CHAPTER 47

Language and mirror neurons

Giacomo Rizzolatti and Laila Craighero

47.1 Introduction

Communication is a process of exchanging information via a common system. There are many natural ways in which individuals may communicate. Besides linguistic communication, which is at the core of human communication, humans communicate using arm gestures, body postures, facial expressions, eye contact, and head and body movements.

Communication may be intentional and non-intentional. In both cases, the sender and the receiver of the messages must have a common code. The difference is that in the case of intentional communication the sender plays the leading role and imposes the communication on the receiver, while in the case of non-intentional communication, the sender sends the message without having any intention to do so. The message is just there. If sender and receiver have a common code, the message reaches the receiver, regardless of the will of the sender.

Of these two types of communication, the non-intentional one is the most basic and primitive. It is evolutionarily necessary because, in social life, individuals have to understand what others are doing, whether or not those others intend to be understood. As will be argued later, it is very plausible that intentional communication is an evolutionarily late development of non-intentional communication.

47.2 Neurophysiological basis of non-intentional communication

47.2.1 General considerations

Actions done by others are probably the most important stimuli of our lives. Most of others' actions do not convey intentional information to the observer. From them, however, we understand what others are doing and we can infer why they are doing it. This involuntary communication is fundamental for interpersonal relations, and is at the basis of social life.

What is the mechanism underlying our capacity to understand others' actions? The traditional view is that actions done by others are understood in the same way as other visual stimuli. Thus, action understanding is based on the visual analysis of the different elements that form an action. For example, when we observe a girl picking up a flower, the analyzed elements would be her hand, the flower, and the movement of the hand towards the flower. The association of these elements and inferences about their interaction enables the observer to understand the witnessed action. The discovery of neurons that code selectively biological motion has better specified the neural basis of this recognition mechanism (Perrett et al., et al. 1989). Actions done by others are coded in a
specific part of the visual system devoted to this task, which in humans includes some extrastriate visual areas and the region of the superior temporal sulcus (STS) (see Allison et al., 2000; Puce and Perrett, 2003).

It has been recently argued that visual information alone does not provide a full understanding of the observed action (Rizzolatti et al., 2001; Jeannerod, 2004). According to this view, a full action understanding can be achieved only when the observed action activates the corresponding motor representation in the observer. In this way, the observed action enters into the observer’s motor network, and enables him or her to relate it to other similar actions and to actions that usually follow the observed one. Without the involvement of the motor system, the visual representation of the action may lead to its recognition, but critical elements for understanding what the action is about (e.g. how it refers to other actions and how to reproduce it) are lacking.

These theoretical considerations received strong support from the discovery that in the motor cortex of the macaque monkey there is a particular set of neurons that discharge both when the monkey observes a given motor act and when it does the same act. These neurons, called “mirror neurons,” represent a system that directly matches observed and executed actions. Their properties will be described in the next section.

### 47.2.2 Mirror system in monkeys

Mirror neurons were originally discovered in the rostral part of the ventral premotor cortex (area F5, see Figure 47.1) of the macaque monkey. Like all neurons of this area, mirror neurons have motor properties. They code mostly distal hand actions such as grasping, holding, tearing, and manipulating. Their defining functional characteristic is that they become active not only when the monkey does a particular action (like grasping an object) but also when it observes another individual (monkey or human) performing a similar action. Mirror neurons do not respond to the sight of a hand mimicking an action or to meaningless intransitive movements. Similarly, they do not respond to the observation of an object alone, even when it is of interest to the monkey (Gallese et al., 1996; Rizzolatti, Fadiga, Fogassi, and Gallese, 1996).

The vast majority of F5 mirror neurons show a marked similarity between the observed

---

**Figure 47.1** Lateral view of the monkey cerebral cortex. The motor areas (F1–F7) are classified according to Rizzolatti et al. (1998), while the parietal areas are named according to the nomenclature of Von Economo (1929). Mirror neurons have been described in area F5 and in the rostral part of the inferior parietal lobule (areas PF and PFG). Abbreviations: AI, inferior arcuate sulcus; AS, superior arcuate sulcus; C, central sulcus. (From Rizzolatti and Craighero, 2004.)
action effective in triggering them and the executed action effective in triggering them. This sensory–motor congruence is occasionally extremely strict. In these cases the effective motor action and the effective observed action coincide both in terms of goal (e.g. grasping) and in terms of how the goal is achieved (e.g. precision grip). For most mirror neurons, however, the congruence is broader and is confined to the goal of the action.

Early studies on mirror neurons examined the upper sector of F5, where hand actions are mostly represented. Recently, a study was carried out on the properties of neurons located in the lower part of F5 where neuron activity is mostly related to mouth actions (Ferrari et al., 2003). The results showed that two classes of mouth mirror neurons could be distinguished: ingestive and communicative mirror neurons. Ingestive mirror neurons, which represent the majority of mouth mirror neurons, respond to the observation of actions related to ingestive functions (e.g. grasping food with the mouth). Virtually all of them show a good correspondence between the effective observed and the effective executed action. More intriguing are the properties of the communicative mirror neurons. For them, the most effective observed action is a communicative gesture, such as lip smacking. However, most of them strongly discharge when the monkey actively performs an ingestive action (Figure 47.2, neuron 76). A few are also active during monkey communicative gestures (Figure 47.2, neuron 33).

This discrepancy between the effective visual input (communicative) and the effective active action (ingestive) is rather intriguing. There is, 

![Figure 47.2](image)

**Figure 47.2** Examples of two mouth mirror neurons. Neuron 76 responds to the observation of a communicative gesture (lip smacking). Its motor activity is related, however, to monkey’s food ingestion. Neuron 33 responds to the observation of a communicative gesture (lip protrusion) and discharges when the monkey does a communicative gesture (lip smacking). (Modified from Ferrari et al., 2003.)
however, evidence suggesting that in evolution, monkey communicative gestures—or at least some of them—derived from ingestive actions (see below). From this perspective, one may argue that the communicative mouth mirror neurons found in F5 reflect a process of corticalization of communicative functions not yet freed from their original ingestive basis.

An issue recently addressed was whether mirror neurons are able to recognize actions from their sound. Kohler et al. (2002) recorded F5 mirror neuron activity while the monkey was observing a “noisy” action (e.g. ripping a piece of paper), or was presented with the sound of the action without seeing it. The results showed that about 15 per cent of mirror neurons responsive to presentation of actions accompanied by sounds also responded to the presentation of the sound alone. Most of them discharged specifically to the sound typical of the observed action. These neurons were dubbed “audiovisual” mirror neurons. The properties of audiovisual neurons strongly suggest that the mirror-neuron system is involved in action recognition whatever the modality through which the action is presented.

### 47.2.3 Monkey mirror neuron system and intention recognition

The motoric coding of the actions of others has an important consequence: It enables the observer to infer future acts of the observed agent by using links among motor acts in his or her motor repertoire. The existence of such an inference mechanism was recently tested by studying the activity of neurons of the inferior parietal lobule (IPL).

As shown by Mountcastle (Mountcastle et al., 1975) and Hyvarinen (1982), many IPL neurons discharge in association with specific motor acts. In a recent study, single neurons were recorded from IPL and tested for their motor properties. Neurons discharging selectively in association with grasping movements were tested in two main conditions. In one, the monkey grasped a piece of food located in front of it and brought it to its mouth, while in the other, the monkey grasped the same object and then placed it into a container (Fogassi et al., 2005). The results showed that for most recorded neurons (about 75 per cent), the discharge during grasping depended on the motor act which followed grasping. A series of control experiments demonstrated that this selectivity was not dependent on force, movement kinematics, or type of grasped stimulus. This selectivity of IPL grasping mirror neurons may appear, at first glance, as uneconomical. In fact, fewer neurons are needed if a single set of grasping neurons are used for all possible actions in which grasping is required. However, the cost in number of neurons is offset by the benefits of efficiency: the IPL organization just described is extremely well suited for providing fluidity in action execution. Each neuron codes a specific motor act and simultaneously, being embedded in a specific chain of motor acts, facilitates execution of the next act. This enables a continuous sequence of motor acts without any pause between them.

In the same experiment, a series of IPL grasping mirror neurons were tested for their visual responses in the two conditions used in the motor task. The actions were performed by one of the experimenters in front of the monkey. The first motor act was always the same—grasping—but there were cues (e.g. presence or absence of the container) that allowed the monkey to predict which action was likely to follow. The results were analogous to those found on the motor side. Namely, for the majority of IPL mirror neurons, the discharge during grasping done by the experimenter depended on whether the following act was bringing the food to the experimenter’s mouth or placing it into the container.

These data suggest a simple mechanism that may mediate the capacity of individuals to read the intentions of others. When an individual starts the first motor act of an action, he or she has clear what is the goal for the entire action. Action intention is set before the beginning of the movements, and, as shown by IPL motor properties, is reflected immediately in the neuron discharge. Because the neurons coding motor acts in different action chains have mirror properties, their activation signals not only grasping but, more specifically, grasping-for-eating, or grasping-for-placing. Thus, by their activation, the monkey is able to predict the goal of the observed action, and in this way to “read” the intention of the acting individual.

### 47.2.4 Mirror system in humans: neurophysiological evidence

A large amount of data indicate that in humans, the observation of actions done by others...
activates cortical areas involved in motor control. Evidence for this effect comes from electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings from motor areas, transcranial magnetic stimulation experiments (TMS), and brain imaging studies.

It is well known from EEG studies that active movements of the recorded individual, such as wrist opening/closure, desynchronize the EEG rhythms of the areas around the central sulcus. EEG and MEG studies showed a similar desynchronization, although less marked, during the observation of action by others (Altschuler et al., 1997; Altschuler et al., 2000; Hari et al., 1998; Cochin et al., 1998; Cochin et al., 1999). A desynchronization of cortical rhythms during observation and execution of finger movements was recently also shown using implanted subdural electrodes, a technique that allows a better signal localization than surface EEG. Most interestingly, the desynchronization during the observation and execution of these movements was present in functionally delimited hand and language motor areas (Tremblay et al., 2004).

Fadiga et al. (1995) recorded motor evoked potentials (MEPs) from the right arm and hand muscles (in normal volunteers), elicited by transcranial magnetic stimulation (TMS) of the hand field of the left primary motor cortex. By using this technique, they were able to assess the motor cortex excitability in various conditions, and thus to establish whether the motor system was activated during action observation. Stimulation was performed while volunteers were observing an experimenter grasping objects (transitive hand actions) or performing meaningless arm gestures (intransitive arm movements). Detection of the dimming of a small spot of light and presentation of objects were used as control conditions. The results showed that the observation of both transitive and intransitive actions produced an increase in the MEPs. For both hand and arm movements, the increase selectively concerned those muscles that the volunteers used when producing the observed movements.

Similar results were also found by other groups (Brighina et al., 2000; Gangitano et al., 2001; Clark et al., 2004; see Fadiga et al., 2005). Among these studies, a particularly interesting one is Gangitano et al. (2001), which showed the presence of a strict temporal coupling between the changes in cortical excitability and the dynamics of the observed action. MEPs, which were recorded from hand muscles at different time intervals during passive observation of grasping, matched the time-course of the kinematics of the observed action.

The TMS results are particularly important for two reasons. First, they show that the observation of actions performed by others specifically activates the sets of neurons that are used to replicate the observed action. Second, they indicate that, unlike in the monkey, the human mirror neuron system also resonates for intransitive actions. Thus, the human mirror system appears to have the capacity, absent in monkeys, to replicate internally actions performed by others even when these actions have no apparent goal.

47.2.5 Mirror system in humans: brain imaging studies

Observation of actions done by others activates a complex network formed, in humans, by occipital and temporal areas and by two cortical regions whose function is predominantly or fundamentally motor (e.g. Rizzolatti et al., 1996; Grafton et al., 1996; Grèzes, Costes, and Decety, 1998; Iacoboni et al., 1999; 2001; Nishitani and Hari, 2000; 2002; Grèzes et al., 2001; Perani et al., 2001; Decety et al., 2002; Grèzes et al., 2003; Koski et al., 2003; Koski et al., 2002; Manthey et al., 2003). These two regions are the rostral part of the inferior parietal lobule and the lower part of the precentral gyrus plus the posterior part of the inferior frontal gyrus (IFG) which basically corresponds to area 44. These regions form the core of the human mirror neuron system (Figure 47.3). Occasionally, other regions are also found to be active during the observation of others’ actions. These regions are the dorsal premotor cortex and the sector of IFG most likely corresponding to area 45.

The areas forming the human mirror neuron system show a somatotopic organization. In an fMRI study, normal volunteers were presented with video clips showing actions done with the mouth, hand/arm and foot/leg. Action observation was contrasted with the observation of a static face, hand, and foot respectively (Buccino et al., 2001). Observations of object-related mouth actions determined signal increase in the lower part of the precentral gyrus and of the pars opercularis of IFG, bilaterally. In addition, two activation foci were found in the parietal lobe. One was located in the rostral part of the...
inferior parietal lobule (most likely area PF), while the other was located in the posterior part of the same lobule. Observations of object-related hand/arm actions determined signal increase in two sectors of the frontal lobe, one located in the pars opercularis of IFG, and the other located in the precentral gyrus. The latter activation was more dorsally located than that found during the observation of mouth movements. As in the case of mouth actions, two activation foci were found in the parietal lobe. The rostral focus was in the rostral part of the inferior parietal lobule, but more posteriorly located than that observed during mouth actions, while the caudal focus was essentially in the same location as that for mouth actions. Finally, observations of object-related foot/leg actions determined an activation of a dorsal sector of the precentral gyrus and an activation of the posterior parietal lobe.

In the second part of the same experiment, volunteers were presented with a mimed version of the object-directed actions. Unlike in the monkey, in which mimed actions do not excite mirror neurons, in humans, mimed actions produced a signal increase in the frontal sector of the mirror neuron system. In contrast, no activation was found in the inferior parietal lobule. This finding is in accord with lack of inferior parietal lobule activation found also in other studies in which intransitive actions were used (e.g. finger movements; Iacoboni et al., 1999; 2001; Koski et al., 2002; 2003). Considering that visual information reaches premotor cortex through the inferior parietal lobule, it is hard to believe that the observation of intransitive actions does not produce any activation of this structure. Instead, activation may be more difficult to detect for intransitive actions compared to transitive actions for the following reason. The inferior parietal lobe receives direct information from the inferotemporal cortex, in addition to that from visual areas responsive to biological motion. Thus it is likely that the presence of objects in transitive actions produces a source of activation in addition to that due to

Figure 47.3 Lateral view of human cerebral cortex. The parietal and frontal lobe regions that are consistently activated during the observation of actions by others are shown in red and yellow, respectively. They form the core of the mirror neuron system. The areas shown in blue (area 45 and dorsal area 6) have been reported to be active during the observation of others’ action in some studies. It is likely that they also contain mirror neurons.
action observation, thereby resulting in a statistically significant activation of inferior parietal lobule for the transitive actions. In mimed actions, where there is no object, the activation does not reach statistical significance.

47.2.6 Which actions activate the mirror neuron system?

Humans recognize actions done not only by humans but also by animals. Are animal actions recognized by the mirror neuron system or is the visual system alone sufficient? Recently, Buccino et al. (2004) addressed this question in an fMRI experiment. Video clips showing silent mouth actions performed by humans, monkeys, and dogs were presented to normal volunteers. Two types of action were shown: biting food, and oral communicative actions (speech reading, lip smacking, barking). As a control, static images of the same actions were presented. The results showed that the observation of biting, regardless of whether performed by a man, a monkey, or a dog, produced a signal increase in the inferior parietal lobule and in the pars opercularis of the IFG plus the adjacent precentral gyrus. The left parietal and the left frontal activations were virtually identical for observation of all three species, whereas the right-side foci were stronger during the observation of actions made by a human. Different results were obtained with communicative actions. Observing a human speak activated the left pars opercularis of IFG, whereas the observation of lip smacking—a monkey communicative gesture—activated a small focus in the right and left pars opercularis of IFG; finally, the observation of barking did not produce any frontal lobe activation. In this case the activation was only present in visual areas.

These results show that actions done by others can be recognized through different mechanisms. Actions belonging to the motor repertoire of the observer, regardless of whether they are done by a conspecific, are mapped onto his or her motor system. Actions that do not belong to this repertoire do not excite the motor system of the observer and are recognized on a visual basis. It is likely that these two different ways of recognizing actions have two different psychological counterparts. In the first case the motor "resonance" translates the visual experience into an internal "first person knowledge," whereas this is lacking in the second case.

Further evidence that the mirror neuron system is strongly activated when the observer sees an action that is part of his or her motor repertoire has been recently provided by Calvo-Merino et al. (2005). They used fMRI imaging to study differences in brain activation between watching an action that one has learnt to do and an action that one has not. Experts in classic ballet, experts in the Brazilian martial art capoeira, and inexpert controls viewed videos showing dancers performing classical ballet or capoeira. The results showed greater bilateral activations in premotor cortex, intraparietal sulcus region, right superior parietal lobe, and left posterior STS when expert dancers viewed movements that they have been trained to perform compared to movements that they had not. These results show that the mirror system is particularly active when it integrates observed actions of others with an individual's personal motor repertoire.

The presence of activations during action observation in an area classically considered to be exclusively related to speech might be surprising. However, in recent years, clear evidence has been accumulated that human area 44 contains (as does monkey area F5, the monkey homolog of area 44) a motor representation of hand movements (Krams et al., 1998; Binkofski et al., 1999; Iacoboni et al., 1999; Gerardin et al., 2000; Ehrsson et al., 2000, 2004), in addition to that of orolaryngeal movements. It appears, therefore, that the posterior sector of IFG is involved in both non-intentional communication (i.e. recognition of object-related actions) and intentional (verbal) communication—a finding that suggests, as we will see in the next section, a link between these two forms of communication.

47.3 Mirror neurons and language

Humans mostly communicate by sounds. Sound-based languages, however, do not represent the only natural way for communicating. Languages based on gestures (signed languages) represent another form of complex, fully structured communication system. By using sign language, people express abstract concepts, learn mathematics, physics, philosophy, and even create poetry (see Corballis, 2002). Nonetheless, the fact that signed languages represent a fully
structured communication system has not changed the view, which many share, that speech is the only natural human communication system, and that the evolutionary precursor of human speech consists of animal calls. The argument goes as follows: humans emit sound to communicate, animals emit sounds to communicate, therefore human speech evolved from animal calls.

The logic of this syllogism is rather shaky. Its weakness becomes apparent when one examines animal calls and human speech more closely. First, the anatomical structures underlying primate calls and human speech are different. Primate calls are mostly mediated by the cingulate cortex and by deep, diencephalic and brain stem structures (see Jürgens, 2002). In contrast, the circuits underlying human speech are formed by areas located around the Sylvian fissure, including the posterior part of IFG. It is hard to imagine how in primate evolution, the call system shifted from its deep position found in non-human primates to the lateral convexity of the cortex where human speech is housed.

Second, speech in humans is not, or is not necessarily, linked to emotional behavior, whereas animal calls are. Third, speech is mostly a dyadic, person-to-person communication system. In contrast, animal calls are typically emitted without a well-identified receiver. Fourth, speech is endowed with combinatorial properties that are absent in animal communication. As Chomsky (1966) rightly stressed, human language is “based on an entirely different principle” from all other forms of animal communication. Finally, humans do possess a “call” communication system like that of non-human primates and its anatomical location is similar. This system mediates the utterances that humans emit when in particular emotional states (cries, yelling, etc.). These utterances, which are preserved in patients with global aphasia, lack the referential character and the combinatorial properties that characterize human speech.

The advocates of the sound-based theory of language origin consider a strong argument in favor of this theory to be the presence of referential information in some animal calls (e.g. Pinker, 1994). The famous study of the alarm calls of vervet monkeys (Cheney and Seyfarth, 1990), as well as other studies that extended these observations to other species and other communicative contexts (social relationship, food, inter-group aggression), showed that evolution tried this pathway. The reason why this attempt did not succeed is the lack of flexibility inherent in any communicative system based on emotions. In a non-emotional communication system the same word, for example the word fire, which is basically an alarm message (“escape”), may assume a completely different meaning. It may indicate, for example, that the fire is ready and we can start to cook our meal (“approach message”), as well as conveying other positive messages. This flexibility cannot occur in an emotional communicative system because a referential meaning cannot indicate a behavior that is in contrast with the emotion that generated it. Thus the same utterance or call cannot convey, in different contexts, an escape and an approach message.

If not animal calls, what could be the origin of human speech? An alternative hypothesis is that the path leading to speech started with gestural communication. This hypothesis, first proposed by the French philosopher Condillac, has recently found several defenders (e.g Armstrong et al., 1995; Corballis, 2002). According to this theory, the initial communicative system in primate precursors of modern humans was based on very simple, elementary gesturing. Sounds were then associated with the gestures and became progressively the dominant way of communication.

The discovery of mirror neurons provided strong support for the gestural theory of speech origin. Mirror neurons create a direct link between the sender of a message and its receiver. Thanks to the mirror mechanism, actions done by one individual become messages that are understood by an observer without any cognitive mediation. The observation of an individual grasping an apple is immediately understood because it evokes the same motor representation in the parieto-frontal mirror system of the observer. Similarly, the observation of a facial expression of disgust is immediately understood because it evokes the same representation in the amygdala of the individual observing it (Gallese et al., 2004).

On the basis of this fundamental property of mirror neurons, and the fact that the observation of actions like hand grasping activates the caudal part of IFG (Broca’s area), Rizzolatti and Arbib (1998) proposed that the mirror mechanism is the basic mechanism from which language evolved. In fact, the mirror mechanism
solved, at a initial stage of language evolution, two fundamental communication problems: parity and direct comprehension. Thanks to the mirror neurons, what counted for the sender of the message also counted for the receiver. No arbitrary symbols were required. The comprehension was inherent in the neural organization of the two individuals.

A criticism of this view is based on the fact that the monkey mirror neuron system is constituted of neurons coding object-directed actions. Thus, the monkey mirror neuron system forms a closed system, which by definition does not appear to be particularly suitable for intentional communication. Yet, if this is true for the monkey, it is not the case for the human mirror system. As reviewed above, TMS and brain imaging studies have shown that activation of the human mirror system is achieved by presentation of intransitive actions (Fadiga et al., 1995; Maeda et al., 2002) as well as during pantomime observation (Buccino et al., 2001, Grèzes et al., 2003).

It is difficult to specify how the shift from a closed system of monkeys to an open, intentionally communicative system, in humans might have occurred. The view, however, that communicative actions derived from a more ancient system of non-communicative gestures is not new. Van Hoof (1967), for example, proposed that many of the most common communicative gestures of the monkey, such as lip smacking, are ritualizations of ingestive actions that monkeys use for affiliative purposes. The fact that mouth mirror neurons respond both to the observation of communicative actions and during the execution of ingestive actions appears to give a neurophysiological basis to this idea (Ferrari et al., 2003; see also above, mirror system in monkeys).

Similarly, Vygotsky (1934) suggested that intransitive actions derive in children from object-directed transitive actions. For example, when objects are located close to a child, the child grasps them. When they are located far from the child, the child extends his or her hands towards the objects. Because the mother understands this gesture, the child uses it again and again and, eventually, attempts to reach objects become communicative gestures. Thus, the transition from object-directed to intentional communicative gesture can be accommodated by the mirror neuron hypothesis of language evolution.

47.3.1 From gestures to sound

It is rather unlikely that the gestural communications, which appeared in the ancestors of Homo sapiens, reached the sophisticated complexity shown by modern sign languages. It is much more plausible, as also suggested by Arbib (2005), that gestures emitted for communicative purposes were soon accompanied by sounds, and that the speech development prevented the occurrence of a fully-fledged sign language.

The major problem in this evolutionary scenario is to understand how sounds, initially meaningless, became associated with gestures which conveyed specific meanings. Onomatopoeia, that is, the similarity between the sound of a word and the noise produced by a corresponding natural event, is one of the suggested possibilities. Another possibility is represented by interjedtional utterances emitted by individuals in certain conditions. The problem with both these hypotheses is that they can account for only a very limited number of words. Thus, although they explain the origin of some words, they lack the generality necessary to explain most of the links between sound and meaning.

An interesting theory attempting to explain this link was advanced many years ago by Paget (1930). This theory, called “schematopoeia”, posits that human communication started with manual gestures. These gestures were accompanied by unintentional, but analogous, movements of tongue, lips, and jaw. Later, the gesticulating individuals discovered that the expiration of air through the oral cavities produced audible gestures. This was the beginning of voiced speech.

Paget gives many examples of parallelism between sound and meaning from a variety of languages. For vowels, he suggests that \( a \) (as in large) often refers to anything that is large, wide open, or spacious; \( i \), especially in its narrow or thin variety (as in mini), often refers to something that is small or pointed; and \( aw \) (as in yawning) generally indicates a cavity. For consonants, he suggests that \( m \) implies a continuous closure, while \( dr \) or \( tr \) denotes running or walking. According to this explanation, the great majority of words were originally pantomimic (Paget, 1930: 159). This pantomimic origin is not readily evident in modern languages because, according to Paget, words are built by addition of separate significant elements in the manner of Chinese ideographs.
The schematopoeia theory is obviously very speculative. Its central tenet, however—that there is a basic, natural link between hand/body gestures and speech gestures—is very ingenious. It suggests a clue as to how an opaque gestural system, like the orolaryngeal system, could convey understandable messages, thanks to the close correspondence between hand/body gestures, intrinsically known to the observers, and oro/laryngeal gestures. Furthermore, it has a clear neurophysiological prediction: if hand/arm and speech gestures are strictly linked, they must have a common neural substrate.

A series of recent studies support this prediction. TMS experiments have shown that right-hand motor excitability increases during reading and during spontaneous speech, whereas no language-related effects are found under these conditions in the left-hand motor area or in the leg representations of either hemisphere (Tokimura et al., 1996; Seyal et al., 1999; Meister et al., 2003). Meister et al. (2003) stressed that the increase of hand motor cortex excitability cannot be due to word articulation because word articulation recruits motor cortex bilaterally, but the observed activation was strictly limited to the left hemisphere. The facilitation appears, therefore, to result from a co-activation of the right hand motor area with the language network.

Similar conclusions were also reached by Gentilucci and his co-workers (2001). In a series of behavioral experiments, they showed normal volunteers one of two 3-D objects, one large and the other small, randomly presented. On the visible face of both objects, either two Xs or a series of dots could appear. The volunteers were required to grasp the objects and, regardless of the symbols written on the object, to open their mouth in all conditions. The kinematics of hand, arm, and mouth movements were recorded. The data showed that in spite of the instruction to open the mouth in the same way in all conditions, lip aperture and the peak velocity of lip aperture increased when the hand movement was directed to the large object (Figure 47.4).

In another experiment, the authors employed a similar experimental procedure, but asked the participants to pronounce a syllable (e.g. gu or ga) instead of simply opening their mouths. The syllables were written on the objects in the same locations as the symbols in the previous experiment. The results showed that lip aperture was larger when the participants grasped a larger object. Furthermore, the maximal power level

![Figure 47.4](image-url)
recorded during syllable emission was also higher when volunteers grasped the large object (Gentilucci et al., 2001). These experiments indicate that both buccal movements and the orolaryngeal synergies necessary for syllable emission are linked to manual gestures. Most importantly, hand actions producing large movements share a neural substrate with large mouth actions, precisely as proposed by Paget’s theory.

Finally, evidence for a link between gesturing and speech also comes from clinical studies. Hanlon et al. (1990) showed that, in aphasics, pointing with the right hand to a screen where objects are presented facilitates object naming. Similarly, Hadar et al. (1998) found that word retrieval is facilitated through gesturing in brain-damaged patients.

Clearly, the reviewed experiments do not prove the schematopoeia theory. Nonetheless, they indicate that the theory is not as bizarre as one may think initially: a link between hand gestures and the speech system is present in modern Homo sapiens.

47.3.2 The appearance of echo-mirror neurons

The association between specific sounds and communicative gestures has obvious advantages, such as the possibility of communicating in the dark or when hands are busy with tools or weapons. Nonetheless, to achieve effective sound communication, the sounds conveying messages previously expressed by gesture (“gesture-related sounds”) ought to be clearly distinguishable and, most importantly, should maintain constant features; they must be pronounced in a precise, consistent way. This requires a sophisticated organization of the motor system related to sound production, and a rich connectivity between the cortical motor areas controlling voluntary actions and the centers controlling the oro-laryngeal tract. The large expansion of the posterior part of the inferior frontal gyrus culminating in the appearance of Broca’s area in the human left hemisphere is, most likely, the results of the evolutionary pressure to achieve this voluntarily control.

In parallel with these modifications occurring in the motor cortex, a system for understanding them should have evolved. We know that in monkey area F5, the homolog of human area 44, there are neurons—the so called “audiovisual neurons” (Kohler et al., 2002; see also above, mirror system in monkeys), that respond to the observation of actions done by others as well as to the sounds of those actions. This system, however, is tuned for recognition of the sound of physical events and not of sounds done by individuals. In order to understand the protospeech sounds, a variant of the audiovisual mirror neuron system tuned to resonate in response to sounds emitted by the orolaryngeal tract should have evolved. A more sophisticated acoustic system, enabling a better discrimination of the gesture-associated sounds, has also probably evolved. Note, however, that an improvement in auditory discrimination would be of little use if the gesture-related sounds did not activate the orolaryngeal gesture representation in the brain of the listener.

47.3.3 Echo-mirror neuron system and the problem of its functional role

Is there evidence that modern humans have a mirror neuron system that responds to sounds produced by the orolaryngeal tract and that naturally accompanies gesticulation? Considering the sophistication of the speech system in our species, it is difficult to prove this point. There is clear evidence, however, that humans are endowed with a motor system that resonates selectively in response to speech sounds (the “echo mirror neuron system”).

Fadiga et al. (Fadiga et al., 2002) recorded MEPs from the tongue muscles in normal volunteers instructed to listen carefully to acoustically presented verbal and non-verbal stimuli. The stimuli were words, regular pseudo-words, and bitonal sounds. In the middle of words and pseudo-words there was either a double \( f \) or a double \( r \). \( F \) is a labio-dental fricative consonant that, when pronounced, requires virtually no tongue movements, whereas \( r \) is a linguo-palatal fricative consonant that, in contrast, requires marked tongue muscle involvement to be pronounced. During the stimulus presentation, the left motor cortex of the participants was stimulated with single pulse TMS. The results showed that listening to words and pseudo-words there was either a double \( f \) or a double \( r \). F is a labio-dental fricative consonant that, when pronounced, requires virtually no tongue movements, whereas \( r \) is a linguo-palatal fricative consonant that, in contrast, requires marked tongue muscle involvement to be pronounced. During the stimulus presentation, the left motor cortex of the participants was stimulated with single pulse TMS. The results showed that listening to words and pseudo-words containing the double \( r \) produced a significant increase of MEP amplitude recorded from tongue muscles compared to listening to bitonal sounds and words and pseudo-words containing the double \( f \) (Figure 47.5).
Results congruent with those of Fadiga et al. (2002) were obtained by Watkins et al. (2003). Using the single pulse TMS technique, they recorded MEPs from a lip muscle (\textit{orbicularis oris}) and a hand muscle (first \textit{interosseus}) in four conditions: listening to continuous prose, viewing speech-related lip movements, listening to non-verbal sounds, and viewing eye and brow movements. Compared to viewing eye and brow movements, listening to and viewing speech enhanced the MEP amplitude recorded from the \textit{orbicularis oris} muscle. Intriguingly, the MEPs obtained when subjects listened to non-verbal sounds also increased compared to viewing eye and brow movements. This increase was very likely due to activation of a human mirror system similar to that of the monkey audio-visual mirror neurons (Kohler et al. 2002).

All of these effects were seen only in response to stimulation of the left hemisphere. No changes of MEPs in any condition were observed following stimulation of the right hemisphere. Finally, the size of MEPs elicited in the first \textit{interosseus} muscle did not differ in any condition.

Taken together, these data suggest that a mirror neuron system for speech sound—an echo-mirror neuron system—exists in humans: When an individual listens to verbal stimuli, there is an automatic activation of his speech-related motor centers. Did this system evolve from the hypothetical gesture-related sounds mirror system discussed in the previous section? There is no doubt that speech is not purely a system based on sounds as such. As shown by Liberman (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000), an efficient communication system cannot be built by substituting tones or combinations of tones for speech. There is something special about speech sounds that distinguish them from other auditory material, and this is their capacity to evoke the motor representation of the heard sounds in the listener’s motor cortex. Note that this property, postulated by Liberman on the basis of indirect evidence, is now demonstrated by the existence of the echo-mirror neuron system. One may argue, however, that this property is only for translating heard sounds into pronounced sounds. In other

\textbf{Figure 47.5} Total areas of the normalized motor evoked potentials (MEPs) recorded from tongue muscles during listening to words and bitonal sounds. “rr” and “ff” refer to verbal stimuli containing a double lingua-palatal fricative consonant “r”, and a double verbal labio-dental fricative consonant “f”, respectively. (Modified from Fadiga et al., 2002.)
words, the basic function of mirror neurons—understanding—will be lost here and only the imitation function, developed on the top of former (see Rizzolatti and Craighero, 2004) would be present.

It is possible that an echo-mirror neuron system would evolve solely for the purpose of translating heard sounds into pronounced sounds. We are strongly inclined, however, to think that the motor link that this system provides to speech sounds has a more profound evolutionary significance. First, as discussed above there is a consistent link in humans between hand actions and orolaryngeal gestures, similar to the one present in the monkey for hand and mouth action. Thus, if these neurons acquired mirror properties, as other types of F5/Broca’s area neuron did, a category of neurons evolve coding orolaryngeal tract gestures simultaneously with body-action gestures. In other words, neurons appeared which coded phonetics simultaneously with semantics. In this way, heard speech sounds produced not only a tendency to imitate the sound but also an understanding of the accompanying body-action gestures (much as audio-visual mirror neurons allow understanding of the actions that produce the sounds). Second, once a primitive sound-to-meaning linkage was established, it served as the base for the development of additional, increasingly arbitrary, links between sounds and actions—i.e. the development of words. These arbitrary links greatly extended the possibilities for rich communication while requiring a lengthy, culturally bound learning period. Finally, given the necessity of distinguishing among more speech sounds in more combinations, the links between heard speech sounds and orolaryngeal gestures became stronger (as in the modern echo-mirror system), whereas there was little pressure to further develop the link between sound and meaning given the success of the learnt system of arbitrary linkages.

47.4 Conclusions

The aim of the present chapter has been to describe the functional properties of mirror neurons systems, and to show how a mechanism originally devoted to non-intentional communication becomes in evolution the mechanism at the basis of intentional communication. The reconstruction of how speech appeared in evolution can only be hypothetical. Nonetheless, evidence collected in the last few years strongly supports the hypothesis that human speech evolved from gestural communication rather than from animal calls.

We have not addressed the problem of the origin of grammar. We recognize the importance of this issue—for many scholars the most important issue—but at present there are not sufficient neurophysiological data to discuss its origin, even at a very hypothetical level. Nor have we discussed the growing behavioral and neurophysiological evidence showing that during speech processing there is activation of non-speech areas and of the motor system in particular. These findings and their implications are dealt with in detail in the excellent chapters in this volume by Glenberg (21) and Pulvermüller (8).

References


Acknowledgements

We are grateful to Arthur Glenberg for his comment on the manuscript. The study was supported by EU Contract IST 2001–35282, Mirrorbot; EU Contract NEST 012738, Neurocom; Cofin 2004 (GR); OMLL, EURO-CORES Programme; EU Contract NEST 5010, “Learning and development of Contextual Action.”