Maurizio Gentilucci*, Claudia Gianelli†, Giovanna Cristina Campione*, and Francesca Ferri*

* Dipartimento di Neuroscienze, Università di Parma, Parma, Italy
† Dipartimento di Psicologia, Università di Bologna, Bologna, Italy

Intersubjectivity and Embodied Communication Systems

Both intersubjectivity and embodied cognitive processes are based on mechanisms for sharing actions, common to the species. The evolution of spoken language and of communication systems in general are good examples of this. In the present review, we propose that, by a process of observation and imitation, the sharing of object-directed actions (i.e., transitive arm/hand actions) and their pantomimes could have been used to progressively construct communication systems capable of representing action meaning (i.e., their goals). Starting from this process of observation and imitation, humans may have constructed progressively more complex communication systems based on iconic and symbolic arm gestures. These communication systems may have gradually been translated into more specific, buccal, gesture-based systems that gave rise to spoken language. In support of these hypotheses, we report evidence showing that the execution and observation of transitive actions and their pantomimes affects the production of phonological units. We describe the effects of the production and observation of symbolic gestures on word pronunciation. Finally, we report evidence that these systems relating gesture to speech have neural correlates in neural circuits located in the frontal cortex, understood to be involved in spoken language.

Keywords: mirror system, transitive actions, gestures, phonemic units, words, human kinematics, vocal spectra.
INTRODUCTION

Intersubjectivity emphasizes how shared cognition and consensus are essential in human relations. Spoken language is intersubjective, just because it evolved as a system for communication: that is, for sharing information. Various forms of intersubjectivity – from the most basic to the most complex – are associated with different stages of the gradual phylogenetic development finally resulting in language. All forms of intersubjectivity originate from a sharing of experience through actions, as common motor intentions or subjective states. Some behaviours such as contagion, neonatal imitation, and mutual gaze testify to the presence of primary intersubjectivity between infants and conspecifics – above all, mothers (Trevarthen 1979, Reddy 2003). Importantly, these behaviours seem also to be present in non-human primates (Ferrari et al. 2009a, 2009b). True imitation (involving self/other differentiation and understanding the goal of the action) and shared attention are higher levels of intersubjective communication, antecedent to the development of spoken language and other complex communication systems based on e.g. iconic and symbolic arm gestures. All these forms of intersubjective experience can be grounded on either a common value or shared action. The embodiment theory of language understanding suggests that neural systems for perception and action are also engaged during language comprehension (Barsalou 1999, Gallese & Lakoff 2005, Glenberg 1997, Lakoff 1987, Pulvermüller 2002, Zwaan 2004, Zwaan & Taylor 2006). Language and other communication systems are clear examples of how both intersubjectivity and embodied cognitive processes are based on shared actions.

Our hypothesis is that the intersubjective nature of spoken language is based primarily on shared representations of arm actions, common to the species. We propose that an initial form of communication based on arm-transitive actions (i.e., guided by an object) and pantomimes of transitive actions was gradually translated to mouth-articulation gestures (Gentilucci & Corballis 2006). In support of these hypotheses, we report evidence that the execution and observation of transitive actions and their pantomimes affect the production of phonological units. We then describe the effects of the observation and production of symbolic gestures on word pronunciation, suggesting that words and symbolic communicative gestures are coded as a single signal by a unique communication system within the frontal cortex. We propose that the same system of translation is also capable of merging speech and symbolic gesture into a single sign, which acquires a new meaning compared to the meaning of the word and the gesture produced separately. This is directly linked with our hypothesis about the social relevance of a similar process: the transfer from arm to mouth of the social intention to interact directly with the interlocutor produces an increase in the intersubjective nature of a word.
EXPERIMENTAL DATA

Translation between gesture and word: anatomical and neurophysiological bases

Gesture is a universal feature of human communication. Speakers of every culture produce gestures. So far as is known, gestures are produced at the same time as speech. The integration of gesture with speech suggests that gesture and speech are under the control of the same planning processes (Kendon 1972, 1980, 2004; McNeill 1992, 2000; Goldin-Meadow 1999).

The frontal regions in monkey and human cortex

In the human cortex, speech and gesture seem at first glance to be controlled by separate, albeit adjacent, systems. The frontal Brodmann Area (BA) 44 – part of Broca’s area – is involved in the control of the mouth articulation gestures by which phonological units of speech are produced (Demonet et al. 1992, Paulesu et al., 1993, Zatorre et al. 1992). BA 44 is separate from but adjacent to frontal premotor BA 6, which is well known to be involved in control of arm and face movements. Rizzolatti and Arbib (1998) propose, on cytoarchitectonic and functional grounds, that the homologue of the monkey frontal premotor area F5 is BA 44, while the homologue of the monkey frontal premotor area F4 is the human lateral BA 6. F5 is involved in controlling hand and mouth movements. Two classes of neurons, which may be the basis on which the system of arm-gesture-based communication was constructed and related to speech, are found here. The first controls grasping actions with the hand and mouth (Rizzolatti et al. 1988). The second, encompassing what are known as mirror neurons, becomes active when the animal observes grasping actions as well as when it carries them out (Gallese et al. 1996). The latter class might similarly be involved in constructing a gestural communication system in human beings: a mirror system has been found in the human frontal areas (Kilner et al. 2009, Gazzola & Keysers 2009).

Double grasp commands to hand and mouth in humans

A system of grasping commands to both the hand and mouth has been observed in humans. In a kinematics study – in which positions, angles, velocities, and accelerations of body segments and joints were measured during motion – participants were required to reach for and grasp either a small or large object while opening the mouth by an arbitrary amount constant between trials (Gentilucci et al. 2001). The different kinematics of grasping the large versus the small object affected the lip-opening kinematics. The fingers were opened more widely when grasping the large object, and the extent of lip opening was likewise greater. Neither the appearance of the object nor the different arm kinematics of reaching to grasp objects of different size accounted for the effects observed on lip opening. In a reverse setting, grasping objects of different size with the mouth, subjects adjusted the
extent to which they opened their fingers, even though they were asked to open their fingers by a fixed degree.

These studies suggest that the same command circuit is involved in controlling grasping movements of both hand and mouth. This could suggest a basis for understanding the origin of the intimate relationship between gesture and speech and the process by which a communication – first evolved as a system of manual gestures – came to be transferred to a system involving the mouth (see the following sub-section).

Many authors have proposed that language evolved from manual actions rather than vocalizations, since manual actions can provide more obvious iconic links to objects and actions in the physical world. Among the early proponents of this view were the 18th-Century philosophers Condillac (1973/1746) and Vico (1953/1744). The idea has been put forward many times, in many forms, since then (e.g., Arbib 2005; Armstrong 1999; Armstrong et al. 1995; Corballis 1992, 2002; Donald 1991; Givôn 1995; Hewes 1973; Rizzolatti & Arbib 1998; Ruben 2005).

More recently, Gentilucci and Corballis (2006) have suggested that the common ancestor of humans was much better pre-adapted to develop a voluntary communication system based on visible gestures than one based on sounds – based on the neurophysiological evidence that non-humans have little cortical control over vocalization (Jürgens 2002), which is critical to speech. Instead, they have greater control over hand movements (see the following sub-section). This ability could have arisen with the emergence of bipedalism in the hominid line, when the hands were freed from locomotion. Once the hands became completely free to bring food to the mouth, mutual adjustments between mouth and hand could be refined – suggesting a hypothesis why such an intimate relationship between the control systems for mouth and hand movements exists. The freeing of the hands from locomotion also provided a potential boost to the evolution of a manual communication system based on transitive actions. The system of double commands involved in oral/manual movements could be co-opted to transfer arm gestures into speech while the manual communication system developed.

This hypothesis is supported by a study in which we required participants to pronounce syllables printed on the visible side of a selection of large and small objects while reaching for and grasping them. Both the maximal power level of the voice and the lip aperture were affected by the different kinematics of grasping large versus small objects. During the initial phase of finger opening, both the peak velocity of finger opening and finger aperture decreased when grasping small compared to large objects (Gentilucci et al. 2001). Another study (Gentilucci et al. 2004a) found that bringing different-sized fruits to the mouth modified formant 2 of the syllable /ba/ in the voice spectra, when the syllable was pronounced simultaneously with carrying out the action. The effect on speech was selective: no effect on voice spectra was found when a vocalization unrelated to the participants’ native language was made simultaneously with bringing-to-the-mouth (Gentilucci et al. 2004a). This means that arm-transitive actions – which potentially also involve the mouth effector – can be functionally related to the articulation organs (lips; velum; larynx; and blade, body, and root of the tongue) used for speech.
The mirror system and speech

In the monkey premotor area F5, another class of neurons – mirror neurons – are activated when the animal executes a grasping action, as well as when it observes the same action performed by another individual. According to Rizzolatti and colleagues (Gallese et al. 1996), the mirror system could be implicated in understanding the meaning of an action by matching observation with execution. It might provide the link between actor and observer that also exists between message sender and receiver.

This system is able to receive, understand, and send messages with the arm. Rizzolatti and Arbib (1998) have suggested that Broca’s area may be the homologue of the monkey premotor area F5, where a mirror system has also been found; and that the system could have served as an initial communication system in language evolution.

They propose that humans – covertly pantomiming actions they observe – come to understand their meaning. When the observed action is of particular interest, the premotor system allows a brief ‘prefix’ of the movement to be exhibited. This prefix is then recognized by the other individual, affecting both actor and observer. The actor recognizes an intention in the observer; while the observer notices that his involuntary response affects the behaviour of the actor. When this occurs, a primitive dialogue between observer and actor is established. This explains the intersubjective character of a communication system based on gesture. Initially, a repertoire based on pantomimes and iconic gesture might be used to exchange information on food or plan predatory activities. By means of the hand/mouth double-command system, this hand/arm gestural repertoire could be shared with – and successfully transferred to – the mouth (Gentilucci & Corballis 2006).

Such a sharing mechanism has been observed in humans. If a subject is asked to pronounce a syllable at the same time he observes another subject grasping either a large or a small object, the manner in which he pronounces the syllable is influenced by the size of the object he sees the other subject grasping (Gentilucci 2003, Gentilucci et al. 2004b). Both lip opening and formant 1 of the voice spectra are greater when the subject observes the grasping of large versus small objects. The actor’s finger shape during the grasping movements correlates with the manner of the observers’ lip opening as he pronounces the syllable /ba/ (Gentilucci et al. 2004b). Visual analysis of the stimulus size and observation of a spot of light (i.e., a non-biological stimulus) moving towards either small or large objects with the same velocity as the actor’s grasping hand do not show comparable effects (Gentilucci 2003). The same variation in formant 1 is found when observers pronounce the syllable /da/ and observe the grasping of fruits of different size. In contrast, no effect is found when observing foot interactions with small versus large objects or objects alone (Gentilucci et al. 2009). When single-pulse Transcranial Magnetic Stimulation (TMS) – which induces a facilitation of the stimulated area – is applied to the tongue motor cortex, it produces facilitation effects when observing graspings (and
pantomimed graspings) but not foot interactions or objects alone (Gentilucci et al. 2009). Once again, the specificity of the relationship between hand and mouth is proven.

Observing the action of bringing large versus small fruits to the mouth likewise affects the simultaneous pronunciation of syllables (Gentilucci et al. 2004a). Observing the action of guiding a large object induces an increase in the formant 2 of the voice spectra compared to observing the action of guiding a small object. The same effects are found when the actor pantomimes the action of bringing the small and large fruits to the mouth without opening it. This shows that neither the presence of the fruit nor the opening of the mouth is responsible for the effect. Observing the action affects the voice in the same way as execution of the action does. We speculate that, by use of the mirror system, an individual observing an arm action automatically and covertly executes (i.e., simulates) the action. This may be the basis on which the meaning of another’s actions are understood.

For manual actions functionally related to orofacial actions, the motor command is also sent to the mouth, reaching the threshold for execution when the mouth is already activated to pronounce the syllable – as our studies show (Gentilucci 2003; Gentilucci et al. 2001, 2004a,b; Gentilucci et al. 2009).

Both the execution and observation of grasping with the hand activates a command to grasp with the mouth. This modifies the posture of the anterior mouth articulation, according to the hand shape used to grasp the object. This, in turn, affects formant 1 of the voice spectra (Gentilucci 2003; Gentilucci et al. 2001; Gentilucci et al. 2004b; Gentilucci et al. 2009), which is related to mouth aperture (Leoni & Maturi 2002). Conversely, both the execution and observation of bringing-to-the-mouth probably induces an internal mouth movement (e.g., chewing or swallowing), which affects tongue displacement according to the size of the object brought to the mouth. This, in turn, modifies formant 2 (Gentilucci et al. 2004a; b), which is related to tongue position (Leoni & Maturi 2002). The same effects are found when observing action pantomimes. On the basis of these results, we propose that pantomimes are used as signs communicating the meaning of actions or property of an object: e.g., taking possession of an object by grasping or showing that the object is edible because it is brought to the mouth. Through the double motor command to hand and mouth, these pantomimes might be associated with activity of particular articulatory mouth organs (Gentilucci & Corballis 2006). Since pantomime affects the configuration of these organs in a specific manner, we suggest that the meaning of the pantomimed action could be carried on by the resultant mouth configuration. The mouth might have been co-opted for speaking due to this transfer (Gentilucci & Corballis 2006; Gentilucci et al. 2006).

From transitive actions to symbolic gestures

Observation of a pantomimed action influences speech in a way that is very similar to observation of the real action. Pantomimes of actions may incorporate iconic gestures that are analogue
representations of objects or actions (Donald 1991). These gestures may lose their analogue features and become arbitrary. The shift from iconic gestures to arbitrary symbols is termed *conventionalization*: conventionalized gestures (symbolic gestures or emblems) do not resemble the physical form of their referents but relate to them in a more abstract way: e.g., connecting the thumb and forefinger in a circle and holding the other fingers straight has acquired the meaning of ‘okay’ even if it seems unrelated to any action or object. The process of *conventionalization* – by which a meaning is attributed to an apparently arbitrary gesture – appears to be common to human and animal communication systems, and is probably driven by increased economy of reference. (For further explanation, see Burling 1999.) In this way, symbolic gestures might have evolved from transitive actions. Concurrently in humans, the same system that relates transitive actions to *phones* (i.e., mouth-articulation postures) might be involved in relating symbolic gestures to words.

According to this point of view, gesture and speech may share the same control system. If true, it leads to the following testable hypothesis: meaningful gestures – but not meaningless ones – should affect words when the two communication signs are produced simultaneously. Conversely, words – but not pseudo-words – should affect gestures.

**Symbolic gesture and communicative words share the same control system**

Bernardis and Gentilucci (2006) presented participants with printed words expressing communicative signs (*CIAO, NO, STOP*). The participants responded to the stimuli using different modalities: repeating the word aloud, executing the corresponding gesture, or repeating the word aloud and simultaneously executing the gesture. In two control conditions, participants pronounced the word and executed a meaningless gesture involving the arm and hand, or pronounced a pseudo-word formed by the same vowels as the three words (*lao*) while randomly executing one of the three symbolic gestures. When pronouncing a word, the parameters of the voice spectra – in particular, formant 2 – increased with simultaneous execution of the corresponding gesture compared to only pronouncing the word. The effect was not observed when the gesture was meaningless. The parameters of the arm kinematics during gesture execution decreased when the gesture was executed simultaneously with word pronunciation compared to when only the gesture was performed. Comparable effects were not observed when participants pronounced a pseudo-word; or, in a follow-up study, when gesture and word were semantically incongruent (Barbieri et al. 2009). The gesture controller seems to send reinforcement commands to the mouth controller involved in pronouncing the words. Once the mouth controller receives the signal, it sends back a block command to the arm controller, because the gesture becomes redundant. Previously, Gentilucci and colleagues (Bernardis & Gentilucci 2006, Gentilucci et al. 2006) have proposed that voice reinforcement may be a consequence of the translation – of the social intention to interact directly with the interlocutor – from arm to mouth (see also below). This produces an increase in the word’s intersubjective nature.
In a second series of experiments, Bernardis and Gentilucci (2006) studied verbal responses to communication gestures presented using different modalities. The aim was to determine whether the observing of/listening to (i.e., interpreting of) communication signs follows the same rules as does action execution. Participants were presented with a word (CIAO, NO, STOP) on a PC display and three different video clips. One showed an actress pronouncing the word; another showed her carrying out a gesture that corresponded in meaning to the word; the third showed her pronouncing the word while performing the corresponding gesture at the same time. The subject responded to each presentation by pronouncing the corresponding word. When the actor was observed pronouncing the word alone or executing the gesture alone, formant 2 of the word’s vowel increased compared to the situation in which the subject merely read aloud the word presented on the screen. A further increase in formant 2 was observed for responses to the simultaneous presentation of word and gesture. Once again, observation of communicative gestures induced the same reinforcement of formant 2 as did execution.

The authors suggest that the observer, responding verbally to the presented gestures, automatically and covertly imitates them through activation of the mirror system. The imitation command is also sent to the mouth controller, inducing modification in the articulation posture that, in turn, affects formant 2. To explain the functional meaning of the formant 2 increase, Bernardis and Gentilucci (2006) argue that placing the tongue further forward in the mouth induces an increase in formant 2 (see also Leoni & Maturi 2002). In non-humans, gestures of both mouth aperture and tongue protrusion accompany gestures typical of approaching relationships: e.g., among monkeys, the tongue protrudes during lip smacking, and face protruding precedes grooming actions (Van Hooff 1962, 1967). When a visible interlocutor is observed producing communicative signs, a responder is induced to a more direct relationship with her: in particular, the responder may configure mouth articulation in such a way as to interact directly. In other words, tongue-brought-forward-in-the-mouth – i.e., an increase in formant 2 – relates to the social intention coded by the communicative sign.

**Broca’s area and the translation of gesture to word**

Previous neuroimaging studies have found activation of Broca’s area when participants observe meaningful arm gestures (Gallagher & Frith 2004, Buccino et al. 2001, Grèzes et al. 1998, Decety et al. 1997, Grafton et al. 1996), with activation of the left ventral premotor area when participants observe and imitate meaningless gestures (Grèzes et al. 1999). Many results show that Broca’s area is involved in representing phonological units, in terms of mouth-articulation gestures (Paulesu et al. 1993, Demonet et al. 1992, Zatorre et al. 1992). On the basis of all this data, we hypothesize that Broca’s area is involved in translating aspects of activated representations of arm gestures into mouth articulation gestures. These aspects may concern the goal of the gesture (Buccino et al. 2001, Buccino et al. 2004) or the intention, or both (Iacoboni et al. 2005). Translation of these aspects should be
blocked by inactivation of the area. To test this hypothesis, Gentilucci and colleagues (2006) induced a brief inactivation of Broca’s area by use of repetitive Transcranial Magnetic Stimulation (rTMS) at low frequency (1 Hz) while subjects performed a task in which they were asked to give verbal responses to presentations of word alone, gesture alone, and word and gesture together – as previously described (Bernardis & Gentilucci 2006). For controls, they stimulated the symmetrical site in the right cortex and performed a sham stimulation of the same site on the left cortex.

As before (Bernardis & Gentilucci 2006), after both sham and right-hemisphere stimulations, formant 2 increased when responding verbally to the actress pronouncing the word or executing the gesture, compared to the word reading. There was a further increase when responding to the actress pronouncing the word and executing the gesture simultaneously. However, with left-hemisphere stimulation, no effect of gesture observation on formant 2 was observed.

In a control experiment, the authors demonstrated that stimulation of Broca’s area produces different effects when responding to communicative versus non-communicative signals (Gentilucci et al. 2006). They presented both the communicative signs used in the previous experiment (CIAO, NO, and STOP) and colour signals: yellow, red, and rose. The colour presentations consisted of printed words of the colours (i.e., GIALLO [yellow], ROSA [rose], and ROSSO [red]), video clips of an actress pronouncing the names of the colours, and video clips of the actress executing the communicative gestures CIAO, NO, and STOP, with coloured spots on the palm of her hand. Subjects responded to the communicative signs by pronouncing the corresponding words and to the colour signals by pronouncing the names of the colours. A different pattern was observed in the colour-presentation responses compared to the communicative-sign responses. No difference in formant 2 was observed when participants read the colour words or responded to the actress pronouncing the colour names after the sham, right-hemisphere, or left-hemisphere stimulations. Formant 2 decreased when participants pronounced the name of the colour presented on the gesturing hand after the sham and right-hemisphere stimulations. The left-hemisphere stimulation removed this effect. We explain this result as follows: observation of the gesture automatically activates pronunciation of the corresponding word, which interferes with verbal response to the colour. Again, Broca’s area seems to be involved in transforming gestures into words.

CONCLUSIONS

Neurophysiological and behavioural data favour the hypothesis that speech and gesture share the same control system. We propose that this system was used to translate an initial form of communication, based on arm gestures, into mouth-articulation gestures related to speech (Gentilucci & Corballis 2006). We believe that the reciprocal influence between speech and gesture we observed in our studies are not just remnants of an ancient process. On the contrary: we propose that the same system of translation is capable of merging speech and symbolic gesture into a single sign. This sign acquires a
new meaning, compared to the meaning of the word and the gesture when produced separately. This connects directly with our hypothesis about the social relevance of a similar process: when the social intention to interact directly with a interlocutor is transferred from arm to mouth, this produces an increase in the intersubjective nature of a word. The data described in the present review suggest that the intersubjective character of spoken language derives primarily from the sharing of arm-action meaning between individuals, and then in the transfer of the social intention of symbolic gestures from hand to mouth.

Acknowledgements

We thank Patricia Gough for her comments on the manuscript. The work was supported by grants from MIUR (Ministero dell’Istruzione, dell’Università e della Ricerca) to M.G.

REFERENCES


