The common language of speech perception and action: a neurocognitive perspective

Jean-Luc Schwartz, Marc Sato, Luciano Fadiga

The common language of speech perception and action: a neurocognitive perspective

Jean-Luc Schwartz[(1)], Marc Sato[(1)], Luciano Fadiga[(2)]
(1) GIPSA-Lab, UMR 5216 CNRS – Univ. de Grenoble, France – Jean-Luc.Schwartz, Marc.Sato@gipsa-lab.inpg.fr
(2) Section of Human Physiology, University of Ferrara and The Italian Institute of Technology, Genova, Italy – fdl@unife.it

Abstract

How do listeners process the speech signal to extract acoustic cues and recover phonetic information? More than 50 years after the appearance of the motor theory of speech perception, recent neurophysiological discoveries challenge the view that speech perception only relies on perceptual auditory mechanisms and suggest that the motor system is also crucial for speech recognition. The aim of the present chapter is to review and discuss these findings in an attempt to define what could be a "common language of perception and action".

Introduction

The question of the auditory vs. motor nature of cognitive representations in speech perception is at the centre of an already old and now quite classical debate. There are actually a number of precursors of the role of action in perception, such as the philosopher Berkeley with his works on vision, or the physicist Helmholtz with the outflow theory pioneering the efferent copy concept. At the beginning of the 20th century, the psychologist and phonetician Raymond Herbert Stetson introduced in his "motor phonetics" the famous claim that "speech is rather a set of movements made audible than a set of sounds produced by movements" (Stetson, 1988), which paves the way towards a view in which gestures are primary and sounds secondary in the linguistic exchange. But the central character in this story is obviously the psychologist Alvin Liberman. His works at the Haskins Labs after the World War 2 lead him to a view in which the link between sounds and phonemes was both complex and unsatisfactory, which drove him progressively at the beginning of the 50s towards what became in the 60s "The Motor Theory of Speech Perception". This theory proposes kind of a dramatic switch in the conception of speech perception, which stimulated both the emergence of new experimental paradigms, the acquisition of many data and … the content of many oral or written debates in the speech communication community. Interestingly, at a time where the debate seemed decreasing – together with the motor theory appeal – the beginning of the 90s lead to a complete reconsideration of the elements of the debate, mainly – though not exclusively, as we shall see – thanks to new techniques and new discoveries about the primate and human brain. The objective of the present chapter is to make the point on the debate and attempt to define some great trends of what could be a "common language of perception and action". The chapter is organised in four parts. Firstly, some basic elements of the 40-years debate will be recalled. Then, the turning point of the "discovery" of mirror neurons and of the introduction of new neurocognitive techniques will be described in the speech perception context. A third section will be devoted to a reanalysis of the perceptuo-motor links in light of new discoveries. The fourth section will be the occasion to present some perspectives. Interestingly in this context, it is noteworthy to mention that the authors of the present paper

participated to this story from different perspectives – though with a common interest and a large number of shared views on the role of perceptuo-motor interactions, and hence, do not necessarily completely converge on a coherent final story.

I. An old debate about auditory vs. motor theories of speech perception

The starting point in Liberman and colleagues' reasoning is the difficulty to describe in a straightforward way the link between phonemes and sounds. This includes two intricate components. Firstly, the complexity of the linguistic message requires to combine phonemes at a high rate, which is made possible by doing several things at the same time, that is combining articulatory gestures in a clever, largely parallel way: this is coarticulation in a broad sense. Secondly, this possibly linear combination of articulatory components is non-linearly transformed into a sequence of sounds in which the link with the sequence of phonemes is not at all transparent. The basic intuition in the motor theory is that the link between phonemes and gestures should be more direct than the link with sounds (Liberman et al., 1962; Liberman and Mattingly, 1985; Liberman and Whalen, 2000).

This switch from an "information processing" acoustic/auditory approach to a motor approach was related to a double functionalist framework. Firstly, in a *linguistic (phonological) framework*, would phonological units be better described in motor rather than in auditory terms (see e.g. the proposals, in the same Haskins Laboratories, of an "articulatory phonology", Browman and Goldstein 1986, 2000)? Secondly, in a *technological (computational) framework*, should informational elements be conceived as articulatory rather than acoustical? The literature about both phonetic/phonological descriptions and speech technologies (recognition, synthesis, coding) was constantly stimulated by this dichotomy during the second part of the 20th.

The debate was then nourished by a large series of experimental works, which stimulated the emergence or promotion of a number of new paradigms, such as categorical perception (Repp, 1984), trading relationships (Repp, 1983), perceptuo-motor adaptation effects (Cooper, 1979), close shadowing and perceptuo-motor interactions (Porter and Lubker, 1980), audiovisual and multisensory integration (Dodd and Campbell, 1987; Campbell et al., 1998), the duplex effect (Liberman et al., 1981), etc. In parallel, the motor theories (with a plural accounting for the emergence of a "second motor theory", the Direct Realist Theory by Carol Fowler also at the Haskins Labs) searched for a large-scale cognitive background: modularism in the case of Liberman and Mattingly's Motor Theory (1985), Gibsonian realism in the case of Fowler (1986).

However, it is fair to say that the debate has progressively somehow decreased in vigour and lost in acuity, while in the same movement the interest and credibility of motor theories declined in the speech community for lack of decisive arguments. In this context, a number of new discoveries and the emergence of new techniques in the field of neurocognition of perception and action produced a spectacular movement, shifting the equilibrium point back towards motor views, introducing new paradigms and, by the way, renewing the interest for old paradigms that had been more or less abandoned.

Before describing this shift in the next section, and analysing its consequences for the perceptuo-motor debate in the following one, let us mention two points that seem to us quite important in the analysis of this debate. Firstly, whatever the position that should be adopted

finally in this debate, the neurocognitive perceptuo-motor shift that we are experiencing has definitely a major interest: it forces to establish or re-establish a solid link between knowledge in speech perception and production, a link which was quasi vanishing in the last twenty years. Perception and production specialists did not go in the workshops or sessions of international conferences and seldom worked together in the same labs. The situation is now quickly changing, and this is in our view a major positive achievement.

A second interesting point to mention is that another functionalist framework has appeared in the reasoning and takes an increasing importance: the search for *language origins*. This theme has been long considered as fragile or even unsound from various perspectives, including the Chomskyan one. The last years have seen on the contrary an increasing interest for the question of language phylogeny (see e.g. the OMLL program launched by the European Science Foundation, http://www.esf.org/activities/eurocores/programmes/omll.html), and the nature of the perceptuo-motor link is at the centre of this question through the concept of *parity*. Therefore, phylogeny now participates to the debate itself, and the authors of the present chapter are all convinced that it is indeed an adequate framework for considering perceptuo-motor interactions in speech communication.

II. Perception as a mirror of action: the neurocognitive shift

A strong empirical support to Liberman's motor theory of speech perception comes from the discovery of mirror neurons in monkey premotor area F5. Area F5 belongs to the ventral premotor cortex and stores a representation of hand and mouth actions, as shown by single neurons and intracortical microstimulation studies (see Rizzolatti et al., 1988).

The specificity of the goal seems to be an essential prerequisite in activating F5 neurons. The same neurons that discharge during grasping, holding, tearing, manipulating, are silent when the monkey performs actions that, although involving a similar muscular pattern, are indeed characterised by a different goal (i.e. grasping to put away, scratching, grooming, etc.).

In addition to the motor properties shared by all F5 neurons, a particular class of F5 neurons discharge also when the monkey *observes* another individual making an action in front of it. These neurons are the "mirror neurons" (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996a), a special class of visuomotor neurons matching others' actions on the observer's motor repertoire. There is a strict congruence between visual and motor properties of F5 mirror neurons: e.g., mirror neurons motorically coding whole hand prehension discharge during observation of whole hand prehension performed by the experimenter but not during observation of precision grasp.

The most likely interpretation for the visual response of visuomotor neurons is that, at least in adult individuals, there is a close link between action-related visual stimuli and the corresponding actions that pertain to monkey's motor repertoire. Thus, every time the monkey observes the execution of an action, the related F5 neurons are addressed and the specific action representation is "automatically" evoked. Under certain circumstances it guides the execution of the movement, under others, it remains an unexecuted representation of it that might be used to understand what others are doing.

Transcranial magnetic stimulation (TMS) (Fadiga et al., 1995; Strafella and Paus, 2000) and brain imaging experiments have demonstrated that a mirror-neuron system is present also in humans: when the participants observe actions made by human arms or hands, motor cortex becomes facilitated (this is shown by TMS studies) and cortical activations are evoked in the ventral premotor/inferior frontal cortex (Rizzolatti et al., 1996b; Grafton et al., 1996; Decety et al., 1997; Grèzes et al., 1998; Iacoboni et al., 1999; Decety and Chaminade, 2003; Grèzes et al., 2003). Grèzes et al. (1998) showed that the observation of meaningful but not that of meaningless hand actions activates the left inferior frontal gyrus (Broca's region). Moreover, two further studies have shown that observation of meaningful hand-object interaction is more effective in activating Broca's area than observation of non goal-directed movements (Hamzei et al, 2003; Johnson-Frey et al, 2003). In addition, direct evidence for an observation/execution matching system has been provided by two experiments, one employing fMRI technique (Iacoboni et al, 1999), the other using event-related MEG (Nishitani and Hari, 2000) that directly compared in the same subjects action observation and action execution.

Taken together, all the fMRI studies on action observation, constantly show that Broca's area (or its right homologue) become active when we observe the actions of another individual. The evidence that Broca's area is activated during time perception and calculation tasks (Gruber et al. 2001), harmonic incongruity perception (Maess et al., 2001), tonal frequency discrimination (Muller et al., 2001), prediction of sequential patterns (Schubotz and von Cramon, 2002a) as well as during prediction of increasingly complex target motion (Schubotz and von Cramon, 2002b), suggests that this area could have a central role in representing syntactically ordered sequential information in several different domains (Lieberman, 1991). This could be crucial for action understanding, allowing the parsing of observed actions on the basis of the predictions of their outcomes.

Others' actions do not generate only visually perceivable signals. Action-generated sounds and noises are also very common in nature. In a recent experiment Kohler and colleagues (2002) have found that 13% of the investigated F5 neurons discharge both when the monkey performed a hand action and when it heard the action-related sound. Moreover, most of these neurons discharge also when the monkey observed the same action, demonstrating that these 'audio-visual mirror neurons' represent actions independently of whether they are *performed*, *heard* or *seen.*

The presence of an audio-motor resonance in a monkey brain region considered to be the cytoarchitectonical homologue of human Broca's area (classically considered as the motor centre for speech) prompts the Liberman's hypothesis on the mechanism at the basis of speech perception. The motor theory maintains that the ultimate constituents of speech are not sounds but articulatory gestures that have evolved exclusively at the service of language. Speech perception and speech production processes could thus use a common repertoire of motor primitives that, during speech production, are at the basis of articulatory gesture generation, 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 her articulatory gestural representations are activated by the listening to verbal sounds. Although this theory is not unanimously accepted, it proposes a plausible model of an action/perception cycle in the frame of speech processing. To investigate if speech listening activates listener's motor representations, Fadiga et al. (2002) administered TMS on cortical tongue motor

representation, while subjects were listening to various verbal and non-verbal stimuli. Motor evoked potentials (MEPs) were recorded from subjects' tongue muscles. Results showed that during listening of words formed by consonants implying tongue mobilisation (i.e. Italian 'R' vs. 'F') MEPs significantly increased. This indicates that when an individual listens to verbal stimuli, his/her speech related motor centres are specifically activated. Moreover, words-related facilitation was significantly larger than pseudo-words related one.

The presence of "audio-visual" mirror neurons in monkeys (Kohler et al., 2002) and the presence of this "speech-related acoustic motor resonance" in humans (Fadiga et al. 2002), suggests that, independently from the sensory nature of the perceived stimulus, the mirror-neuron resonant system retrieves from the action vocabulary (stored in the frontal cortex) the stimulus-related motor representations. It is however unclear if the activation of the motor system during speech listening is causally related to speech perception, or if it is a mere epiphenomenon due, for example, to an automatic compulsion to repeat without any role in speech processing. Empirical evidence suggests that the first hypothesis might be correct (Wilson et al. 2004, Pulvermuller et al., 2006, Meister et al., 2007). A recent experimental work done in our laboratory (Kotz et al., submitted, D'Ausilio et al., in preparation) shows that the application of TMS on speech-related areas specifically interferes with different speech components. Indeed, whereas the application of TMS on motor centres induces clear-cut phonological interference effects, the TMS-induced virtual lesion of Broca's area seems to have effects only (if any) on the discrimination of the lexical properties of auditorily presented verbal stimuli. The recent finding that Broca's aphasics show a specific deficit in pragmatically representing the actions performed by others (Fazio et al., submitted) further strengthens the possibility that speech evolved on a premotor substrate originally devoted to motor understanding. This provides further support to the idea that Broca's area became a speech centre because of its premotor origins (Fadiga et al., 2006).

III. Reanalysis of the perceptuo-motor link in speech perception

As previously described, the properties of mirror neurons in the monkey brain and of a putative mirror neuron system in humans have provided evidence pointing to a close connection between perception and action systems during action observation. By indicating a neurophysiological mechanism that might create 'motor parity' between communicating individuals, the discovery of the human mirror neuron system has been interpreted as a strong empirical support to one of the main claims of Liberman's motor theory of speech perception, that is, perceiving speech is perceiving gestures.

Since then, besides the involvement of temporal auditory regions, brain areas involved in the planning and execution of speech gestures (i.e., the left inferior frontal gyrus, the ventral premotor and primary motor cortices) and areas subserving proprioception related to mouth movements (i.e., the somatosensory cortex), have been repeatedly found to be activated during 'passive' auditory, visual and/or auditory-visual speech perception (e.g., Möttonen et al., 2004; Wilson et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Skipper et al. , 2005; Pulvermuller et al., 2006; Wilson and Iacoboni, 2006; Skipper et al., 2007). As previously mentioned, recent TMS studies also demonstrated that motor-evoked potentials recorded from the lips or tongue muscles are enhanced during both passive speech listening and viewing, when stimulating the corresponding area of the left primary motor cortex (Sundara et al., 2001; Fadiga et al., 2002; Watkins et al., 2003; Watkins and Paus, 2004; Roy et al., 2008). Importantly, this speech motor 'resonance' mechanism appears to be articulatory specific,

motor facilitation being stronger when the recorded muscle and the presented speech stimulus imply the same articulator (Fadiga et al., 2002; Roy et al., 2008). The specificity of this speech motor resonance mechanism is also suggested by two recent fMRI studies showing similar somatotopic patterns of motor activity in the superior portion of the ventral premotor cortex during both producing and listening to or viewing lips- and tongue-related phonemes (Pulvermuller et al., 2006; Skipper et al., 2007). Altogether, these studies thus suggest that speech perception involves an automatic and specific mapping from the speaker's articulatory gestures into the listener's motor plans.

Most recent neurobiological models of speech and language understanding also claim for a tight connection between perception and production systems. These models have in common to postulate that the links between articulatory and perceptual mechanisms look like or derive from action–perception links that are observed for a range of non-linguistic actions (Aboitiz and Garcia, 1997; Aboitiz et al., 2006; Hickok and Poeppel, 2000, 2004, 2007; Scott and Johnsrude, 2003; Rizzolatti and Arbib, 1998; Arbib, 2005; Wilson and Iacoboni, 2006; Skipper et al., 2007).

One influential model is the dual-stream model of Hickok and Poeppel (2000, 2004, 2007). It is proposed that early cortical stages of speech processing involve auditory fields in the superior temporal gyrus. This cortical processing system then diverges into a ventral stream, which is involved in mapping sound onto meaning, and a dorsal stream, which is involved in mapping sound onto articulatory-based representations. The ventral stream projects ventro-laterally toward the inferior temporal cortex, which is assumed to contain widely distributed conceptual representations. The dorsal stream projects first towards a region at the parietal–temporal boundary, which serves as a sensorimotor interface, and then to frontal motor regions. Bi-directionality in the dorsal pathway between auditory temporal and motor frontal regions is assumed to provide a mechanism for the development and maintenance of parity between auditory and motor representations of speech, especially in infancy during speech acquisition. In adults however, the dorsal stream is not considered to be a critical component of speech perception under normal listening conditions. Rather, this dorsal circuit would play a functional role only when the listener has to explicitly use articulatory-based processes to keep auditory-based representations active, as in phonological tasks and verbal working memory tasks. Finally, another role of this sensorimotor loop is to allow rapid articulatory adjustments in speech production, by helping to distinguish the sensory consequences of our own actions from sensory signals due to changes in the outside world (see Guenther, 2006 for a review).

Another model proposed by Skipper and colleagues (2007) also explains speech perception by means of feedforward and feedback projections. In this model, early multisensory speech representations in the superior temporal gyrus and derived from acoustical and/or visual signal, can be thought of as multisensory hypotheses about the phonemes produced by the speaker. These hypotheses are then translated onto motor control commands localized in the inferior frontal gyrus and which, based on past articulatory experience, could generate corresponding motor actions in the ventral premotor and motor cortex. Activated motor commands would predict the acoustic and somatosensory consequences of executing a speech movement through efference copy, or feedback control commands, to both the left superior temporal sulcus/gyrus and somatosensory cortices, respectively. Finally, these internally generated sensory consequences are thought to constrain the ultimate phonetic interpretation of the incoming sensory information.

Critically, these two models not only argue against the view that speech perception relies exclusively on the auditory system and the acoustic properties of speech, but also that speech perception is determined only through feedforward, direct mapping, mechanisms from auditory to motor regions: they discard, in a symmetric way, both "pure" auditory theories and "pure" motor theories (see Schwartz et al., 2002, 2007, for a review). In addition, despite accumulating evidence that passive speech perception induces motor cortical activity, both models question a possible mediating role of the motor system under normal listening conditions. In the dual-stream model (Hickok and Poeppel, 2000, 2004, 2007), the primary function of the dorsal auditory-motor circuit is thought to serve speech development and the acquisition of a new vocabulary. When the child learns to articulate speech sounds, it may provide a mechanism by which sensory representations of speech can be stored and compared against its articulatory production, with this comparison being used to shape future productions. Although in adults motor representations of speech can still be activated, they are thought to be used strategically to assist in working memory and sub-lexical task performance, that is whenever translation of phonological information to an articulatory code is required to support maintenance and comparison of speech segments (e.g., Démonet et al., 1992, 1994; Zatorre et al., 1992; Paulesu et al., 1993, 1996; Burton et al, 2000 - for a review, see Poldrack et al., 1999; Démonet et al., 2005; Vigneau et al., 2006). In Skipper et al.'s model, the speech motor centres are thought to be strongly recruited depending on the modality of the presentation and on the ambiguity of the sensory inputs that is when the mapping between sensory information and phonetic categories is not sufficiently clear. This proposal is indirectly supported by some fMRI studies showing an increased activation of the speech motor centres during auditory-visual and visual speech perception compared with auditory presentation alone (Skipper et al., 2005, 2007), during the audiovisual observation of phonetically conflicting compared to matching vowels/syllables (e.g., Jones and Callan, 2003; Pekkola et al., 2005; Ojanen et al., 2005; Skipper et al., 2007), during the auditory identification of non-native versus native phonemes (e.g., Callan et al., 2004; Wilson and Iacoboni, 2006), and of intelligible versus masked or distorted speech (e.g., Binder et al., 2004; Zekveld et al., 2006).

In sum, whether the motor system might mediate speech perception through the internal generation of candidate articulatory categorizations under normal listening conditions is largely debated. Actually, it is important to note that while previous brain imaging and single-pulse TMS studies clearly demonstrate the recruitment of the motor system during passive speech perception, the results are intrinsically correlational and cannot be used to address causality. From this view, both electrocortical stimulation studies during neurosurgical operations, repetitive transcranial magnetic stimulation studies (rTMS) and clinical data from frontal aphasic patients are inconclusive regarding a possible functional role of the motor system in speech processing under normal listening conditions. Temporarily disrupting the activity of the opercular part of the inferior frontal gyrus (ie., Broca's area) or the superior portion of the ventral premotor cortex, by means of either repetitive TMS or electrocortical stimulation during neurosurgery, has been shown to disrupt subjects' ability to perform 'complex' phonological tasks that require segmentation processes and working memory demands (Boatmann, 2004; Nixon et al., 2004; Romero et al.,, 2006; Sato et al., in preparation). However, no interference effects were observed in syllable identification/discrimination tasks that could be performed without need for phonemic segmentation (Boatmann, 2004; Boatman and Miglioretti, 2005; Sato et al., in preparation), except in the case where syllables were embedded in white noise (Meister et al., 2007). Despite inherent limitations of both rTMS and electrocortical stimulation techniques, these results nevertheless appear in line with the above-mentioned models of speech perception,

indicating that the speech motor centers are actively recruited depending on the ambiguity/complexity of the speech stimuli and on the use of segmentation and working memory processes (Hickok and Poeppel, 2004, 2007; Skipper et al., 2007).

Alternatively, if the motor system does not play a critical role in speech processing under normal listening conditions in adults, then what could be the function of the motor activity observed during passive speech perception? One possibility is that this activity is not strictly intrinsic to speech comprehension but may rather facilitate conversational exchange by contributing to setting a common perceptuo-motor ground between speakers. In that case, speech motor resonance may represent dynamic sensorimotor adaptation under the influence of the other talker's speech patterns, and in return may facilitate conversational interactions by helping adaptive convergent behaviours.

As a matter of fact, previous studies have highlighted a strong tendency by a speaker to imitate a number of phonetic characteristics in another speaker's speech in the course of a conversational interaction. Such a behavioral tendency necessarily involves complex sensorimotor interactions that allow speakers to compare the phonetic characteristics of the utterances they hear with their own speech auditory and motor repertoire. Previous studies have shown that this interfacing process is displayed in a variety of ways. Some of them involve natural settings, as during conversational exchange when exposure to the speech of other talkers leads to phonetic convergence with that speech (e.g., Sancier and Fowler, 1997; Pardo, 2006). Some are special to experimental manipulations, as when seeing a video of an articulating mouth influences the production of similar or dissimilar articulations (Kerzel and Bekkering, 2000; Gentilucci and Cattaneo, 2005; Gentilucci and Bernardis, 2007).

Evidence for strong sensorimotor interactions in speech also comes from studies showing the existence of perceptuo-motor adaptation mechanisms. For instance, while manipulation of the auditory feedback during speech production leads to rapid motor corrections to counteract the effect of perturbation, an after-effect or adaptation is observed when the perceptual manipulation is removed (e.g., Houde & Jordan, 1998, 2002; Jones & Munhall, 2005; Purcell & Munhall, 2006; Villacorta et al., 2007). The fact that this persistence or learning in the motor system does not disappear immediately, likely reflects a change in motor representations due to a global remapping of the auditory-motor relationship. Crucially, it has been shown that real-time alteration of the auditory feedback related to the speaker's own voice, causes not only compensatory changes in the *production* but also in the *perception* of speech that persist once the feedback alteration has ended (Shiller et al., submitted). In addition, Cooper and Lauritsen (1974) demonstrated that repetitive listening to a CV syllable with an initial voiceless stop consonant caused subjects to produce a shorter voice onset time for voiceless stop consonants in CV syllables. Because adaptive changes in perceptual speech sound representations occur during repetitive listening of a speech sound (see Eimas & Corbit, 1973; Samuel, 1986; for discussion on this selective adaptation phenomenon), this perceptuo-motor adaptation might represent the fatiguing of specialised phonetic feature detectors that mediate both speech perception and production. These latter studies thus provide behavioural evidence for a functional and plastic change involving both input and output processes simultaneously, during both speech production and perception.

IV. Conclusive remarks and proposals

So, where are we now, more than 50 years after the first sketches of the Motor Theory as soon

as e.g.. Liberman et al. (1952) ("*we should expect that the relation between perception and articulation will be considerably simpler than the relation between perception and acoustic stimulus*", p. 513)? The "visionnaire" nature of his intuitions appears strikingly reinforced by the neurocognitive turn of the 90s, described in detail in the present chapter. There is no doubt that perceptual and motor representations are connected in the human brain, motor areas being active in speech perception while on the other way round auditory areas are active during speech motor control (see Guenther, 2006; Guenther et al., 2006). This connection should be crucial in the learning of perceptuo-motor representations in the course of speech development, during the first years of life. The dorsal route also seems actively involved in all explicit phonological tasks, which strongly suggests that the phoneme is really a *perceptuo-motor unit* built in through language development and stored somewhere inside the temporo-parieto-frontal circuit as an *amodal* network of connections between multisensory and motor representations.

The functional role of the perceptuo-motor connection in online speech perception is less clear. Between the "high" hypothesis of an automatic "translation" of speech sensory inputs into articulatory gestures, emerging from a number of papers from the Haskins Labs, and the "low" hypothesis of a more or less complete separation between the dorsal and the ventral streams, largely removing perceptuo-motor links from comprehension in the model by Hickok & Poeppel, there is a large space for theoretical elaboration, computational models, and, above all, new experiments. The authors of the present chapter would probably converge on setting the needle somewhere between these two extreme positions, around two basic ideas. Firstly, motor representations should play a crucial role in *shaping* perceptual units, that is *extracting* the adequate components, *predicting* future sensory events or *integrating* events in a hopefully smart way, *complementing* them with adequate articulatory information (possibly through "procedural knowledge", see Viviani and Stucchi, 1992). Secondly, there seems to exist an implicit, probably unconscious but largely autonomous *tuning* of the speaking/listening partners in a dialog, and this tuning, quite likely emerging from something like mirror neurons, should play an important role in social interaction, if not in active understanding.

The speech representations in this framework could be *perceptuo-motor* rather than merely auditory or merely motor. This is the view defended by Schwartz et al. (2002, 2007) in the "Perception-for-Action-Control Theory" (PACT) in which a speech gesture is not considered as a pure articulatory unit, but rather as a motor coordination shaped by motor-to-sensory nonlinearities (as in Stevens's Quantal Theory, 1972, 1989). This enables to take into account the "perceptual value" of an articulatory gesture, which leads to a number of efficient predictions about the shape of sound inventories in human languages, as shown by Lindblom with the Dispersion Theory (Liljencrants and Lindblom, 1972; Lindblom, 1986). PACT, centered on the *co-structuring of the perception and action systems in relation with phonology*, is clearly different from both an auditory theory in which the sensory-interpretative chain is considered independently of the patterning of sounds by speech gestures, in the search of some *"direct link" between sounds and phonemes*; and from a motor theory in which perception is nothing but a mirror of action, in the claim of a *"direct link" between sounds and gestures*. It is rather focused on *multimodal percepts regularized by motor constraints*; or *speech gestures shaped by multimodal processing*.

Altogether, and whatever the needle position, there seems to be indeed a common language of perception and action, shaping speech communication and human language. The experimental and theoretical challenges for speech communication researchers are strongly renewed and

enhanced in this now widely accepted framework. Progress in most dimensions of speech research, including perception, production, development, phylogeny, and technology, should derive from this fascinating perspective.

ACKNOWLEDGEMENTS:
Luciano Fadiga was supported by CE grants Robotcub, Poeticon and Contact. Marc Sato and Jean-Luc Schwartz benefited from a support from the ISCC CNRS program (Cog-Speech project).

References

Aboitiz, F. & Garcia, V. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. Brain Research Reviews, 25: 381-396.

Aboitiz, F., Garcia, V., Bosman, C. & Brunetti, E. (2006). Cortical memory mechanisms and language origins. Brain and Language, 98(1): 40-56.

Arbib, M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behavioral and Brain Sciences, 28(2): 105-124.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A. & DouglasWard, B. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci., 7: 295-301.

Boatman, D.F. & Miglioretti D.L. (2005). Cortical sites critical for speech discrimination in normal and impaired listeners. J. Neuroscience, 25(23):5475–5480.

Boatmann, D.F. (2004). Cortical bases of speech perception: evidence from functional lesion studies. Cognition, 92: 47-65.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. Phonology Yearbook, 3, 219-252.

Browman, C.P. & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. Bulletin de la Communication Parlée, 5, 25-34.

Burton, M.W., Small, S.L. & Blumstein, S.E. (2000). The role of segmentation in phonological processing: an fMRI investigation. Journal of Cognitive Neuroscience, 12(4): 679-690.

Callan, D.E., Jones, J., Callan, A. & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. NeuroImage, 22: 1182-1194.

Campbell, R., Dodd, B., & Burnham, D. (eds.) *Hearing by eye, II. Perspectives and directions in research on audiovisual aspects of language processing* (pp. 85-108). Hove (UK): Psychology Press

Cooper, W. (1979). Speech perception and production: Studies in selective adaptation. Norwood, NJ: Ablex.

Cooper, W.E. & Lauritsen, M.S. (1974). Feature processing in the perception and production of speech. Nature, 252:121-123.

Decety, J., & Chaminade, T. (2003). When the self represents the other: a new cognitive neuroscience view on psychological identification. Conscious Cogn, 12(4), 577-596.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. Brain, 120 (Pt 10), 1763-1777.

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A. & Frackowiak, R.S.J. (1992). The anatomy of phonological and semantic processing in normal subjects. Brain, 115: 1753-1768.

Démonet, J.-F., Price, C., Wise, R. & Frackowiak, R.S.J. (1994). A pet study of cognitive strategies in normal subjects during language tasks: influence on phonetic ambiguity and sequence processing on phoneme monitoring. Brain, 117(4): 671-682.

Démonet, J.F., Thierry, G. & Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. Physiol. Rev., 85 (1): 49–95.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. Exp Brain Res, 91(1), 176-180.

Dodd, B., & Campbell, R. (Eds.), *Hearing by eye: the psychology of lipreading* (pp. 3-51). Lawrence Erlbaum Associates, London.

Eimas, P.D. & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4: 99–109.

Fadiga, L., Craighero, L., Buccino, G., & 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., & Roy A.C. (2006). Broca's area: a speech area? In Grodzinsky Y, Amunts K (editors) Broca's region (pp. 137-52). New York: Oxford University Press.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol, 73(6), 2608-2611.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A.C., Calzolari, A., Pozzo, P., Granieri, E., & Fadiga, L. Encoding of Human Action in Broca's area. Submitted.

Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *J. Phonetics, 14*, 3-28.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain, 119 (Pt 2), 593-609.

Gentilucci, M. & Bernardis, P. (2006). Automatic audiovisual integration in speech perception. Neuropsychologia.

Gentilucci, M. & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. Exp Brain Res., 167: 66-75.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. Exp Brain Res, 112(1), 103-111.

Grezes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. Neuroimage, 18(4), 928-937.

Grèzes, J., Costes, N., & Decety, J. (1998). Top-down effect of strategy on the perception of human biological motion: a PET investigation. Cog Neuropsych, 15, 553-582.

Gruber, O., Indefrey, P., Steinmetz, H., & Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. Cereb Cortex, 11(4):350-359.

Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, 39, pp. 350-365.

Guenther, F.H., Ghosh, S.S., and Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, pp. 280-301.

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C., & Buchel, C. (2003). The human action recognition system and its relationship to Broca's area: an fMRI study. Neuroimage, 19, 637-44.

Hickok, G. & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends in Cognitive Science, 4(4): 131-138.

Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition, 92: 67-99.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8: 393-402.

Houde, J.F. & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science, 279 (5354) : 1213–1216.

Houde, J.F. & Jordan, M.I. (2002). Sensorimotor adaptation of speech: I. Compensation and adaptation. J. Speech Lang. Hear. Res., 45: 295–310.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. Science, 286(5449), 2526-2528.

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

Jones, J. & Callan, D.E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. NeuroReport, 14(8): 1129-1133.

Jones, J.A. & Munhall, K.G. (2005). Remapping auditory–motor representations in voice production. Curr. Biol., 15 (19): 1768–1772.

Kerzel, D. & Bekkering, H. (2000). Motor activation from visible speech: evidence from stimulus response compatibility, J. Exp. Psychol. Hum. Percept. Perform., 26: 634-647.

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

Kotz, S.A., Begliomini, C., Craighero, L., D'Ausilio, A., Fabri-Destro, M., Raettig, T., Zingales, C., Haggard, P., & Fadiga, L. Is Broca's area involved in phonological perception? Submitted.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 65, 497-516.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1962). A motor theory of speech perception. Proceedings of the Speech Communication Seminar, Stockholm.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics, 30, 133-143.

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liberman, A.M. & Whalen, D.H. (2000). On the relation of speech to language. Trends in Cognitive Sciences, Vol 4, No. 5, 187-196.

Lieberman, P. (2001). Human language and our reptilian brain. Perspectives in Biology and Medicine, 44(1), 32–51

Maess, B., Koelsch, S., Gunter, T.C., & Friederici, A.D. (2001). Musical syntaxprocessed in Broca's area: an MEG study. Nat Neurosi, 4, 540-545.

Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., & Iacoboni, M. (2007) The essential role of premotor cortex in speech perception. Curr. Biol. 17:1692-1696.

Möttönen, R., Järveläinen, J., Sams, M. & Hari, R. (2004). Viewing speech modulates activity in the left SI mouth cortex. NeuroImage, 24: 731-737.

Muller, R.A., Kleinhans, N., & Courchesne, E. (2001). Broca's area and the discrimination of frequency transitions: a functional MRI studies. Brain Lang, 76, 70-76.

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. Proc Natl Acad Sci U S A, 97(2), 913-918.

Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P. & Passingham, R. (2004). The inferior frontal gyrus and phonological processing: an investigation using rTMS. Journal of Cognitive Neuroscience, 16(2): 289-300.

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I.P., Joensuu, R., Autti, T. & Sams, M. (2005). Processing of audiovisual speech in Broca's area. NeuroImage, 25: 333-338.

Pardo, J. (2006). On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America, 119 (4): 2382-2393.

Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal components of working memory. Nature, 362: 342-344.

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jaaskelainen, L.P., Kujala, T. & Sams, M. (2006). Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3T. NeuroImage, 29(3): 797-807.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H. & Gabrieli, J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage, 10: 15–35.

Porter, R. J.[, Jr.], & Lubker, J. F. (1980). Rapid reproduction of vowel–vowel sequences: evidence for a fast and direct acoustic–

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci U S A, 103, 7865-70.

Purcell, D.W. & Munhall, K.G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. Journal of the Acoustical Society of America, 119: 2288–2297.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-361.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 10). (pp. 243-335). New York: Academic Press.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & 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(3), 491-507.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996a). Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res, 3(2), 131-141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996b). Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp Brain Res, 111(2), 246-252.

Romero, L., Walsh, V. & Papagno, C. (2006). The neural correlates of phonological short-term memory: a repetitive transcranial magnetic stimulation study. Journal of Cognitive Neuroscience, 18(7): 1147-1155.

Roy, A.C., Craighero, L., Fabbri-Destro, M. & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. J. Physiol. Paris, 102(1-3): 101-105.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. Cognitive Psychology, 18, 452–499.

Sancier, M., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of Phonetics, 25: 421–436.Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. Journal of Communication Disorders, 39: 350–365.

Sato, M., Tremblay, P. & Gracco, V. (submitted). A mediating role of the premotor cortex in speech segmentation.

Schubotz, R.I., Friederici, A.D., & von Cramon, D.Y. (2000). Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. Neuroimage, 11(1), 1-12.

Schubotz, R.I., & von Cramon, D.Y. (2002a). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. Neuroimage. 15(4), 787-796.

Schubotz, R.I., & von Cramon, D.Y. (2002b). A blueprint for target motion: fMRI reveals perceived sequential complexity to modulate premotor cortex. Neuroimage, 16(4), 920-935.

Schwartz, J.L., Abry, C., Boë, L.J., and Cathiard, M.-A. (2002). Phonology in a Theory of Perception-for-Action-Control. In J. Durand and B. Laks (Eds.) *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, 254-280.

Schwartz, J.L., Boë, L.J., & Abry, C. (2007). Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception-for-Action-Control Theory (PACT). In M.J. Solé, P. Beddor & M. Ohala (eds.) *Experimental Approaches to Phonology* (pp. 104-124). Oxford University Press.

Scott, S.K. & Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. Trends Neurosci., 26: 100–107.

Shiller, D., Sato, M., Gracco, V. & Baum, S. (submitted). Perceptual recalibration of speech sounds following motor learning.

Skipper, J.I., Nusbaum, H.C. & Small, S.L. (2005). Listening to talking faces: Motor cortical activation during speech perception. NeuroImage, 25: 76-89.

Skipper, J.I., Van Wassenhove, V., Nusbaum, H.C. & Small, S.L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cerebral Cortex, 17(10): 2387-2399.

Stetson, R.H. (1988). Motor Phonetics. A Retrospective Edition. J.A. Scott Kelso & K.G. Munhall (eds.). Boston, Toronto, San Diego : Little, Brown & Co.

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E.E.Davis Jr. and P.B.Denes (Eds.), *Human Communication: A Unified View*. New-York: Mc Graw-Hill, 51-66.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17: 3-45.

Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. Neuroreport, 11(10), 2289-2292.

Sundara, M., Namasivayam, A.K. & Chen, R. (2001). Observation-execution matching system for speech: A magnetic stimulation study. Neuroreport, 12(7): 1341-1344.

Vigneau, M., Beaucousin, V., Hervé, P.Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B. & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. NeuroImage, 30(4): 1414–1432.

Villacorta, V.M., Perkell, J.S. & Guenther, F.H. (2007). Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception. Journal of the Acoustical Society of America, 122: 2306–2319.

Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 603-623.

Watkins, K.E. & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. Journal of Cognitive Neuroscience, 16(6): 978-987.

Watkins, K.E., Strafella, A.P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41(3): 989-994.

Wilson, S. M. & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. NeuroImage, 33(1): 316-25

Wilson, S.M., Saygin, A.P., Sereno, M.I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nat Neurosci, 7, 701-702.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. Science, 256, 846–849.

Zekveld, A.A., Heslenfeld, D.J., Festen, J.M. & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. NeuroImage, 32, 1826-1836.