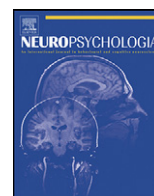




Contents lists available at [SciVerse ScienceDirect](#)

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects

Alessandro D'Ausilio^{a,b}, Joanna Jarmolowska^b, Pierpaolo Busan^b, Ilaria Bufalari^b, Laila Craighero^{b,*}

^a IIT, The Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

^b Dep. S.B.T.A., Section of Human Physiology, University of Ferrara, via Fossato di Mortara 17/19, 44121 Ferrara, Italy

ARTICLE INFO

Article history:

Received 25 October 2010

Received in revised form

14 September 2011

Accepted 15 September 2011

Available online xxx

Keywords:

Speech listening

Perceptual restoration

Phoneme expectation

Coarticulation

Transcranial magnetic stimulation

Tongue corticospinal excitability

ABSTRACT

Humans perceive continuous speech through interruptions or brief noise bursts cancelling entire phonemes. This robust phenomenon has been classically associated with mechanisms of perceptual restoration. In parallel, recent experimental evidence suggests that the motor system may actively participate in speech perception, even contributing to phoneme discrimination. In the present study we intended to verify if the motor system has a specific role in speech perceptual restoration as well. To this aim we recorded tongue corticospinal excitability during phoneme expectation induced by contextual information. Results showed that phoneme expectation determines an involvement of the individual's motor system specifically implicated in the production of the attended phoneme, exactly as it happens during actual listening of that phoneme, suggesting the presence of a speech imagery-like process. Very interestingly, this motoric phoneme expectation is also modulated by subtle coarticulation cues of which the listener is not consciously aware. Present data indicate that the rehearsal of a specific phoneme requires the contribution of the motor system exactly as it happens during the rehearsal of actions executed by the limbs, and that this process is abolished when an incongruent phonemic cue is presented, as similarly occurs during observation of anomalous hand actions. We propose that altogether these effects indicate that during speech listening an attentional-like mechanism driven by the motor system, based on a feed-forward anticipatory mechanism constantly verifying incoming information, is working allowing perceptual restoration.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

In everyday life, noise often reduces the intelligibility of speech. However, despite interference from background noises, we usually perceive speech to be continuous through interruptions. This phenomenon has been demonstrated in studies in which single phonemes were replaced by an extraneous sound. Most listeners reported that the utterance was intact suggesting that they had restored the missing phoneme (Samuel, 1981). These and similar results (Elman & McClelland, 1988; Ganong, 1980; Warren, 1970; Warren & Obusek, 1971) have been interpreted as the evidence that speech perception depends upon the bottom-up confirmation of expectations, and that phonemic restoration depends upon the interplay between the listener's expectations and the acoustic signal. In fact, increasing listeners' expectations of a phoneme by, for example, priming the word, enhances perceptual restoration (Samuel, 1981). Therefore, context information may be used in an anticipatory or "predictive" manner at multiple levels. During

processing of a sentence the most likely candidates are generated anticipating semantic, lexical, or even perceptual features (McClelland & Elman, 1986; McClelland & Rumelhart, 1981).

The ability to predict others' action outcomes has a very important adaptive function not only with speech-related actions. For instance we anticipate forthcoming motor sequences when observing handwriting gestures (Kandel, Orliaguet, & Boë, 1994). When writing two letters (e.g. "ll", "le", "ln") the movement time and the letter shape of the first letter is constrained by the execution of the second one – in analogy with coarticulation in speech production. Coarticulation is the well known phenomenon of merging the production of two consecutive phonemes: the auditory spectral components of a given phoneme are dramatically altered by the articulatory requirements of neighbouring phonemes (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Kandel et al. (1994) showed that subjects can predict the identity of the second letter ("l", "e" or "n") by viewing the production of the first one ("l"). These results demonstrate that kinematic information about written coarticulation supports this anticipatory ability (Orliaguet, Kandel, & Boë, 1997). Similarly, a series of studies have shown the ability to predict the goal of a grasping action from hand pre-shaping during reaching (Orliaguet, Viallon, Kandel, & Coello, 1996),

* Corresponding author. Tel.: +39 0532 455928; fax: +39 0532 455242.
E-mail address: crh@unife.it (L. Craighero).

from grasp postures assumed before starting the action (Fischer, Prinz, & Lotz, 2008) or from the intrinsic properties of the to-be-grasped object (Craighero et al., 2008).

One account of these phenomena is given by the direct matching hypothesis which claims that action understanding and prediction results from a mechanism that maps a perceived action onto motor representations of that action (see Rizzolatti & Craighero, 2004). Each time an individual observes an action done by another individual an analogous motor representation, usually generated during action execution, is activated. Such re-enactment allows the retrieval of the action's motor details and, therefore, facilitates the action's outcome prediction. This view is corroborated by using a developmental approach investigating action perception in children unable to perform the observed action. The first example is given by the study of the so called proactive gaze behaviour (Flanagan & Johansson, 2003). This effect consists of the fact that when subjects observe a block stacking task, the coordination between their gaze and the actor's hand position is predictive, rather than reactive, exactly replicating the gaze–hand coordination shown by the observers when performing the task themselves. This eye predictive capability is absent in children unable to perform the observed action themselves. Infants begin to master the block stacking task at around 7–9 months of life, and it has been shown that 12-month old infants focus on goals in the same way as adults do, whereas 6-month-olds do not. This implies that the development of proactive eye movements might depend on hand action development (Falck-Ytter, Gredebäck, & von Hofsten, 2006). Similarly, handwriting gestures prediction is at chance level between 7 and 9 years, whereas at the age of 11, children are equivalent to adults (see Kandel, Orliaguet, & Boë, 2000). This indicates that perceptual anticipation appears when handwriting control becomes more stable and written coarticulation is clearly observed also at the production level. Altogether these studies confirm that critical cues provided by gestures cannot be perceptually exploited if they cannot be linked to individual's motor competence.

Recent data suggest that principles of the direct matching hypothesis apply also to the perception of speech. Speech perception induces a somatotopic activation (Pulvermüller et al., 2006) of the motor representations relative to the production of the listened phonemes (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Watkins, Strafella, & Paus, 2003). Furthermore, the selective interference with speech production centers (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Mottonen & Watkins, 2009; Sato, Tremblay, & Gracco, 2009), the manipulation of somatosensory feedback typically associated to the articulation of specific phonemes (Gick & Derrick, 2009; Ito, Tiede, & Ostry, 2009) and jaw dynamic perturbation (Nasir & Ostry, 2009) proved effective in altering subject's performance in several speech discrimination tasks.

The aim of the present paper is to show that phoneme expectation determines an involvement of the individual's motor system specifically implicated in the production of the attended phoneme. Furthermore, we aim to demonstrate that phoneme expectation is induced not only by explicit contextual information but also by subtle coarticulation cues of which the listener is not consciously aware.

We submitted participants to a task in which each trial consisted in the presentation of two consecutive pseudo-words separated by a 1000 ms interval. Each pseudo-word began with the syllable BI and could either continue with a tongue-produced phoneme [r], [l] or not [v], [f].

According to the International Phonetic Alphabet (IPA) and in particular according to publications on cross-linguistic phonetics (Ladefoged, 2001) and on Italian phonology (Rogers & d'Arcangeli, 2004), articulators and place of articulation of the considered

consonants vary: [r] and [l] are classified as dental/alveolar while [v] and [f] are classified as labio/dental fricatives. Alveolar consonants are articulated with the tongue against the superior alveolar ridge. In particular [r] is an alveolar trill produced by vibrating the tip of the tongue against the alveolar ridge. Labio/dentals are made by the lower lip acting as the active articulator against the lower edge of the front upper teeth. Magnetic resonance images of the vocal tract revealed that during sustained production of the labio/dental [f] and [v] the tongue is involved only in its posterior body which shows concave cross-sectional shapes (Narayanan, Alwan, & Haker, 1995). Furthermore, the stimuli used in the present experiment required the production of a doubled consonant which in Italian shortens the preceding vowel and lengthens the consonant itself. Therefore, to pronounce the 'rr' of the stimulus "birro" the tip of the tongue must vibrate 2 or 3 times against the alveolar ridge. Similarly, there is a strong impact between the tip of the tongue and the alveolar ridge in producing the sound represented by 'll' in "billo" (Norman, 1937).

These data clearly indicate that, although tongue activity is always necessary during the production of all consonants, tongue tip is much more articulatorily recruited during the realization of [r] and [l] then [f] and [v]. Our experimental rationale was based on this assumption. A similar rationale motivated previous studies recording tongue and lips cortico-spinal excitability (Fadiga et al., 2002; Roy et al., 2008; Watkins et al., 2003) as well as research dealing with the motor somatotopy of speech perception (D'Ausilio et al., 2009; Mottonen & Watkins, 2009).

The structure of each presented stimulus was: [BI]-[double consonant]-[O] (i.e.: BI-RR-O, BI-LL-O, BI-VV-O, BI-FF-O). We induced phoneme expectation by manipulating the percentage of trials in which the two pseudo-words were the same (75%). Consequently, we defined the first pseudo-word as "prime" and the second one as "target".

We applied transcranial magnetic stimulation (TMS) on tongue motor representation in primary motor cortex and we measured tongue evoked motor potentials (MEPs) during a sound gap of 400–450 ms inserted between [BI] and [double consonant] of the target. Therefore, if the motor system is involved in phoneme expectation, tongue corticospinal excitability should be specifically enhanced after presentation of a tongue-related prime (e.g., BI-RR-O) than after a no tongue-related prime (e.g., BI-FF-O), exactly as it would happen if the listener actually perceived the primed phoneme (see Fadiga et al., 2002). Furthermore, we explored if coarticulatory features, in analogy with the work of Kandel et al. (1994), are also able to influence phoneme expectations during speech listening. Following this evidence, in our stimuli, some of the articulatory features of target double consonant necessarily determine specific subtle acoustic features in the previous BI token. Consequently, we included an orthogonal manipulation such that the BI of the target could (e.g., when [BI] is extracted from the pronounced BIRRO) or could not be (e.g., when [BI] is extracted from BIFFO) coarticulated for the following double consonant (e.g., [RR]).

Therefore, the present experiment was designed to verify (1) if tongue corticospinal excitability is specifically modulated during phoneme expectation when the probability of the target presentation is cued by an explicitly presented priming stimulus, (2) if corticospinal excitability is influenced by coarticulatory features determined by an incoming phoneme, and (3) if these two effects interact.

2. Materials and methods

2.1. Subjects

Thirty-one healthy subjects were recruited after obtaining full information about the study and giving their informed consent. They were paid for participation. None had any history of neurological disease, trauma or psychiatric syndrome

Table 1

This table shows the trial types. Prime column indicates the prime signaling the target with 75% probability of presentation, subdivided into Tongue Prime when the pseudo-word contains a tongue-related phoneme, and No-Tongue Prime when the pseudo-word does not contain it. Target column indicates the target split into its two parts: the [BI] presented before (Pre-TMS) and the [double consonant-O] presented after TMS administration (Post-TMS). Regarding the [BI] stimulus, in parenthesis is shown the pseudo-word from which it belonged to. Consequently, the Coarticulation column indicates if coarticulatory features of [BI] correspond (Correct) or not (Wrong) with the primed target. The Same/Different column specifies if the prime and the target were the same or different. Repetitions column shows the number of repetitions for each trial and, in parenthesis, the number of those trials in which TMS was administered. Asterisks indicate those trials whose MEP data were entered into the analysis.

Prime	Target		Co-articulation	Same/Different	Repetitions	
	Pre-TMS	Post-TMS				
Tongue Prime						
BILLO	BI(llo)	LLO	Correct	Same	6(4*)	
	BI(ffo)	LLO	Wrong	Same	6(4*)	
	BI(llo)	FFO	Correct	Different	2(1)	
	BI(ffo)	FFO	Wrong	Different	2(1)	
	BIRRO	BI(rro)	RRO	Correct	Same	6(4*)
		BI(vvo)	RRO	Wrong	Same	6(4*)
		BI(rro)	VVO	Correct	Different	2(1)
		BI(vvo)	VVO	Wrong	Different	2(1)
		No-Tongue Prime				
BIFFO	BI(ffo)	FFO	Correct	Same	6(4*)	
	BI(llo)	FFO	Wrong	Same	6(4*)	
	BI(ffo)	LLO	Correct	Different	2(1)	
	BI(llo)	LLO	Wrong	Different	2(1)	
BIVVO	BI(vvo)	VVO	Correct	Same	6(4*)	
	BI(rro)	VVO	Wrong	Same	6(4*)	
	BI(vvo)	RRO	Correct	Different	2(1)	
	BI(rro)	RRO	Wrong	Different	2(1)	

and had normal hearing. All subjects were Italian native speakers. Procedures were approved by the local ethical committee of the University of Ferrara. The complete experiment could be performed only in eleven subjects showing clear and ss tongue MEPs (mean age, 23.9; SD, 4.7; 4 females; see Section 2.4 for more details).

2.2. Procedure

After signing the informed consent and having familiarized with the experimental procedure, subjects entered the TMS mapping procedure (see Section 2.4 for details). Upon a successful mapping procedure (25–35 min), we let subjects sit on a comfortable chair in front of a table and a computer monitor (about 75 cm distance). Stimuli were acoustically presented through semi-professional headphones (AKG). Subjects' responses were acquired by a custom-made response box consisting of two buttons (one on the left, one on the right side), while both stimuli presentation and behavioural data recordings were controlled by an E-Prime script (Psychology Software Tools, Inc.). Subjects' responses were given with the left hand, ipsilateral to the TMS scalp administration. We did so in order to avoid motor programming interference or any TMS interference to response selection. The correct synchronization between auditory stimuli and TMS occurrence was preliminary tested by feeding both the PC sound-card output and the TMS trigger to an external A/D board with an internal hardware clock (CED, micro1401). Subjects first completed 16 training trials and then the experiment started. The experiment lasted roughly 15 min including 64 trials, 40 of which had TMS (see Table 1). The whole experimental session lasted about 1 h.

2.3. Stimuli and trial structure

Stimuli consisted in 4 pseudo-words read by a female actress. Pseudo-words were phonotactically legal Italian pronounceable sequences of CVCCV sounds. All pseudo-words included an initial syllable "BI" followed by either "RRO", "LLO" (containing a double consonant requiring tongue involvement during articulation) and "FFO" or "VVO" (not containing a double consonant requiring tongue involvement during articulation). Stimuli lasted ~950 ms.

Each trial consisted in the presentation of two consecutive pseudo-words with a 1000 ms interval. The two stimuli were the same in 75% of trials. At the end of the second pseudo-word presentation an on-screen message prompted the subject to perform one out of two tasks. In some trials they had to press one of two buttons at will, whereas in the remaining trials they had to decide whether the two pseudo-words were the same or different and congruently press the corresponding button. Response buttons position was counterbalanced between-subjects. These tasks were devised in order to have the subject carefully listening to the stimuli, without forcing the preparation of any response strategy (50% of trials in each task, randomly).

Furthermore, we included a stimulus manipulation such that the BI syllable of the second pseudo-word could or could not be correctly coarticulated to the subsequent double consonant. By using an audio editing software (Audacity) we isolated the syllable BI from the four original sound tracks. Subsequently, we mixed the four obtained BI syllables assembling the following stimuli: BI(llo)LLO, BI(ffo)FFO, BI(rro)RRO, BI(vvo)VVO, defined as correctly coarticulated because the BI syllable

was extracted from the presented target (in parentheses), and BI(ffo)LLO, BI(llo)FFO, BI(vvo)RRO, BI(rro)VVO, defined as wrongly coarticulated because the BI syllable was not extracted from the presented target. A random 300–350 ms gap between the BI syllable and the subsequent double consonant was inserted.

Therefore, we could have four different experimental conditions:

- "same" condition (the first and the second pseudo-word are the same), correct coarticulation in the second pseudo-word (e.g., BILLO–BI(llo)LLO)
- "same" condition, wrong coarticulation in the second pseudo-word (e.g., BILLO–BI(ffo)LLO)
- "different" condition (the first and the second pseudo-word are different), correct coarticulation (e.g., BILLO–BI(ffo)FFO)
- "different" condition, wrong coarticulation (e.g., BILLO–BI(llo)FFO).

See Table 1 for all possible trial types.

To verify whether coarticulation features in the BI stimuli could be easily detected by subjects we run a control behavioural experiment. Six subjects (who did not participate in the TMS experiment) listened, via headphones, to a sequence of BI extracted from the four pseudo-words (BIRRO, BILLO, BIVVO, BIFFO) presented in the main experiment. After each stimulus presentation subjects were forced to indicate which of two visually presented pseudo-words the acoustic stimulus belonged to. The two pseudo-words were presented in a balanced order on the right and left side of a screen placed in front of subjects. There were four experimental conditions, presented in a randomized order: (i) Acoustic BI(rro) – visual BIRRO and BIVVO; (ii) Acoustic BI(vvo) – visual BIRRO and BIVVO; (iii) Acoustic BI(llo) – visual BILLO and BIFFO; (iv) Acoustic BI(ffo) – visual BILLO and BIFFO. Subjects were asked to press one of two buttons, with their right index and middle finger, spatially corresponding to the selected visual pseudo-word. The total number of trials was 80, twenty repetitions for each experimental condition. We applied a series of single-sample *t*-tests to compare each of the four conditions against the chance level. Results indicated that responses were not different than chance level for all stimuli, separately (BI(llo): $t(5) = 0.1281$, $p = 0.9$; BI(rro): $t(5) = -0.9455$, $p = 0.39$; BI(vvo): $t(5) = 1.9897$, $p = 0.1$; BI(ffo): $t(5) = 1.7112$, $p = 0.15$). This indicates that, in a forced-choice paradigm, subjects were not aware of the presence of coarticulation information in the stimuli presented in the main experiment.

2.4. TMS and EMG

TMS stimulation was delivered through a figure-eight 70 mm coil and a Magstim 200 stimulator (Magstim, Whitland, UK) to the tongue motor representation in the left hemisphere. The tongue motor representation was found by searching the First dorsal Interosseus (FDI) area, and the FDI resting motor threshold by using standard protocols (Rossini et al., 1994). Then the coil was moved about 4 cm laterally and 1 cm anteriorly from the FDI hot spot. Stimulator output was then increased in 5% steps until a tongue MEP could be shown (max 70% of the stimulator output). After the first few MEPs could be reliably recognized, the coil was parametrically moved in a 2 by 2 cm radius and the coil orientation was rotated roughly in 10° steps in order to maximize stability and amplitude of MEPs. The stimulation intensity was then set in

Table 2

This table shows raw data. Averaged peak-to-peak MEP size in all four conditions. Values are in millivolt plus/minus standard error of the mean.

	Correct coarticulation	Wrong coarticulation
Tongue	0.277 ± 0.075	0.261 ± 0.062
No-Tongue	0.220 ± 0.057	0.256 ± 0.073

order to have 5 out of 5 MEPs, clearly discernible from the background EMG activity (between 100 and 200 μ V; stimulation range from 60 to 68% of maximal stimulator output). MEPs were recorded with a wireless EMG system (Aurion, ZeroWire EMG) since this system proved to be optimal in reducing the TMS artefact in facial muscle recordings. This system automatically provides the difference in electric potential between two electrodes as a measure of muscle electrical activity. Two Ag/AgCl cup electrodes were placed on the right dorsal surface of the tongue (10–15 mm from midline, 10 mm from tongue tip) with an inter-electrode distance of 10–20 mm, using a surgical glue (Hystoacryl, B. Braun Surgical SA). The electrodes' position allowed the best recording of EMG activity of the anterior part of the tongue. The signal was band-pass filtered (50–1000 Hz) and digitized (2 kHz).

Such electrode placing and TMS mapping procedure could not be completed in all 31 subjects for several reasons. First of all, in 9 subjects it was not possible to run the full experiment because of technical problems (no stable electrode contact, strong basal EMG activity, the stimulation was not well tolerated due to the high intensities at such lateral scalp sites, unstable MEP recordings). Finally and most importantly, 11 subjects were discarded because artefacts spoiled the signal. Specifically we noted at least 2 kinds of recurrent artefacts. The first was characterized by an average latency of 4–6 ms with a MEP-like morphology, including both a positive and a negative deflection. Due to its early latency, this was likely caused by an ipsilateral cranial nerve stimulation (*hypoglossus*) rather than a cortical stimulation. The second kind of artefact, instead, starting later around 13–16 ms, was slower and in most cases included only a positive component. This late wave, considering its latency and morphology, was possibly mediated by the electrodes' displacement caused by an ipsilateral masseter muscle contraction. We therefore analyzed the data only of those subjects showing no EMG pre-activation and reliable MEPs responses with a latency between 8 and 11 ms (Cruccu, Inghilleri, Berardelli, Romaniello, & Manfredi, 1997; Paradiso, Cunic, Gunraj, & Chen, 2005; Svensson, Romaniello, Arendt-Nielsen, & Sessle, 2003; Svensson, Romaniello, Wang, Arendt-Nielsen, & Sessle, 2006).

TMS was delivered 100 ms after the end of presentation of the BI syllable, followed by a random 300–350 ms gap before double consonant presentation.

2.5. Design, data analysis and statistics

We had a same-different paradigm with a 75–25% ratio in order to let the first pseudo-word be sufficiently predictive of the second one. Consequently, a total of 64 trials of which 48 same, and 16 different were presented (see Table 1 for more details). TMS was delivered in 32 out of 48 “same” trials, and in 8 out of 16 “different” trials. We submitted to analysis the “same” trials only. Of the 32 TMS-stimulated “same” trials, orthogonally to the priming condition, 16 trials were correctly articulated with the expected phoneme and 16 trials were wrongly articulated with the expected phoneme (see Table 1). Therefore, the experiment consisted of a 2 (Articulation [Tongue vs. No-Tongue]) \times 2 (Coarticulation [Correct vs. Wrong]) design.

The dependent variable was the mean MEP peak-to-peak amplitude (32 MEPs, 8 each condition). Table 2 contains the raw MEP amplitude data in all the experimental conditions. These single MEP amplitudes were transformed into z-scores for each subject and then averaged for every condition.

3. Results

The data for all subjects were submitted to an analysis of variance (ANOVA) considering as within-subjects factors Articulation (Tongue vs. No-Tongue) and Coarticulation (Correct vs. Wrong). The two-way ANOVA reported a significant main effect of Articulation ($F(9,1) = 23.55$; $p = 0.0009$) only: Tongue corticospinal excitability was enhanced after presentation of a prime including a tongue-produced phoneme (Tongue Prime: BILLO, BIRRO).

Furthermore, we searched for whether an implicit elaboration of coarticulatory features might have modulated such effect. Positive and negative z-values significantly different from zero represent increased and reduced modulation with respect to the average individual responses amplitude, respectively. Therefore, mean values for each condition were tested (planned comparisons) against the no-modulation hypothesis that, in normalized values, is represented by zero value. This analysis reported that after presentation of both Tongue and No-Tongue primes, results were different from

zero only when the BI syllable was coarticulated at the same articulation site as the prime was (Tongue-Correct: $t_9 = 2.72$, $p = 0.023$; No-Tongue-Correct: $t_9 = -2.59$, $p = 0.028$; Tongue-Wrong: $t_9 = 1.45$, $p = 0.18$; No-Tongue-Wrong: $t_9 = -1.16$, $p = 0.27$). For instance a Tongue prime led to a significant increase of tongue MEPs only when the BI syllable in the target pseudo-word was coarticulated for a tongue-produced sound.

Our results indicate that, after presentation of a phonological prime, tongue corticospinal excitability is enhanced when a tongue-involving phoneme is attended with respect to when a non tongue-involving phoneme is attended. However, this modulation is influenced by coarticulatory cues: only when the BI syllable is extracted from the attended pseudo-word (either tongue or non tongue-involving) corticospinal excitability is significantly modulated (Fig. 1).

4. Discussion

The exact role played by the motor system in speech perception is a matter of recent debate (Lotto, Hickok, & Holt, 2009; Toni, de Lange, Noordzij, & Hagoort, 2008). Some authors claim that speech sounds encoding predominantly require the ventral language pathway in the temporal cortex, thus not requiring a motor component (Scott & Johnsrude, 2003). On the other hand, a recent series of studies supported a motor recruitment, and thus a leading role for the dorsal language route (D'Ausilio et al., 2009; Meister et al., 2007; Mottonen & Watkins, 2009; Sato et al., 2009). In particular, Meister et al. (2007) showed that the application of repetitive TMS to the premotor cortex disrupts subjects' ability to perform a phonetic discrimination task, and D'Ausilio et al. (2009) found a double dissociation after TMS administration to motor cortex controlling lips and tongue during the discrimination of lip- and tongue-articulated phonemes. However, speech perception is a very complex cognitive ability which requires multiple computations surely involving more than a motor recruitment when phoneme recognition is used for word comprehension. In particular, low-level phonetic representations involved in phoneme discrimination may not be the same as the low-level phonetic representations used in word comprehension (see Hickok & Poeppel, 2007). Nevertheless, Devlin and Aydelott (2009) commented that “Taken together, these two TMS studies (D'Ausilio et al., 2009; Meister et al., 2007) provide the strongest evidence to date that the motor system is not only activated during speech perception, but this activation also plays a role in discriminating specific phonemes” (pg. R199). However, they suggested that “speech production regions may be recruited to aid speech comprehension, perhaps using a form of implicit motor simulation” (pg. R199). With the term “implicit motor simulation” the authors possibly refer to the meaning attributed to it by Gallese (2003) as a direct, automatic, and unconscious process of simulation determined by an external event of which, however, the observer have to be aware. Therefore, it might be questioned to which extent the motor contribution to speech perception may also be based on an automatic process determined by stimuli of which subjects are unaware.

Generally speaking, the present study reveals new experimental evidence, namely that tongue corticospinal excitability is specifically enhanced during expectation of a tongue-involving phoneme. This effect may be attributed to a motor imagery process or, in other words, to a voluntary motor simulation: Subjects may have mentally reiterated the stimulus to accomplish the task. In such a view, we were indeed recording a motor excitability-enhancement induced by a specific speech imagery process addressing a specific motor representation. To our knowledge, speech motor imagery has been only investigated in fMRI studies showing mixed results as far as the primary motor cortex involvement is concerned (Callan,

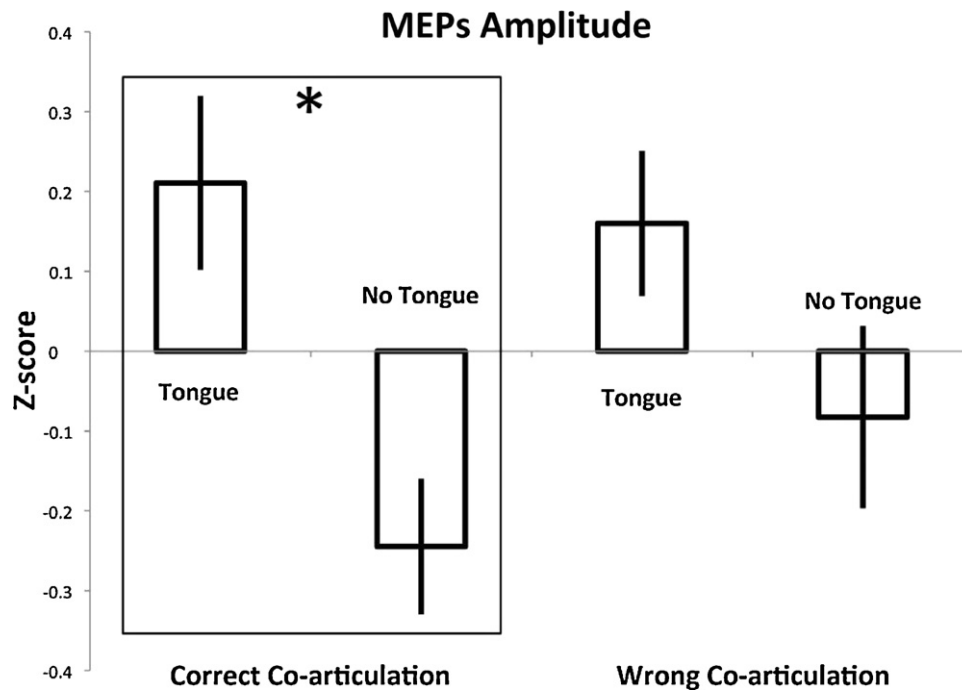


Fig. 1. Results. Z-score MEPs amplitudes in four conditions. MEPs in tongue-articulated (Tongue; 'r' and 'l') and control stimuli (No-Tongue; 'f' and 'v') are shown with and without correct coarticulation. Bars represent standard error of the mean and the asterisk show significant effects in the planned comparisons analysis.

Jones, Callan, & Akahane-Yamada, 2004; Kleber, Birbaumer, Veit, Trevorrow, & Lotze, 2007; Shergill et al., 2001). Our TMS procedure let us demonstrate that speech imagery-like process is supported by activity in the motor system exactly as it happens during motor imagery of other effectors (Fadiga et al., 1999; see Fadiga & Craighero, 2004).

However, speech imagery was abolished when an incongruent phonemic cue was presented. In fact, wrong coarticulation cues cancelled the modulation of tongue corticospinal excitability. Similar results were also found during observation of hand actions. Gangitano, Mottaghy, and Pascual-Leone (2004) had subjects watching a video clip of a hand approaching and grasping a ball in which maximal finger aperture was substituted with an unpredictable closure. These authors showed that FDI motor excitability was suppressed during the observation of the incongruent action, as if the activated motor plan was discarded when features of the presented movement ceased to match those of the attended one.

Interestingly enough, listeners were at chance level when asked to explicitly decode coarticulatory features. Therefore, present results seem to indicate that the involvement of the motor system during speech perception is mainly based on contextual cues that may be further modulated by the unconscious detection of subtle coarticulation cues. As a consequence, here we show for the first time that during speech listening the motor command is reproduced in great detail, even in its coarticulation characteristics.

Moreover, some further striking similarities can be found in the modulation of motor excitability during observation of hand actions (Borroni, Montagna, Cerri, & Baldissera, 2005). In fact, in Borroni's study participants watched a cyclic flexion-extension movement of the wrist while MEPs were elicited in their right forearm extensor and flexors muscles. Results showed that the pattern of flexor-extensor excitability had the same period of the observed movement and was phase advanced, as it happens in muscle activation respect to real movement. These subjects have only access to visual kinematic cues and thus they can infer muscle pattern timing only using a simulation strategy. Much in the same way,

speech coarticulation cues are not consciously perceived but are simulated in the listener's motor system. In fact, motor rehearsal of the attended phoneme is sufficient to allow the prediction and automatic detection of an incongruently coarticulated stimulus by the motor system despite the fact that individuals were not able to tell them apart.

We hypothesize that a feed-forward anticipatory mechanism (a speech imagery-like mechanism) based on contextual cues on next target probability is constantly verifying incoming information elaborated at an unconscious level (coarticulation detection) as an online feed-back based control strategy. Such control strategy is indeed not new in motor neuroscience. In fact, the relationship between action and perception, at the level of motor control, is thought to be organized around similar principles. People may use internal predictive models to generate goal-directed actions (Desmurget & Grafton, 2000). More specifically, during goal-directed action, internal models provide sensory expectations that are used to monitor and control movements. Analogously, it has been argued that the same internal modelling mechanisms are reused when we encode another's action in terms of our own motor repertoire (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Fazio et al., 2009; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Craighero, 2004).

The computational advantage of internal models is that they offer a clear mechanism for the anticipation of future sensory-motor or goal states and, as already mentioned, there is ample evidence at multiple levels regarding motor anticipatory mechanisms in perceptual tasks. Subjects can indeed anticipate the next motor event from subtle kinematic changes (Fischer et al., 2008; Kandel et al., 1994; Orliaguet et al., 1996) or object features (Craighero et al., 2008). Similarly, the observer's oculo-motor behaviour anticipates the goal location of observed actions (Flanagan & Johansson, 2003). More interestingly, this anticipatory capability appears only after subjects skilfully master the behaviour of interest (Falck-Ytter et al., 2006; Kandel et al., 2000). Therefore, it is possible that motor experience on a given task enables a rich sensory-motor encoding of that skill.

Similar models have been proposed in the domain of speech perception (McClelland & Elman, 1986; McClelland & Rumelhart, 1981; Pickering & Garrod, 2007). Here, internal modelling generates predictions at all levels, including the phonological, syntactic and semantic ones. Predictions are then compared with the auditory input, and corrections can be done to finally generate an adequate interpretation, much in the same way we observed in our present study, at a psychophysiological level. The behavioural advantage of such a mechanism could be that of perceptual restoration and might emerge when dealing with degraded or missing information. Phonemic restoration may take advantage of motor representation rehearsal occurring as soon as a specific phoneme candidate reaches the probability threshold to be perceived. In fact, several studies showed that anterior language areas might be recruited for sensory decisions and completion during sub-optimal listening conditions (Binder, Liebenthal, Possing, Medler, & Ward, 2004; Boatman & Miglioretti, 2005; Moineau, Dronkers, & Bates, 2005) or the illusory gap filling phenomenon (Shahin, Bishop, & Miller, 2009).

In this sense, the motor system might furnish an attentional-like mechanism able to prime perceptual processes (Rizzolatti, Riggio, Dascola, & Umiltà, 1987; Rizzolatti, Riggio, & Sheliga, 1994; Rizzolatti & Craighero, 1998). Anticipatory processes, guided by articulatory gestures, may be firstly activated by a partial auditory feature extraction and subsequently, using contextual information and probability maps, may be employed for sensory completion of degraded speech (Shahin et al., 2009). Sensory completion might be mediated by anticipatory mechanisms such as those proposed for general sensory-motor control (Wolpert & Kawato, 1998). Forward-inverse couples are based upon the ability of the system to predict either a sensory state given the motor command, or the motor state given the sensory state. These couples are built during development via active movement production and sensory feedback recording – such as the speech-babbling phase (Guenther, Ghosh, & Tourville, 2006). After development, these sensory-motor maps might be used to cope with a natural context where we are continuously exposed to incomplete or noisy sensory information. Therefore, we envisage speech perception as an active process searching for relevant features among several sources of noise. This search might be directed towards salient features via an attentional-like mechanism, driven by the motor system.

Acknowledgments

This study was supported by the following grants: MIUR PRIN 2008, FAR 2008 and FAR 2009 from the University of Ferrara to L.C. and by the E.C. project Poeticon.

References

- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A. & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7, 295–301.
- Boatman, D. F. & Miglioretti, D. L. (2005). Cortical sites critical for speech discrimination in normal and impaired listeners. *The Journal of Neuroscience*, 25, 5475–5480.
- Borrioni, P., Montagna, M., Cerri, G. & Baldissera, F. (2005). Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement. *Brain Research*, 1065, 115–124.
- Callan, D. E., Jones, J. A., Callan, A. M. & 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.
- Craighero, L., Bonetti, F., Massarenti, L., Canto, R., Fabbri Destro, M. & Fadiga, L. (2008). Temporal prediction of touch instant during observation of human and robot grasping. *Brain Research Bulletin*, 75, 770–774.
- Crucchi, G., Inghilleri, M., Berardelli, A., Romaniello, A. & Manfredi, M. (1997). Cortical mechanisms mediating the inhibitory period after magnetic stimulation of the facial motor area. *Muscle Nerve*, 20, 418–424.
- D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C. & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 381–385.
- Desmurget, M. & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, 4, 423–431.
- Devlin, J. T. & Aydelott, J. (2009). Speech perception: Motoric contributions versus the motor theory. *Current Biology*, 19, R198–R200.
- Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27, 143–165.
- Fadiga, L., Buccino, G., Craighero, L., Fogassi, L., Gallese, V. & Pavesi, G. (1999). Corticospinal excitability is specifically modulated by motor imagery: A magnetic stimulation study. *Neuropsychologia*, 37, 147–158.
- Fadiga, L. & Craighero, L. (2004). Electrophysiology of action representation. *Journal of Clinical Neurophysiology*, 21, 157–169.
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *The European Journal of Neuroscience*, 15, 399–402.
- Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608–2611.
- Falck-Ytter, T., Gredeback, G. & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, 9, 878–879.
- Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., et al. (2009). Encoding of human action in Broca's area. *Brain*, 132, 1980–1988.
- Fischer, M. H., Prinz, J. & Lotz, K. (2008). Grasp cueing shows obligatory attention to action goals. *Quarterly Journal of Experimental Psychology*, 61, 860–868.
- Flanagan, J. R. & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424, 769–771.
- Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358, 517–528.
- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593–609.
- Gangitano, M., Mottaghy, F. M. & Pascual-Leone, A. (2004). Modulation of premotor mirror neuron activity during observation of unpredictable grasping movements. *European Journal of Neuroscience*, 20, 2193–2202.
- Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.
- Gick, B. & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462, 502–504.
- Guenther, F. H., Ghosh, S. S. & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96, 280–301.
- Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Ito, T., Tiede, M. & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1245–1248.
- Kandel, S., Orliaguet, J.-P. & Boë, L.-J. (1994). Visual perception of motor anticipation in the time course of handwriting. In C. Faure, P. Keuss, G. Lorette, & A. Vinter (Eds.), *Advances in handwriting and drawing: A multidisciplinary approach* (pp. 379–388). Paris: Europaia.
- Kandel, S., Orliaguet, J.-P. & Boë, L.-J. (2000). Detecting anticipatory events in handwriting movements. *Perception*, 29, 953–964.
- Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T. & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, 36, 889–900.
- Ladefoged, P. (2001). *Vowels and consonants. An introduction to the sounds of languages*. Malden, MA, USA: Blackwell.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Lotto, A. J., Hickok, G. S. & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13, 110–114.
- McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. I. An account of basic findings. *Psychological Review*, 88, 375–407.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D. & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17, 1692–1696.
- Moineau, S., Dronkers, N. F. & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech, Language and Hearing Research*, 48, 884–896.
- Mottonen, R. & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29, 9819–9825.
- Narayanan, S. S., Alwan, A. A. & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 98, 1325–1347.
- Nasir, S. M. & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 20470–20475.
- Norman, H. L. (1937). Reduplication of consonants in Italian pronunciation. *Italica*, 14, 57–63.

- Orliaguet, J.-P., Kandel, S. & Boë, L.-J. (1997). Visual perception of cursive handwriting: Influence of spatial and kinematic information on the anticipation of forthcoming letters. *Perception*, *26*, 905–912.
- Orliaguet, J.-P., Viallon, S., Kandel, S. & Coello, Y. (1996). Perceptual anticipation in sequential grasping movements. In *Proceedings of the XXVI international congress of psychology* Montreal, Hove, East Sussex: Psychology Press.
- Paradiso, G. O., Cunic, D. I., Gunraj, C. A. & Chen, R. (2005). Representation of facial muscles in human motor cortex. *Journal of Physiology*, *567*, 323–336.
- Pickering, M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science*, *11*, 105–110.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O. & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7865–7870.
- Rizzolatti, G. & Craighero, L. (1998). Spatial attention: Mechanisms and theories. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science: Vol. 2. Biological and cognitive aspects* (pp. 171–198). East Sussex, England: Psychology Press.
- Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
- Rizzolatti, G., Riggio, L. & Sheliga, B. M. (1994). Space and selective attention. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 231–265). Cambridge, MA: MIT Press.
- Rizzolatti, G., Riggio, L., Dascola, I. & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*, 31–40.
- Rogers, D. & d'Arcangeli, L. (2004). Italian. *Journal of the International Phonetic Association*, *34*, 117–121.
- Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: Basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalography and Clinical Neurophysiology*, *91*, 79–92.
- Roy, A. C., Craighero, L., Fabbri-Destro, M. & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology, Paris*, *102*, 101–105.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology. General*, *110*, 474–494.
- Sato, M., Tremblay, P. & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*, 1–7.
- Scott, S. K. & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*, 100–107.
- Shahin, A. J., Bishop, C. W. & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage*, *44*, 1133–1143.
- Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M. & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*, 241–253.
- Svensson, P., Romaniello, A., Arendt-Nielsen, L. & Sessle, B. J. (2003). Plasticity in corticomotor control of the human tongue musculature induced by tongue-task training. *Experimental Brain Research*, *152*, 42–51.
- Svensson, P., Romaniello, A., Wang, K., Arendt-Nielsen, L. & Sessle, B. J. (2006). One hour of tongue-task training is associated with plasticity in corticomotor control of the human tongue musculature. *Experimental Brain Research*, *173*, 165–173.
- Toni, I., de Lange, F. P., Noordzij, M. L. & Hagoort, P. (2008). Language beyond action. *Journal of Physiology, Paris*, *102*, 71–79.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392–393.
- Warren, R. M. & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, *9*, 358–362.
- Watkins, K. E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989–994.
- Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*, 1317–1329.