

SHORT COMMUNICATION

Speech listening specifically modulates the excitability of tongue muscles: a TMS study

Luciano Fadiga,¹ Laila Craighero,^{1,2} Giovanni Buccino² and Giacomo Rizzolatti²

¹Dipartimento di Scienze Biomediche e Terapie Avanzate, Sezione di Fisiologia Umana, Università di Ferrara, via Fossato di Mortara 17/19, 44100 Ferrara, Italy

²Istituto di Fisiologia Umana, Università di Parma, via Volturno 39, 43100 Parma, Italy

Keywords: mirror neurons, motor-evoked potentials, motor system, motor theory of speech perception

Abstract

The precise neural mechanisms underlying speech perception are still to a large extent unknown. The most accepted view is that speech perception depends on auditory-cognitive mechanisms specifically devoted to the analysis of speech sounds. An alternative view is that, crucial for speech perception, it is the activation of the articulatory (motor) gestures that generate these sounds. The listener understands the speaker when his/her articulatory gestures are activated (motor theory of speech perception). Here, by using transcranial magnetic stimulation (TMS), we demonstrate that, during speech listening, there is an increase of motor-evoked potentials recorded from the listeners' tongue muscles when the presented words strongly involve, when pronounced, tongue movements. Although these data do not prove the motor theory of speech perception, they demonstrate for the first time that word listening produces a phoneme specific activation of speech motor centres.

Introduction

There are two major accounts on how speech is perceived. The first posits that speech is formed by sounds that rely for their production and perception on separate processes, neither of which is distinctly linguistic (Kuhl & Miller, 1975; Klatt, 1979; Massaro & Cohen, 1983, 1990; Sussman, 1989). The second account (motor theory of speech perception) maintains that the ultimate constituents of speech are not sounds, but articulatory gestures that have evolved exclusively for the service of language (Lieberman *et al.*, 1967; Lieberman & Mattingly, 1985; Lieberman & Wahlen, 2000). A cognitive translation into phonology is not necessary because the articulatory gestures are phonologic in nature. Furthermore, speech perception and speech production processes use a common repertoire of motor primitives that, during speech production, are at the basis of articulatory gesture generation, while during speech perception, are activated in the listener as the result of an acoustically evoked motor 'resonance'.

The first account is the most widely accepted. It explains speech perception in accord with the classical view that the brain is a hierarchically organized structure with a top level formed by associative cortices and low levels formed by sensory (for perception) and motor (for action) cortices. The second account, on the contrary, maintains the rather contrainuitive view that the motor system intervenes in perception. This account has been often criticized as too speculative and devoid of firm experimental evidence. More recently, however, the plausibility of the motor theory of speech perception has been reinforced by a series of new experimental data, derived from the neurophysiology of the motor system.

Single neuron studies showed that, in the ventral premotor cortex of the monkey, there are neurons that discharge both when the monkey performs a specific action and when it observes an individual making a similar action (Gallese *et al.*, 1996). Subsequent brain imaging studies (Grafton *et al.*, 1996; Rizzolatti *et al.*, 1996a; Grèzes *et al.*, 1998; Iacoboni *et al.*, 1999; Nishitani & Hari, 2000; Buccino *et al.*, 2001) and neurophysiological investigations of the motor cortex excitability in human subjects (Fadiga *et al.*, 1995; Strafella & Paus, 2000) showed that, in humans, the observation of motor actions also activates the motor circuits involved in the generation of the seen movements. It has been proposed that this observation/execution matching system (mirror system) is the physiological expression of a brain mechanism involved in 'understanding' the actions made by others; actions are recognized because both the agent and the observer share the same motor repertoire (Rizzolatti *et al.*, 1996b). The homology between this newly discovered visuo-motor mechanism and the mechanism postulated by the motor theory of speech perception is obvious.

In the present study we tested whether, as the motor theory of speech perception postulates, the passive listening of verbal material induces in the listener the automatic activation ('resonance') of the cortical centres involved in speech production.

Materials and methods

General procedures

Nine normal, right handed, volunteers participated in the experiment after giving their informed consent. All experimental procedures were approved by the local Ethical Committee. Participants were seated on an armchair with the elbow flexed at 90° and their hands half pronated in a totally relaxed position. The backrest was inclined at

Correspondence: Professor Luciano Fadiga, ¹Dipartimento di Scienze Biomediche, as above.
E-mail: fdl@unife.it

TABLE 1. Verbal stimuli used in the experiment

Labiodental fricative consonant, 'r'		Lingua-palatal fricative consonant, 'ff'	
Words	Pseudo-words	Words	Pseudo-words
birra (bier)	berro	baffo (moustache)	biffo
carro (cart)	firra	beffa (hoax)	ciffo
cirro (cirrus)	forro	buffo (funny)	leffa
farro (spelt)	furra	ceffo (snout)	meffa
ferro (iron)	marro	coffa (crow's nest)	paffo
mirra (myrrh)	merro	goffo (clumsy)	peffa
morra (morra)	parro	muffa (mold)	poffa
porro (leek)	perro	puffo (smurf)	seffa
serra (greenhouse)	vorro	tuffo (dive)	viffo
terra (ground)	virro	zaffo (plug)	voffo

For words and pseudo-words, each category of stimuli (double 'r', "rr" and double 'f', "ff") are shown. Words pertaining to each category were balanced according to their frequency in Italian language. Pseudo-words had phonetic structure and spectral complexity comparable to those of words and were regular for Italian language pronunciation.

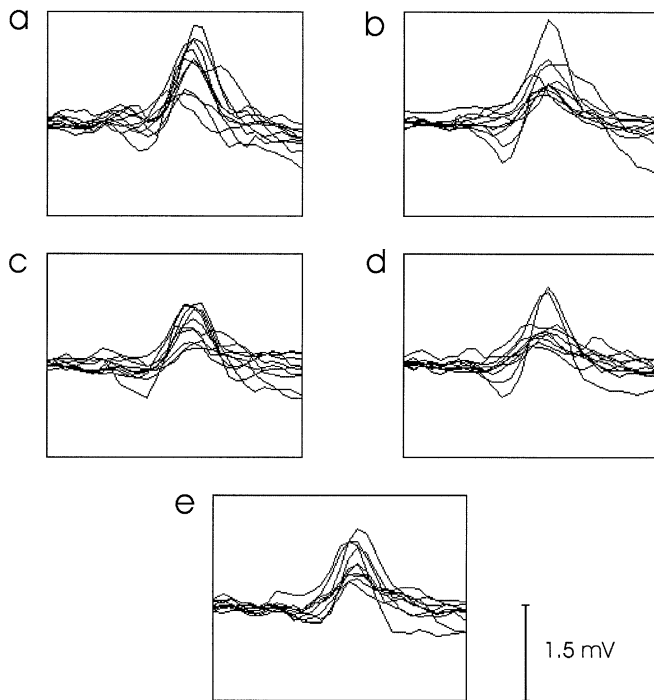


FIG. 1. Motor-evoked potentials recorded from the tongue muscles of one subject in all experimental conditions. (a) Words with embedded double 'r' consonants; (b) pseudo-words with embedded double 'r' consonants; (c) words with embedded double 'f' consonants; (d) pseudo-words with embedded double 'f' consonants; (e) bitonal sounds. In each panel, all MEPs are shown superimposed in the interval 6–20 ms after TMS. Note the difference in amplitude between conditions (a) and (b) with respect to conditions (c–e). The observed motor facilitation is under-threshold for overt movement generation, as assessed by high sensitivity electromyography showing that during the task the participants' tongue muscles were absolutely relaxed.

40° and the participants' head lay on a headrest in order to maintain a comfortable and stable head position. Subjects were instructed to listen carefully to a sequence of acoustically presented verbal and nonverbal stimuli. In order to keep an adequate level of attention, participants were sometimes asked during the experiment to report if the last verbal stimulus heard was a word or a pseudo-word.

Acoustic stimuli were delivered through a loudspeaker connected to a personal computer that was used to present them in a randomized sequence. Stimuli consisted of 20 disyllabic words, 20 disyllabic pseudo-words and 20 bitonal sounds of equivalent intensity (peak, 78–80 dB) and duration. Words and pseudo-words were selected according to a consonant–vowel–consonant–consonant–vowel (cvccv) scheme (see Table 1). The embedded consonants in the middle of words and of pseudo-words were either a double 'f' (labiodental fricative consonant that, when pronounced, requires slight tongue tip mobilization) or a double 'r' (lingua-palatal fricative consonant that, when pronounced, requires strong tongue tip mobilization). Bitonal sounds lasted about the same time as verbal stimuli and their intonation pattern replicated that of the verbal stimuli.

Participants' left motor cortex was magnetically stimulated by using a single pulse TMS (ESAOTE, Biomedica, Italy). Magnetic stimuli were delivered through an eight-shaped coil placed on the skull with the handle positioned in a medio-lateral orientation. The experiment was subdivided into a mapping session and an experimental session. In the mapping session, the stimulation was made by applying magnetic stimuli on predetermined positions on a grid with a resolution of one-centimeter, drawn on a bathing cap wore by the participants. The coordinate origin was located at the Cz reference point determined according to the international 10–20 EEG system, Jasper (1958). The cortical representation of anterior tongue muscles and of hand right opponens pollicis (OP) muscle was mapped by moving the centre of the coil by one centimeter-steps according to the grid. Stimulus intensity was adjusted in order to determine the motor threshold for the recorded muscles. Mapping was then performed by using stimulus intensity at 120% of the measured motor threshold. The OP muscle map was assessed in order to have a reference point for subsequent localization of tongue muscle representation. Motor-evoked potentials (MEPs) were recorded from OP muscle by using Ag–AgCl surface electrodes (diameter 6 mm) glued to the participants' skin according to a bipolar disposition. In order to record MEPs from tongue muscles, a silicon cast of each participant's mouth was prepared. On each cast Ag–AgCl electrodes (diameter 3 mm) were placed, on the surface in contact with the tongue, 1 cm medial to the print of the premolar teeth bilaterally. The tongue muscles representation was located more anteriorly and laterally with respect to the OP representation.

In the experimental session, which generally occurred within three days after the mapping session, the excitability of motor cortex during

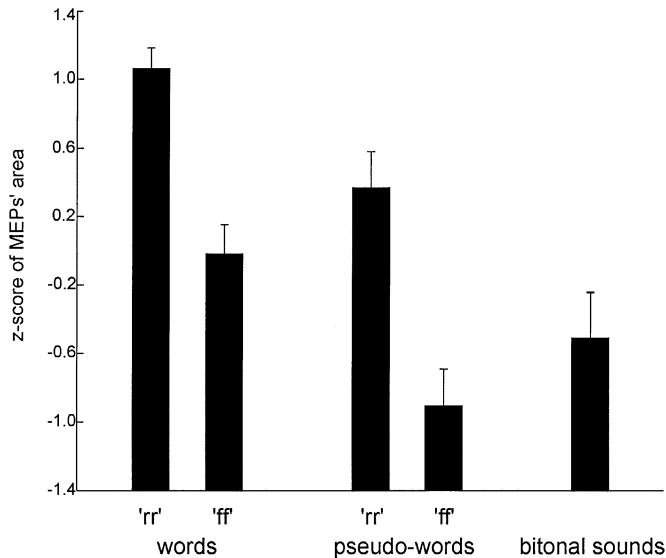


FIG. 2. Average value (+ SEM) of intrasubject normalized MEPs total areas for each condition. Data from all subjects; 'rr' and 'ff' refer to verbal stimuli containing a double lingua-palatal fricative consonant 'r', and containing a double labio-dental fricative consonant 'f', respectively.

speech listening, in correspondence to tongue movement representations, was assessed. The lowest threshold point on the skull was selected according to the map assessed previously for tongue muscles. The coil was kept in a stable position by means of an articulated arm held by a tripod and participants were requested to keep a stable, relaxed position during the whole experiment.

During acoustic stimuli presentation, TMS was automatically delivered by the computer 100 ms after the beginning of the double consonant part of the word or 100 ms after the change in frequency of bitonal sounds. MEPs recorded in different conditions were band-pass filtered (20–1000 Hz), digitized (sampling frequency 2000 Hz) and stored on a computer. After rectification, the area underlying MEPs was calculated for each trial and was used for successive statistical analysis. The pre-TMS electromyographic activity, recorded from 100 ms before TMS, was also acquired in all trials. The analysis of pre-TMS rectified data showed no difference among conditions.

Results

Figure 1 shows the MEPs of one representative subject recorded during the five stimulus conditions. Note that the words and pseudo-words containing a double 'r' (panels 'a' and 'b', respectively) showed on average an increase in MEPs amplitude with respect to all other conditions (panels 'c'–'e').

To assess statistically the effect of the various stimulus conditions on MEPs amplitude, a one-way analysis of variance (ANOVA) was performed on the MEPs' areas after intrasubject normalization (z -scores) (see Fig. 2). The considered factor was 'condition' [words(rr), words(ff), pseudo-words(rr), pseudo-words(ff), bitonal sounds]. The main effect was significant ($F_{(4,32)} = 11.32$, $P < 0.001$). *Post hoc* analysis (Newman–Keuls test, level of significance $P < 0.05$) revealed that words(rr) was significantly larger than all other levels. Both, words(ff) and pseudo-words(rr) were significantly larger than pseudo-words(ff). Bitonal sounds were not statistically different from both words(ff) and pseudo-words(ff) but were significantly smaller than both words(rr) and pseudo-words(rr). In summary, listening to

words containing a double 'r' consonant determines an increase of tongue MEPs significantly larger than that determined by listening to all other presented stimuli. Tongue MEPs recorded during listening of bitonal sounds are not different from those recorded during listening to words and pseudo-words containing a double 'f' consonant.

To further investigate whether the effect due to the presence of meaning in the presented verbal stimuli interacts with the effect due to the presence of consonants requiring strong tongue tip mobilization, data were submitted to a two-way ANOVA. The considered factors were 'meaning' (words, pseudo-words) and 'tongue involvement' [strong (rr), moderate (ff)]. Both factors were significant ('meaning', $F_{(1,8)} = 10.3$, $P < 0.05$; 'tongue involvement', $F_{(1,8)} = 99.1$, $P < 0.001$) without interaction between them. This result demonstrates that two independent effects are present: (i) when the stimuli contain a double 'r' there is a strong facilitation of tongue muscles independently of whether stimulus is meaningful or not; (ii) the meaningfulness of the stimuli induces a facilitation of tongue MEPs that is independent of stimuli phonological characteristics.

Discussion

The present experiment demonstrates that the passive listening to words that involve tongue mobilization induces an automatic facilitation of the listener's motor cortex. This facilitation, that is under threshold for overt movement generation, is revealed by TMS of tongue cortical motor representation. Furthermore, the effect is stronger in the case of words than in the case of pseudo-words suggesting a possible unspecific facilitation of the motor speech centre due to recognition that the presented material belongs to an extant word. The precise localization of the facilitation site cannot be established with the technique used in the present study. Nevertheless, the presence of the meaningfulness-related effect, together with recent data obtained with a paired-pulse TMS study during action observation (Strafella & Paus, 2000), strongly suggests, by analogy, a cortical origin of the motor facilitation here described. Furthermore, even if the MEPs increase were due to speech listening-induced, subcortical facilitation (Baldissera *et al.*, 2001), it is very unlikely that this facilitation could be due to a noncortical input only, given the linguistic properties of the presented stimuli.

What is the physiological meaning of the present results? The present data show that when an individual listens to verbal stimuli there is an activation of the speech related motor centres. Most interestingly, they show also that this activation is highly specific. Phonemes that require in production a strong activation of tongue muscles, automatically produce, when heard, an activation of the listener's motor centres controlling tongue muscles. This acoustic/motor 'resonance' mechanism strictly resembles the observation/execution matching system involved in action recognition (mirror system). As in the case of the mirror system, this mechanism could be involved in phonetically 'understanding' others' speech: phonemes are recognized because both the speaker and the listener share the same articulatory motor repertoire.

The results also show that the activation is stronger in the case of words than in the case of pseudo-words. This effect could be due to a further facilitation of the speech motor centres induced by the presence of meaning, possibly inferred on the basis of the first syllable. However, one cannot exclude that this facilitation could be due to the familiarity of the presented words in respect to the pseudo-words. It is important to note that in both cases, the speech motor centres are influenced by the semantics of the presented stimuli.

The presence of phoneme specific resonance of the motor speech centres may subservise speech perception as suggested by Liberman (Liberman & Mattingly, 1985; Liberman & Whalen 2000), or it may be a mechanism involved in imitation of vocal gestures. The present data do not allow one to differentiate between these two hypotheses. They show, however, that this resonance occurs automatically and in conditions in which there is no reason whatsoever to repeat the perceived stimuli. It is hard to understand what could be the function of this 'echolalic' mechanism during normal verbal communication. Thus, although the present data by no means prove the motor theory of speech perception, yet they appear to be best explained by this theory.

Acknowledgements

We wish to thank *ad memoriam* A. Liberman for the support given to this work in its preliminary phases. We acknowledge A. Bello, M. Gangitano and G. Melli for their collaboration in data acquisition and G. Cossu and G. Spidalieri for their helpful comments on the manuscript. This research was supported by Human Frontier Science Program and Italian University Ministry grants.

Abbreviations

MEP, motor-evoked potential; OP, opponens pollicis; TMS, transcranial magnetic stimulation.

References

- Baldissera, F., Cavallari, P., Craighero, L. & Fadiga, L. (2001) Modulation of spinal excitability during observation of hand actions in humans. *Eur J. Neurosci.*, **13**, 190–194.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G. & Freund, H.J. (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.*, **13**, 400–404.
- Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1995) Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.*, **73**, 2608–2611.
- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain*, **119**, 593–609.
- 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**, 103–111.
- Grèzes, J., Costes, N. & Decety, J. (1998) Top-down effect of strategy on the perception of human biological motion: a PET investigation. *Cogn. Neuropsychol.*, **15**, 553–582.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C. & Rizzolatti, G. (1999) Cortical mechanisms of human imitation. *Science*, **286**, 2526–2528.
- Jasper, H.H. (1958) Ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, **10**, 371–373.
- Klatt, D.H. (1979) Speech perception: a model of acoustic-phonetic analysis and lexical access. *J. Phonetics*, **7**, 279–285.
- Kuhl, P.K. & Miller, J.D. (1975) Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, **190**, 69–72.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev.*, **74**, 431–461.
- 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 Cogn. Neurosci.*, **4**, 187–196.
- Massaro, D.W. & Cohen, M.M. (1983) Integration of visual and auditory information in speech perception. *J. Exp. Psychol. Hum. Percept. Perform.*, **9**, 753–771.
- Massaro, D.W. & Cohen, M.M. (1990) Perception of synthesized audible and visible speech. *Psychol. Sci.*, **1**, 55–63.
- Nishitani, N. & Hari, R. (2000) Temporal dynamics of cortical representation for action. *Proc. Natl Acad. Sci. USA*, **97**, 913–918.
- Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996b) Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.*, **3**, 131–141.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D. & Fazio, F. (1996a) Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp. Brain Res.*, **111**, 246–252.
- Strafella, A.P. & Paus, T. (2000) Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport*, **11**, 2289–2292.
- Sussman, H. (1989) Neural coding of relational invariance in speech: human language analogs to the barn owl. *Psychol. Rev.*, **96**, 631–642.