. and Fazio, F. (1996b) Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp Brain Res*, 111, 246-252.

Rizzolatti, G., Fogassi, L. and Gallese, V. (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci*, 2, 661-670.

Rizzolatti, G. and Gentilucci, M. (1988) Motor and visual-motor functions of the premotor cortex. . In Rakic, P. and Singer, W. (eds), *Neurobiology of neocortex*. Wiley, Chichester, 269-284.

Sakata, H., Taira, M., Kusunoki, M., Murata, A. and Tanaka, Y. (1997) The TINS Lecture. The parietal association cortex in depth perception and visual control of hand action. *Trends Neurosci*, 20, 350-357.

Salmelin, R. and Hari, R. (1994) Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. *Neuroscience*, 60, 537-550.

Strafella, A.P. and Paus, T. (2000) Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport*, 11, 2289-2292.

Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C. and Rizzolatti, G. (2001) I know what you are doing. a neurophysiological study. *Neuron*, 31, 155-165.

Watkins, K.E., Strafella, A.P. and Paus, T. (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41, 989-994.

Wolpert, D.M. and Miall, R.C. (1996) Forward Models for Physiological Motor Control. *Neural Networks*, 9, 1265-1279.

Figure legends

Figure 1. Lateral view of monkey left hemisphere. Area F5 occupies the caudal bank of the arcuate sulcus (green) and the convexity immediately posterior to it (orange). Area F5 is bidirectionally connected with the inferior parietal lobule (areas AIP - anterior intra-parietal, PF and PFG). Within the frontal lobe, area F5 is connected with hand/mouth representations of primary motor cortex (area F1, labelled in bold in the figure).

Figure 2. A typical F5 mirror neuron discharging during the visual presentation of the same grasping movement, in the left visual hemifield (leftmost panel), centrally (central panel) or in the right visual hemifield (rightmost panel). Note the substantial equivalence of the responses. The vertical bars across rasters and histograms indicate the instant at which the experimenter touched the object. Ordinates, spikes per second. Histograms bins, 20 ms. (modified from Gallese et al, 1996)

Figure 3. A, Simplified model of action representation. The actions level (area F5) is represented in the leftmost part of the figure. Action representations, if activated, activate in turn a set of motor synergies, here depicted in orange (mov, F1 level). Action execution does not produce consequences only on the external environment. Indeed, a series of afferent signals come back, from the periphery to the brain. These proprioceptive, visual, auditory signals (Perceivable consequences, in the figure), are constantly monitored by the brain and used to control the development of the ongoing action, signalling also the goal achievement. The hypothesis we suggest is that proprioceptive and motor information, biologically invariant by definition during the actuation of a same motor command, are used by the brain to generalize (and to validate) the visual inputs related to the ongoing action. These visual inputs, that continuously vary depending on the position of the head with respect to the acting hand, are forcedly considered as homologues because they are generated by the same (or very similar) motor program.

B, Simplified model of action recognition. Two individual 'brains' are shown, each one organized according to the scheme of A. When the individual on the left grasps an object

her motor system receives a visual description of the ongoing movement that could be used to control its correct execution. At the same time, however, the observer's 'brain' on the left sees the same scene (with some changes of perspective). Due to the visuomotor coupling she created to control her own movements through the process previously described, this visual representation of the seen action gains the access to the correspondent motor representation (following the dotted line). This is, in our view, the 'recognition' played by mirror neurons

Figure 4. Cortical activation pattern during observation of animal hand shadows and real animals. Significantly activated voxels ($P < 0.001$) in the moving animal shadows (red clusters) and moving real animals (green clusters) conditions after subtraction of the static controls. In the middle part of the figure, the experimental time-course for each contrast is shown (i.e. C1, moving; C2, static). Note the almost complete absence of frontal activation for real animals in comparison to animal shadows, which bilaterally activate the inferior frontal gyrus (arrows). (Modified from Fadiga et al, 2006).

Figure 5. Reaction times (RTs, msec \pm SEM) for the lexical decision during the phonological priming task with and without transcranial magnetic stimulation (TMS) between prime and target. Solid line: conditions in which prime and target share a rhyme. Dashed line: no rhyme. W-W, prime-word/target-word; W-PW, prime-word/target-pseudoword; PW-W, prime-pseudoword/target-word; PW-PW, prime-pseudoword/target-pseudoword.

Figure 6. Model schematics (see text for explanations). A, organization of the motor system during execution. B, the same model working during action observation.

	Motor space session	Visual space session
	<i>Training</i>	
# of sequences	24 (+ VMM)	64
# of points of view	1	4
Classification rate (on the training set)	98%	97%
	<i>Test</i>	
# of sequences	96	32
# of points of view	4	4
Classification rate	97%	80%

Table 1. Summary of the results of the experiment testing the model. Motor space session: motor information is available to the classifier. Visual space session: only visual information is used by the classifier. Training was always performed with all available sequences after partitioning the data into equallysized training and test set.

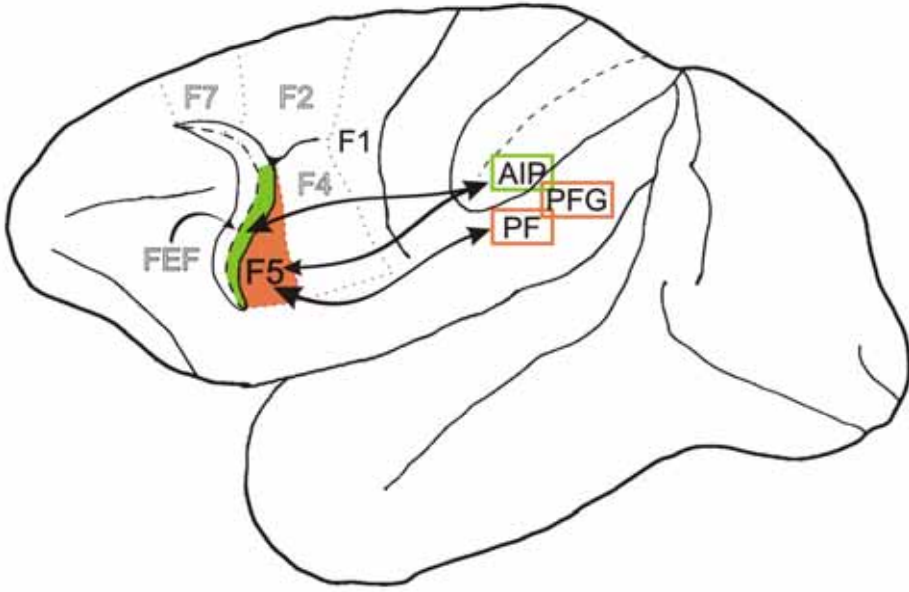


Figure 1, Craighero et al.

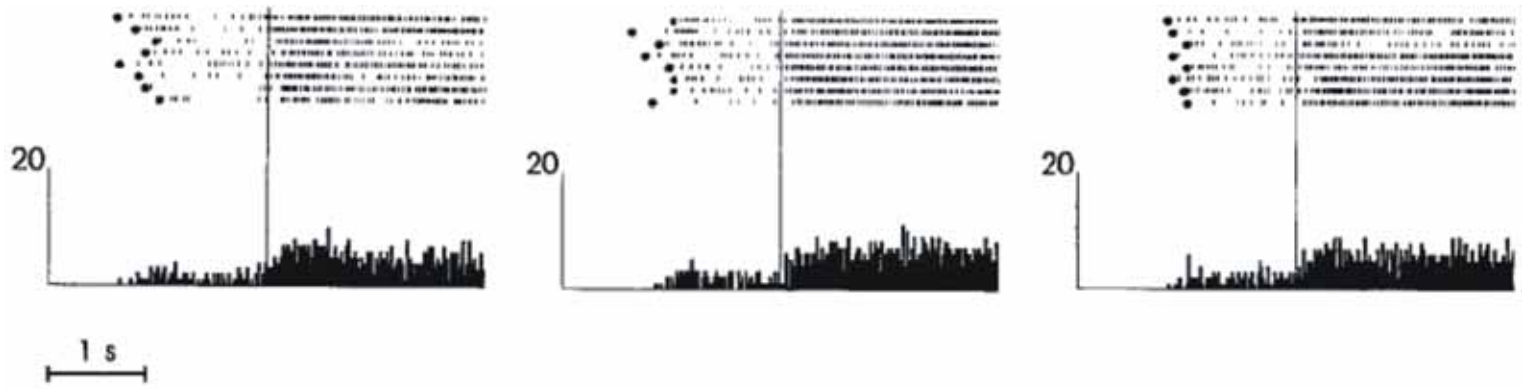


Figure 2, Craighero et al.

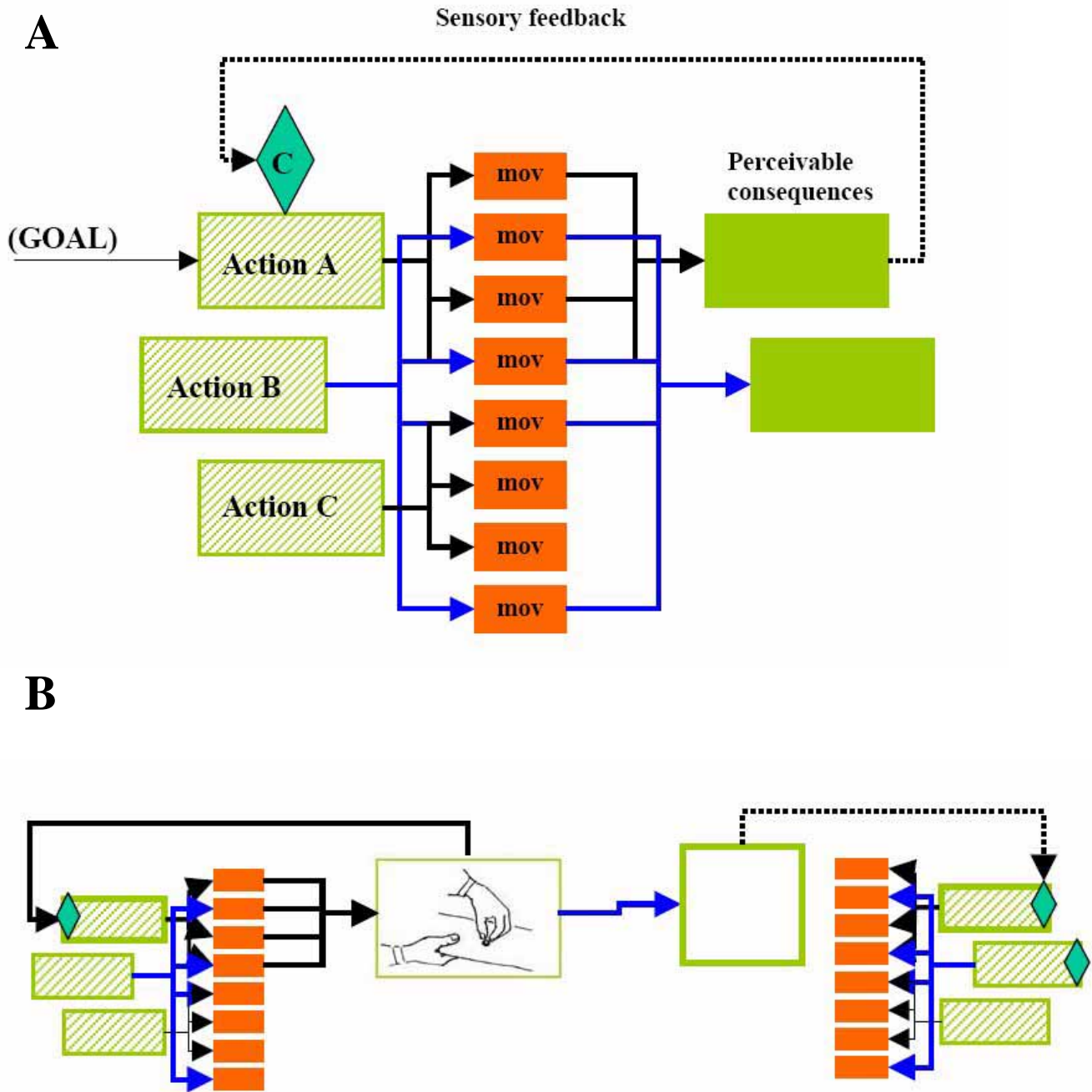


Figure 3, Craighero et al.

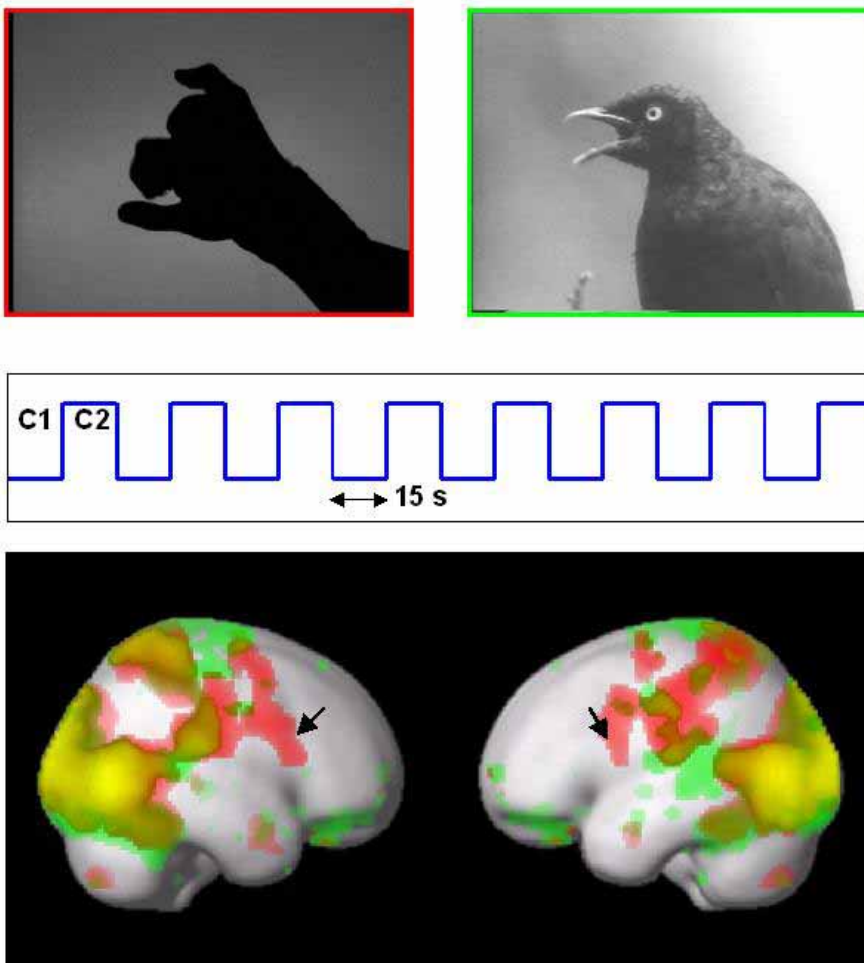


Figure 4, Craighero et al.

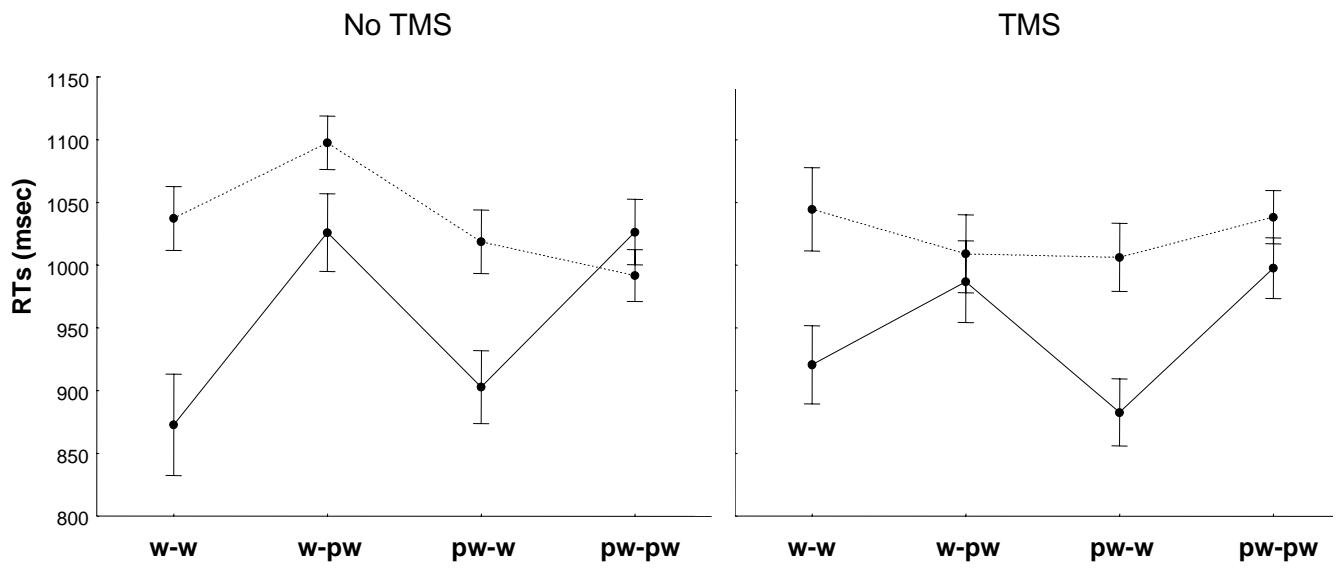


Figure 5, Craighero et al.

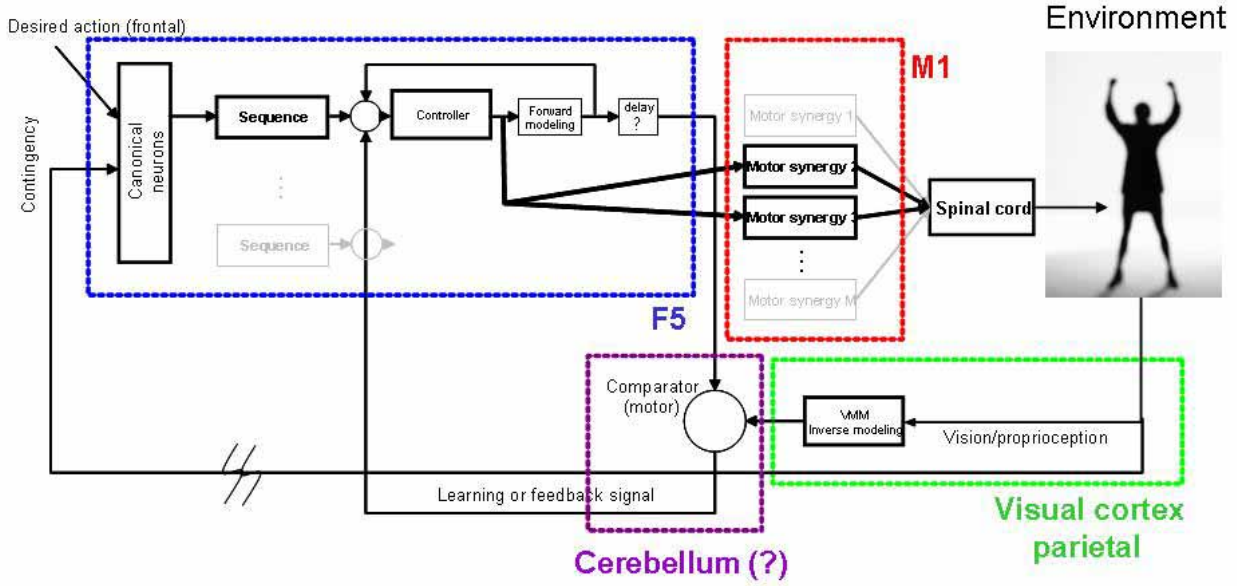
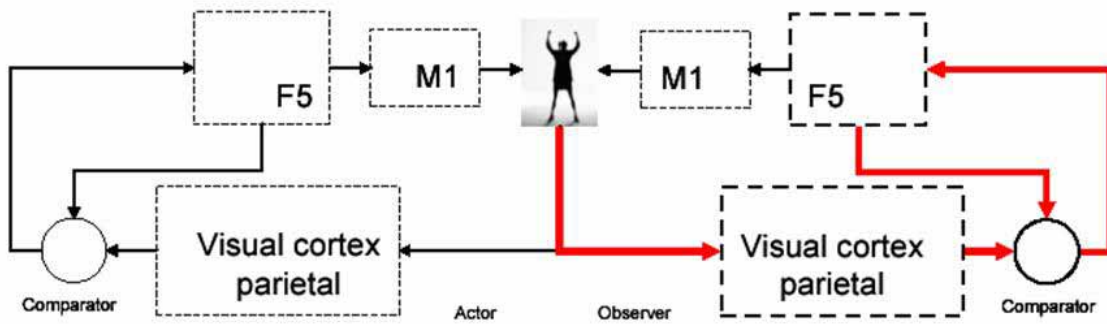
A**B**

Figure 6, Craighero et al.