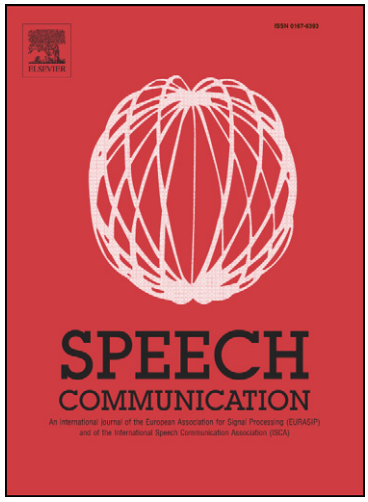
On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception

Marc Sato, Giovanni Buccino, Maurizio Gentilucci, Luigi Cattaneo



On the tip of the tongue: modulation of the primary motor cortex during audio-visual speech perception

Marc Sato, Giovanni Buccino, Maurizio Gentilucci, Luigi Cattaneo

PII: S0167-6393(09)00179-4
DOI: 10.1016/j.specom.2009.12.004
Reference: SPECOM 1851

To appear in: *Speech Communication*

Received Date: 9 April 2009
Revised Date: 2 December 2009
Accepted Date: 3 December 2009

Please cite this article as: Sato, M., Buccino, G., Gentilucci, M., Cattaneo, L., On the tip of the tongue: modulation of the primary motor cortex during audiovisual speech perception, *Speech Communication* (2009), doi: 10.1016/j.specom.2009.12.004

ON THE TIP OF THE TONGUE: MODULATION OF THE PRIMARY MOTOR CORTEX DURING AUDIOVISUAL SPEECH PERCEPTION

Marc Sato[1,CA], Giovanni Buccino[2], Maurizio Gentilucci[2], Luigi Cattaneo[2]

[1] GIPSA-LAB, Département Parole & Cognition, UMR 5216, CNRS & Grenoble Universités, France

[2] Departimento di Neuroscienze, Sezione di Fisiologia, Università di Parma, Italy

CA Corresponding author.

GIPSA-LAB, UMR CNRS 5216, Grenoble Universités

1180 avenue centrale, BP 25, 38040 Grenoble Cedex 9, France.

Email: marc.sato@gipsa-lab.inpg.fr

ABSTRACT

Recent neurophysiological studies show that cortical brain regions involved in the planning and execution of speech gestures are also activated in processing speech sounds. These findings suggest that speech perception is in part mediated by reference to the motor actions afforded in the speech signal. Since interactions between auditory and visual modalities are beneficial in speech perception and face-to-face communication, we used single-pulse transcranial magnetic stimulation (TMS) to investigate whether audiovisual speech perception might induce excitability changes in the left tongue-related primary motor cortex and whether acoustic and visual speech inputs might differentially modulate motor excitability. To this aim, motor-evoked potentials obtained with focal TMS applied over the left tongue primary motor cortex were recorded from participants' tongue muscles during the perception of matching and conflicting audiovisual syllables incorporating tongue- and/or lip-related phonemes (i.e., visual and acoustic /ba/, /ga/ and /da/, visual /ba/ and acoustic /ga/, visual /ga/ and acoustic /ba/). Compared to the presentation of congruent /ba/ syllable, which primarily involves lip movements when pronounced, exposure to syllables incorporating visual and/or acoustic tongue-related phonemes induced a greater excitability of the left tongue primary motor cortex as early as 100-200 ms after the consonantal onset of the acoustically presented syllable. These results provide evidence that both visual and auditory modalities specifically modulate activity in the tongue primary motor cortex at an early stage during audiovisual speech perception. Because no interaction between the two modalities was observed, these results suggest that information from each sensory channel is recoded separately in the primary motor cortex at that point of time. These findings are discussed in relation to theories assuming a link between perception and action in the human speech processing system and theoretical models of audiovisual interaction.

Key Words: Audiovisual speech perception, Transcranial magnetic stimulation, Motor system, Mirror-neuron system, Motor theory of speech perception, McGurk effect.

INTRODUCTION

Although humans are proficient to extract phonetic features from the acoustic signal alone and, to a less extent, capable to partly follow speech gestures when audition is lacking, interactions between auditory and visual modalities are beneficial in speech perception. For instance, seeing speaker's mouth movements enhances speech intelligibility in noisy conditions (Sumby and Pollack, 1954; MacLeod and Summerfield, 1987; Benoît, Mohamadi and Kandel, 1994). Even without any noise, speech-reading may improve language comprehension (Reisberg, McLean and Goldfield, 1987) and benefits hearing-impaired listeners (Grant, Walden and Seitz, 1998). On the other hand, seeing incongruent articulatory gestures may modify auditory speech perception. This refers to the well-known McGurk effect (McGurk and McDonald, 1976) in which a visual /ga/ dubbed with an acoustic /ba/ is sometimes perceived as /da/ or /tha/, while a visual /ba/ dubbed with an acoustic /ga/ is sometimes be perceived as /bga/ (for a review of experimental replications or refinements of the McGurk effect, see Green, 1998). Because the visual input may change the perceiver's auditory experience, these results provide strong evidence for audiovisual integration in speech processing.

Complementing these behavioral findings, functional magnetic resonance imaging (fMRI) studies show that multiple subcortical structures and cortical regions play a key role in audiovisual integration of speech (for a review, see Möttönen, 2004; Ojanen, 2005). Notably, activity within sensory-specific visual and auditory regions (the visual motion-sensitive cortex, V5/MT, and the Heschl's gyrus at the junction of primary and secondary auditory cortices) as well as within multisensory regions (the posterior parts of the left superior temporal gyrus/sulcus, pSTS/STG) are modulated during audiovisual speech perception, when compared to auditory and visual unimodal conditions (Calvert, Campbell and Brammer, 2000; Callan et al., 2003, 2004; Sekiyama et al., 2003; Skipper, Nusbaum and Small, 2005). Because supra- and sub-additive responses to congruent or incongruent stimuli has been observed in pSTS/STG, it has been proposed that acoustic and visual speech inputs are first integrated in these high-level multisensory integrative regions and that modulations of activity within the sensory-specific cortices would then be caused by feedback projections from these multisensory regions and would represent the physiological correlates of the perceptual changes experienced after multisensory integration (Calvert, Campbell and

Brammer, 2000). Alternatively, several magneto-encephalographic (MEG) and electro-encephalographic studies (EEG) of audiovisual speech perception challenge this hypothesis by demonstrating that visual speech input modulates activity in the primary and secondary auditory cortices at an early stage in the cortical speech processing hierarchy (Sams et al., 1991; Möttönen et al., 2002, 2004; Klucharev et al., 2003; see also Besle et al., 2004; van Wassenhove et al., 2005). From these results, both early non-phonetic activation of auditory areas depending upon visual motion cues and a later speech-specific left-lateralized response mediated by backward-projections from multisensory areas have also been suggested (Klucharev et al., 2003; Hertrich et al., 2007). There is either no current agreement between theoretical models of audiovisual speech perception about the processing level at which the acoustic and visual speech signals are integrated (for a review, see Schwartz, Robert-Ribes and Escudier, 1998). While early integration models assume that the visual signal is combined with the auditory signal and processed as single input early in the speech processing hierarchy (Green, 1998), late integration models argue that visual and acoustic inputs are matched separately against unimodal phonetic prototypes, the integration occurring after modality-specific processing (Massaro, 1998).

Apart from unisensory/multisensory regions, recent neurophysiological studies suggest that audiovisual integration of speech might be in part mediated by the speech motor system. In fact, brain areas involved in the planning and execution of speech gestures (i.e., the posterior part of the left inferior frontal gyrus, the premotor and primary motor cortices) are activated during both auditory, visual and audiovisual speech perception (e.g., Nishitani and Hari, 2002; Calvert and Campbell, 2003; Paulesu et al., 2003; Wilson et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Skipper, Nusbaum and Small, 2005; Pulvermuller et al., 2006; Wilson and Iacoboni, 2006). In addition, recent repetitive or double-pulse TMS studies demonstrate that stimulating the left premotor cortex or the orofacial primary motor cortex might impair auditory syllable identification (Meister et al., 2007; d'Aussilio et al., 2009; Sato, Tremblay and Gracco, in press). These results appear in keeping with the long-standing proposal that speech perception and speech production are closely linked processes, as first detailed in the motor theory of speech perception (Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000; for reviews, see Galantucci, Fowler and Turvey, 2006; Schwartz, Sato and Fadiga, 2008), and with more

recent neurophysiological perspectives based on the existence of a mirror-neuron system in humans (Rizzolatti and Arbib, 1998; Rizzolatti and Craighero, 2004; Arbib, 2005; Gentilucci and Corballis, 2006). Further studies also suggest that audiovisual integration of speech might in part be mediated by reference to the motor actions afforded in the speech signals. Indeed, increased activity within speech motor regions has been observed during audiovisual speech perception, compared to auditory and visual unimodal conditions (Skipper et al., 2005, 2007), as well as during audiovisual speech perception under adverse listening or viewing conditions (Callan et al., 2003, 2004). Furthermore, increased activity and sub-additive responses in Broca's area have also been reported during the perception of incongruent auditory-visual speech stimuli, compared to congruent audiovisual or unimodal conditions (Calvert, Campbell and Brammer, 2000; Jones and Callan, 2003; Sekiyama et al., 2003; Ojanen et al., 2005; Pekkola et al., 2005). From these results, speech motor regions appear as good candidates for brain areas where acoustic and visual speech signals can interact.

In keeping with these later findings, the present study was designed to further test a possible involvement of the tongue primary motor cortex in audiovisual integration of speech using single-pulse TMS. When applied to the primary orofacial motor cortex, this technique can be used to monitor changes in the excitability of the cortical motor representations of tongue or lip muscles with a relatively high temporal resolution (around 10 ms). This can be done by measuring lip or tongue motor-evoked potentials (MEPs) elicited by TMS from the corticobulbar pathway under various experimental conditions (Fadiga, Craighero and Etienne, 2005). Previous single-pulse TMS studies show that lip or tongue MEPs are enhanced during passive speech listening or viewing, when stimulating the corresponding area of the left primary motor cortex (Sundara, Namasivayam and Chen, 2001; Fadiga et al., 2002; Watkins, Strafella and Paus, 2003; Watkins and Paus, 2004; Roy et al., 2008). Furthermore, this speech motor 'resonance' mechanism (i.e., the automatic activation of the cortical centres involved in speech production during speech listening; see Fadiga et al., 2002) appears at an early stage, around 100ms after the stimulus onset, and is likely to be articulator specific, motor facilitation being stronger when the recorded muscle activity and the auditory speech stimulus reflect the same articulator (Fadiga et al., 2002; Roy et al., 2008). Importantly, a previous TMS study (Sundara, Namasivayam and Chen, 2001) did not find any

increase of MEP responses recorded from the lip muscles during auditory and incongruent audiovisual speech perception, but during visual and congruent audiovisual perception of speech. However, this result appears at odds with some above-mentioned studies demonstrating that passive auditory speech perception induces cortical activity in the speech motor system and its stronger involvement during the perception of incongruent auditory-visual stimuli, compared to congruent ones.

Because methodological aspects of the Sundara and colleagues' study (2001) might have impact their results, notably the fact that the TMS pulses were not time-locked to the stimuli, we further investigate whether audiovisual speech perception might induce excitability changes in the left tongue primary motor cortex and whether acoustic and visual speech inputs might differentially modulate motor excitability early in the speech processing hierarchy. To this aim, native Italian participants were presented to temporally matched audiovisual syllables, consisting of congruent (i.e., /ba/, /ga/, /da/) and incongruent (i.e., visual /ba/ and acoustic /ga/, and vice-versa) audiovisual syllables, while MEPs obtained with focal single-pulse TMS were recorded from the tongue muscles. Magnetic stimuli were delivered either 100 ms or 200 ms after the onset from the time corresponding to the consonant release of the acoustically presented syllable. For each time-pulse, the mean size of the MEPs for the sound-vision pair /ba/-/ba/, /ba/-/ga/, /ga/-/ga/ and /ga/-/ba/ stimuli was expressed as a percentage of the mean size of the MEPs obtained for the congruent audiovisual /da/ stimulus (which primarily involves tongue movements when pronounced and here served as a control condition). This allows to test whether both auditory and visual features of the syllables, as well as their interaction, might induce greater motor excitability in the left tongue-related primary motor cortex. In addition, a behavioral experiment was run after the TMS experiment in order to evaluate participants' syllable identification of the stimuli.

According to the above-mentioned studies, we hypothesized that exposure to syllables incorporating visual and/or acoustic tongue-related phonemes (i.e., sound-vision pair /ga/-/ga/, /ga/-/ba/, /ba/-/ga/) should induce a greater excitability of the left tongue primary motor cortex compared to the presentation of congruent /ba/ syllable (which primarily involves lip movements when pronounced) but not compared to the congruent audiovisual /da/ stimulus. Another goal of this study was to test a possible interaction between the auditory and visual modalities in order to determine if motor information from

each sensory channel are recoded separately, or rather integrated, at an early stage in the primary motor cortex. Finally, we also tested whether the excitability degree of the tongue motor cortex might correlate with participants' syllable identification of the incongruent audiovisual stimuli. Given that the incidence of the McGurk effect has been shown to vary across individuals (Brancazio and Miller, 2005), a correlation between the excitability degree of the tongue motor cortex and the corresponding perceptual identification of the incongruent audiovisual stimuli (i.e., a greater excitability of the tongue motor excitability for participants who mainly reported syllables incorporating tongue-related phonemes) would likely suggest a mediating role of the motor system in audiovisual speech perception.

METHOD

Participants

Ten healthy adults (five males; mean age ± SD: 28 ± 4 years), native Italian speakers, participated in the study. All were right-handed, according to a standard handedness inventory (Oldfield, 1971), had normal or corrected-to-normal vision and reported no history of speaking or hearing disorders. Participants were screened for neurological, psychiatric, and other medical problems, and contraindications to TMS (Wasserman, 1998). Informed consent was obtained for all participants and they were paid for their participation. The protocol was approved by the Parma University Ethical Committee and was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Electromyography

Continuous electromyography (EMG) recordings from the tongue muscles were acquired with a 1902 CED amplifier (Cambridge Electronic Design, Cambridge, U.K.) at 1000 x amplification and digitized with a CED Micro 1401 analog-to-digital converting unit (Cambridge Electronic Design, Cambridge, U.K.) at a sampling rate of 8 kHz. The signal was band-pass filtered (60-4000 Hz) and stored on a computer for offline analysis. The tongue muscles were recorded with a bipolar montage, with a pair of Ag-AgCl surface electrodes (diameter 3mm) mounted on a 1 cm x 1 cm plastic plate and fixed on an amagnetic metal clip device. Though it is well-known that the cortical representation of the tongue muscles is mainly bilateral (Muellbacher, Mathis and Hess, 1994; Urban et al., 1996), it is a matter of debate whether unilateral tongue motor responses can be safely recorded devoid of volume-conducted potentials from the contralateral side (Muellbacher and Mamoli, 1997; Chen, Wu and Chu, 1999). For this reason, we decided to record the tongue on the midline, instead of a unilateral recording. Accordingly, the active and reference electrodes were placed on the dorsal surface and the ventral aspect of the tongue, respectively, approximately 2 cm caudal to the tongue apex. In a pilot test, visual monitoring of the EMG signal confirmed the involvement of the tongue muscles as recorded with our electrode montage during the production of /da/ and /ga/ (incorporating a bilabial and an apico-dental consonant, respectively), while the production of /ba/ involved only slight tongue mobilization (incorporating a velar consonant).

Transcranial Magnetic Stimulation

The left motor cortex was magnetically stimulated by means of monophasic single pulse TMS delivered through a figure-of-eight coil (ESAOTE, Biomedica, Italy). The coil was moved over the scalp in order to determine the optimal site from which maximal amplitude MEPs were elicited in the tongue muscles. For optimal activation of the cortico-hypoglossal projections, the intersection of the coil was placed tangentially to the scalp with the handle pointing backward parasagitally (Muelbacher, Mathis and Hess, 1994; Rodel, Laskawi and Markus, 2003). The tongue motor cortex was located by first identifying the hand area (where application of TMS elicited a visible twitch in the contralateral hand) and then moving the coil ventrally and slightly anteriorly until maximal amplitude MEPs were recorded in the tongue muscles. The resting motor threshold of the tongue muscles was determined according to standard methods as the minimal intensity capable of evoking MEPs in 5 out of 10 trials of the relaxed tongue muscles with amplitude of at least 50 μV (Rossini *et al.*, 1994). The intensity of the stimulator was then set to 120% of the resting motor threshold during the experimental session.

Stimuli

Multiple utterances of /ba/, /da/ and /ga/ syllables were individually recorded by a female actor, native Italian speaker, using a digital video camera. The actress pronounced each syllable naturally, maintaining an even intonation, tempo and vocal intensity while producing the speech sequence, and kept their lips closed between each utterance. Video digitizing (the speaker's full face was presented against a grey background) was done at 25 frames/s in 720 × 576 dots. Audio digitizing was done at 44.1 kHz in 16 bits. One clearly articulated token of each syllable was selected in order to achieve two goals. The first goal was for the three syllables to be temporally aligned at their consonantal onset (corresponding to the bilabial /b/, apico-dental /d/ and velar /g/ releases after the prevoicing period; mean value ± SD: 528 ms ± 3) and matched for acoustic duration and intensity. The second goal was that the speaker initiated and finished each utterance from a neutral closed-mouth position. With this editing procedure the consonant release occurred for both speech sequences in the 14th frame of the movie, each movie being 25 frames long (1 s). Five distinct stimuli were then created, consisting of three congruent audiovisual clips (corresponding to /ba/, /da/ and /ga/ syllables) and two incongruent

audiovisual clips (corresponding to a /ba/ video track dubbed with a /ga/ audio track and vice-versa).

Procedure

The experiment was programmed using Matlab (The Mathworks Inc., Natick, MA), Cogent (Functional Imaging Laboratory, Queen Square, London) and Signal (Cambridge Electronic Design, Cambridge, U.K.) software to control the stimulus presentation and to trigger the TMS and EMG recordings. The video track was delivered on a 19 inch computer monitor with a viewing distance of approximately 60 cm and the audio track was presented at a comfortable sound level through two loudspeakers placed on either side of the monitor.

TMS experimental session

During the TMS experimental session, participants were seated on an armchair, their head lying on a headrest in order to maintain a comfortable and stable position. They were instructed to carefully listen-to and watch audiovisual clips of a speaker. In addition, they were also asked to maintain their tongue in a relaxed position during the whole session. Each trial started with a fixation cue (the '+' symbol) presented during 500 ms, immediately followed by the stimulus for 1000 ms and then by a blank screen for 4000 ms. TMS pulses were automatically delivered either 100 ms or 200 ms from the time corresponding to the consonantal onset of the acoustically presented syllable (see Figure 1). These delays were extrapolated on the basis of a previous TMS study showing an increase of excitability of the left tongue motor cortex during speech listening as early as 100 ms after the onset of the stimuli (Fadiga et al., 2002).

Insert Figure 1 about here

For each movie (i.e., congruent audiovisual /ba/, /da/, /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli) and each time-pulse (100 and 200 ms), every stimulus was presented 10 times in a pseudo-randomized sequence for a total of 100 trials. In order to avoid habituation, 20 additional trials consisting of a 1s movie combining visual noise and white noise were also presented. These trials were not included in the analyses.

Behavioral experimental session

In order to test participants' performance identification of the stimuli, a behavioral experimental session was run after the TMS experimental session (the sessions were run separately to minimize/avoid

motor activity due to key response and/or possible covert speech in the syllable decision task; see Fadiga et al., 2002). The procedure and apparatus were the same as described previously. Participants were instructed to carefully listen-to and watch audiovisual stimuli of a speaker and were asked to write what they hear. Every stimulus (i.e., congruent audiovisual /ba/, /da/, /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli) was presented 10 times in a pseudo-randomized sequence for a total of 50 trials. Each trial started with a fixation cue presented during 500 ms, immediately followed by the stimulus for 1000 ms and then by a question mark (the '?' symbol) for 4000 ms. The question mark was the signal to write the response.

Data Analysis

TMS experimental session

For each participants and each trial, the EMG trace was rectified and the area under the curve corresponding to the MEP and to the baseline EMG activity (100 ms) preceding the pulse were calculated. Even though participants were asked to maintain their tongue in a relaxed position during the session, visual inspection of the baseline EMG activity showed slight activation of the tongue muscles. Because MEP size is known to be related to the amount of baseline EMG activity, an analysis of covariance was used to adjust the MEP size for the corresponding baseline EMG activity in each trial for each participant (Watkins, Strafella and Paus, 2003; Watkins and Paus, 2004). For each participant and each time-pulse, the mean size of the adjusted MEPs for the sound-vision pair /ba/-/ba/, /ba/-/ga/, /ga/-/ga/ and /ga/-/ba/ stimuli was then expressed as a percentage of the mean size of the adjusted MEPs obtained for the congruent audiovisual /da/ stimulus which served as a control condition (see Figure 2). A three-way analysis of variance (ANOVA) was performed on these MEP ratios in order to test each sensory channel as well as their possible interaction. The considered within-subjects factors were related to the auditory (/ba/, /ga/) and visual (/ba/,/ga/) inputs and to the time-pulse (100 ms, 200 ms). Importantly, a three-way ANOVA performed with the same factors on the baseline EMG activity ratios showed neither significant effects nor interactions (all F's< 1.22 and p's> .29).

Insert Figure 2 about here

Behavioral experimental session

Responses were averaged across participant and conditions and grouped in four categories (i.e., /ba/, /da/, /ga/ and others responses – see Figure 3). Because of technical problems, two participants did not perform the behavioral experimental session. A one-way ANOVA was performed on the percentage of reported responses based on the acoustically presented syllable. The considered factor was related to the stimulus presentation (i.e., congruent audiovisual /ba/, /da/, /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli).

Insert Figure 3 about here

For all the analyses, the significance level was set at $p = .05$ and Greenhouse-Geisser corrected when appropriate. When required, post-hoc analyses were conducted with Newman-Keuls tests.

RESULTS

TMS experimental session

The results of the ANOVA showed a significant main effect of both the acoustic channel ($F_{(1,9)}$ = 4.69, p = .05) and the visual channel ($F_{(1,9)}$ = 10.92, p = .01). For both the acoustic and visual channels, the MEP ratios were largest for /ga/ than for /ba/ syllables (on average, 103% vs. 94% and 105% vs. 93%, respectively). No reliable effect of the time-pulse ($F_{(1,9)}$ = 0.06) nor interactions between the acoustic and the visual channels ($F_{(1,9)}$ = 0.07), between the acoustic channel and the time-pulse ($F_{(1,9)}$ = 0.22), between the visual channel and the time-pulse ($F_{(1,9)}$ = 1.03) and between the three factors ($F_{(1,9)}$ = 2.45) were observed.

In order to test whether the MEPs ratios (averaged over the two two time-pulses) observed for congruent audiovisual /ba/ and /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli differed significantly from 100% (i.e., the congruent audiovisual /da/ stimulus which served as a control condition), student's paired t-tests with a Bonferroni correction were performed. The MEPs ratio related to congruent audiovisual /ba/ stimulus differed significantly from 100% (t(9) = -3.19, p = .04) but not those related to the congruent audiovisual /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli (t(9) = -1.53, t(9) = -0.10, t(9) = -0.40, respectively).

Behavioral Experiment

The congruent audiovisual stimuli were almost perfectly recognized, with on average 98% of correct responses. For the sound-vision pair /ba/-/ga/ stimulus, the participants' responses fell into three distinct categories: auditory-based responses (i.e., /ba/ - on average: 61%), 'fusion' responses (i.e., /da/ - on average 26%) and others responses (i.e., /a/ - on average 13%). For the sound-vision pair /ga/-/ba/ stimulus, the participants' responses fell into two distinct categories: auditory-based responses (i.e., /ga/ - on average: 86%) and 'combination' responses (i.e., /bga/ - on average 13% – see Figure 3). The results of the ANOVA showed a significant effect of the stimulus presentation ($F_{(4,28)}$ = 3.19, p = .02). Post-hoc analysis revealed significantly lower auditory-based response scores during the presentation of the sound-vision pair /ba/-/ga/ stimulus than during the presentation of the congruent audiovisual /ba/, /ga/, /da/ and the sound-vision pair /ga/-/ba/ stimuli (p = .03, p = .04, p = .02, p = .05, respectively). No other

significant differences were observed (all *p*'s > .49).

Correlation between MEP ratios and perceptual scores

In order to evaluate a possible relationship between the amount of motor excitability and the perceptual scores obtained during the presentation of incongruent sound-vision pair stimuli, two Pearson tests were performed on the MEP ratios (averaged on the two time-pulses) and the percentage of acoustically-based responses of the participants (i.e., /ba/ and /ga/ for the sound-vision pair /ba/-/ga/ and /ga/-/ba/, respectively). No significant correlation between the two measures was found for the sound-vision pair /ba/-/ga/ (r = 0.24, p = .29) and /ga/-/ba/ (r = -0.18, p = .34) stimuli.

DISCUSSION

Compared to the presentation of congruent audiovisual /ba/ syllable, which primarily involves lip movements when pronounced, exposure to syllables incorporating visual and/or acoustic tongue-related phonemes induced a greater excitability of the left tongue primary motor cortex as early as 100 ms after the stimulus onset. Importantly, the possibility that this modulation reflects overt motor activity is quite unlikely. Indeed, the analysis of the baseline EMG activity failed to show any significant effect and a covariance analysis was used in order to adjust the MEP size for the corresponding baseline EMG activity. Rather, the present results likely suggest that information from both the visual and auditory modalities specifically modulate the excitability of the tongue primary motor cortex at an early stage during audiovisual speech perception.

Although no unimodal conditions were tested in the present study, these results are in keeping with previous fMRI and single-pulse TMS studies showing that cortical speech motor regions are activated during auditory, visual and audiovisual speech perception (e.g., Fadiga et al., 2002; Calvert and Campbell, 2003; Paulesu et al., 2003; Watkins, Strafella and Paus, 2003; Watkins and Paus, 2004; Wilson et al., 2004; Ojanen et al., 2005; Pekkola et al., 2005; Skipper, Nusbaum and Small, 2005; Pulvermuller et al., 2006; Wilson and Iacoboni, 2006; Roy et al., 2008). Both the time-course and articulatory specific modulation of the observed tongue MEPs are also in line with the results from two recent single-pulse TMS studies (Fadiga et al., 2002; Roy et al., 2008), in which an early facilitation of the tongue motor cortex was observed during listening to bisyllables including consonants which require strong tongue mobilization when pronounced (i.e., /r/, /l/) as compared to bisyllables including consonants which require slight tongue mobilization when pronounced (i.e., /f/, /p/, /b/, /m/). The specificity of this speech motor 'resonance' mechanism (Fadiga et al., 2002) is also suggested by two recent fMRI studies showing similar somatotopic patterns of motor activity in 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). However, it is worthwhile noting that a previous single-pulse TMS study (Sundara, Namasivayam and Chen, 2001) did not find any increase of MEP responses recorded from the lip muscles during auditory and incongruent audiovisual speech perception, but during visual and congruent

audiovisual perception of speech. Methodological aspect of the Sundara et al.'s study (2001), most notably the fact that the TMS pulses "were not time-locked to the visual or auditory stimuli" during the experiment (Sundara, Namasivayam and Chen, 2001; pp. 1342), may explain this negative finding.

Although the present results do not imply that perceiving speech is solely mediated by an articulatory code, as claimed in the Motor Theory of Speech Perception (Liberman et al., 1967; Liberman and Mattingly 1985; Liberman and Whalen, 2000), they clearly support the idea that speech perception involves a specific mapping from the speaker's articulatory gestures into the viewer's/listener's motor plans. From this view, they appear consistent with recent neurobiological models of speech perception (Callan et al., 2004; Wilson and Iacoboni, 2006; Skipper et al., 2007) which postulate that both sensory and motor brain regions participate in speech perception by means of sensory-to-motor feedforward and motor-to-sensory feedback projections. In these models, multisensory inputs interact with activity from the motor system involved in speech production, the role of which is to constrain phonetic interpretation of the incoming sensory information. For example, Skipper and colleagues (2007) proposed that multisensory speech representations in the left pSTS/STG are mapped onto speech motor control commands localized in Broca's area. These speech motor control commands are subsequently mapped to the motor commands in the ventral premotor and motor cortices that specifically code the actual dynamics of the movement of the required effector(s). In return, these activated motor commands then predict the acoustic and somatosensory consequences of executing a speech movement through reafferent feedback to both the left superior temporal sulcus/gyrus and somatosensory cortices, respectively. These internally generated sensory consequences are then thought to constrain the ultimate phonetic interpretation of the incoming sensory information. Importantly, Skipper and colleagues (2007) postulate that frontal motor areas, together with the left pSTS/STG, participate in audiovisual integration of speech. Indeed, using fMRI, they showed that perception of a conflicting audiovisual syllable (i.e., auditory /pa/ dubbed on visual /ka/) evoked two different patterns of activity in speech motor areas regarding subject's percept (i.e., 'illusory' /ta/ or visually driven /ka/ syllables) which resembles that evoked by the congruent audiovisual syllable that corresponds to participants' perception of the stimulus. This result appears also in line with the above-mentioned fMRI studies showing increased activity and sub-additive responses in

Broca's area during the perception of incongruent auditory-visual speech stimuli, compared to congruent audiovisual or unimodal conditions (Calvert, Campbell and Brammer, 2000; Jones and Callan, 2003; Sekiyama et al., 2003; Ojanen et al., 2005; Pekkola et al., 2005). Note however that fMRI do not provide the temporal resolution necessary to track activity in localized cortical regions in a period of time restricted to basically a few hundred of milliseconds. From these results, it is therefore difficult to determine the precise time when the acoustic and visual speech signals are integrated in the speech processing hierarchy. Conversely, in the present study, no interaction was observed between the auditory and visual modalities when stimulating the tongue primary motor cortex as early as 100-200ms after the stimulus onset. Given that information from both the visual and auditory modalities specifically modulates the excitability of the tongue primary motor cortex, this rather suggests that motor information from each sensory channel is recoded separately, at an early stage, before audiovisual integration of speech.

Compliant with this finding, a greater excitability of the tongue muscles was observed during the perception of the sound-vision /ba/-/ga/ stimulus as compared to the congruent /ba/ syllable, while participants mainly reported /ba/ responses in the incongruent condition during the behavioral identification task. This would suggest that even when the McGurk 'fusion' effect did not occur, motor information was nevertheless recoded from the visual modality. However, given that the TMS and behavioral sessions were performed separately and that participants were instructed to report what they hear in the behavioral session (which may be critical because directing attention to a particular modality is known to affect the size of the McGurk effect; see Massaro, 1998), this result should be taken with caution. Despite these limitations, this result appears nevertheless quite reminiscent of a previous study by Gentilucci and Cattaneo (2005) who asked participants to repeat aloud a perceived utterance after the presentation of congruent and incongruent audiovisual stimuli. For the incongruent audiovisual stimuli, when participant's responses relied on the acoustical input alone (e.g., /ba/), voice spectra and lip kinematics analyses showed that they were nevertheless influenced by the information provided in the visual modality. By including a visual rate manipulation in a McGurk paradigm, Brancazio and Miller (2005) also showed that even when the McGurk effect does not occur, perceivers may have nevertheless used visual information in phonetic processing. From these results, it has been argued that differences in

the categorization of incongruent audiovisual stimuli might reflect differences in how acoustical and visual information are mapped onto phonetic categories (Brancazio and Miller, 2005). In that case, non-McGurk responses arise if later audiovisual integration gives rise to a percept which does not perfectly match one of the phonetic categories specified visually or acoustically, but rather falls between these categories (which could take a variety of forms including an articulatory one, see Schwartz, Robert-Ribes and Escudier, 1998). As a result of a late phonetic-decision stage, auditory-based responses (or visually-based responses) might be reported even though information from both modalities has been processed at an early stage.

In conclusion, although additional experiments are required to further determine when and how activity within the speech motor regions might constrain the phonetic interpretation of the sensory speech inputs, the present results provide evidence that both visual and auditory modalities specifically modulate activity in the tongue primary motor cortex. Given that no interaction between the two modalities was observed, these results also suggest that information from each sensory channel is recoded separately in the primary motor cortex at an early stage during audiovisual speech perception.

REFERENCES

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.

Besle, J., Fort, A., Delpuech, C. & Giard, M.H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20: 2225-34.

Benoît, C., Mohamadi, T. & Kandel, S.D. (1994). Effects on phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37: 1195-1203.

Brancazio, L. & Miller, J.L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception & Psychophysics*, 67(5): 759-769.

Callan, D.E., Jones, J.A., Munhall, K.G., Callan, A.M., Kroos, C. & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*, 14: 2213-2217.

Callan, D.E., Jones, J.A., Munhall, K.G., Callan, A.M., Kroos, C. & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, 16: 805-16.

Calvert, G.A & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1): 57-70.

Calvert, G.A., Campbell, R. & Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11): 649-657.

Chen, C.H., Wu, T. & Chu, N.S. (1999). Bilateral cortical representation of the intrinsic lingual muscles. *Neurology*, 52: 411-413.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C. & Fadiga, L. (2009). The motor somatotopy of speech perception. Current Biology, 19(5): 381-385.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15*: 399-402.

Fadiga, L., Craighero, L. & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current Opinion in Neurobiology*, 15: 213-218.

Galantucci, B., Fowler, C.A. & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3): 361-377.

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

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews,* 30:949-960.

Grant, K., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, 103, 2677–2690.

Green, K.P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In Campbell, R., Dodd, B. & Burnham, D. (Eds.), *Hearing by eye, II. Perspectives and directions in research on audiovisual aspects of language processing*. Psychology Press, Hove (U.K.), pp. 3-25.

Hertrich, I., Mathiak, K., Lutzenberger, W., Menning, H. & Ackermann, H. (2007). Sequential audiovisual interactions during speech perception: a whole-head MEG study. *Neuropsychologia*, 45(6): 1342-1354.

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

Klucharev, V., Möttönen, R & Sams M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Brain Research. Cognitive Brain Research*, 18: 65-75.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 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 in Cognitive Science*, 3(7): 254-264.

MacLeod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21: 131-141.

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(19): 1692-1696.

Massaro, D.W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264: 746-748.

Möttönen, R., Krause, C.M., Tiippana, K. & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res Cognitive Brain Research*, 13: 417-25.

Möttönen, R. (2004). *Cortical Mechanisms of Seeing and Hearing Speech*. Unpublished PhD Thesis, Helsinki University of Technology, Laboratory of Computational Engineering, Espoo, Finland.

Möttönen, R., Schurmann, M. & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters*, 363: 112-5.

Muellbacher, W., Mathis, J. & Hess, C.W. (1994). Electrophysiological assessment of central and peripheral motor routes to the lingual muscles. *J. Neuro.l Neurosurg. Psychiatry*, 57: 309-315.

Muellbacher, W. & Mamoli, B. (1997). The course of cortico-hypoglossal projections in the human brainstem: functional testing using transcranial magnetic stimulation. *Brain*, 120: 1909-1910.

Nishitani, N. & Hari, R. (2002). Viewing lip forms: Cortical dynamics. *Neuron*, 36: 1211-1220.

Ojanen, V. (2005). *Neurocognitive Mechanisms of Audiovisual Speech Perception*. Unpublished PhD Thesis, Helsinki University of Technology, Laboratory of Computational Engineering, Espoo, Finland.

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.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9: 97-114.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, A.A., De Giovanni, U., Sensolo, S. & Fazio, F. (2003). A functional-anatomical model for lipreading. *Journal of Neurophysiology*, 90: 2005-2013.

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.

Pulvermüller, 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. USA*, 103(20):7865-70.

Reisberg, D., McLean, J. & Goldfield, A. (1987). Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In Campbell, R. & Dodd, B. (Eds.), *Hearing by eye: The psychology of lipreading*. Lawrence Erlbaum Associates, London (U.K.), pp. 97-113.

Rizzolatti, G. & Arbib, M.A. (1998). Language within our grasp. *Trends in Neurosciences*, 21: 188-194.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27: 169-192.

Rodel, R.M., Laskawi, R. & Markus, H. (2003). Tongue representation in the lateral cortical motor region of thehuman brain as assessed by transcranialmagnetic stimulation. *Ann. Ot ol. Rhinol. Laryngol.*, 112(1): 71-76.

Rossini, P.M., Barker, A.T., Berardelli, A., Caramia, M.D., Caruso, G., Cracco, R. Q., Dimitrijevic, M. R., Hallett, M., Katayama, Y. & Lucking, C.H. (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. *Electroencephalogr. Clin. Neurophysiol.*, 91(2):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. *J. Physiol. Paris*, 102(1-3): 101-105.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O.V., Lu, S.T.& Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127: 141-145.

Sato, M., Tremblay, P. & Gracco, V.L. (in press). A mediating role of the premotor cortex in phoneme segmentation. Brain and Language.

Schwartz, J.-L., Robert-Ribes, J. & Escudier, P. (1998). Ten years after Summerfield: A taxonomy of

models for audio-visual fusion in speech perception. In Campbell, R., Dodd, B. & Burnham, D. (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* . Hove, U.K.: Psychology Press, pp. 85-108.

Schwartz, J.-L., Sato, M. & Fadiga, L. (2008). The common language of speech perception and action: a neurocognitive perspective. Revue *Française de Linguistique Appliquée*, 13(2): 9-22.

Sekiyama, K., Kanno, I., Miura, S. & Sugita, Y. (2003). Audio-visual speech perception examined by fMRI and PET. *Neuroscience Research*, 47: 277-287.

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.

Sumby, W.H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of Acoustical Society of America*, 26: 212-215.

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

Urban, P.P., Hopf, H.C., Connemann, B., Hundemer, H.P. & Koehler, J. (1996). The course of cortico-hypoglossal projections in the human brainstem. Functional testing using transcranial magnetic stimulation. *Brain*, 119: 1031-1038

van Wassenhove, V., Grant, K.W. & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. USA*, 102: 1181-6.

Wassermann, E.M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: Report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5-7, 1996. *Electroencephalogr. Clin. Neurophysiol.*, 108(1): 1-16.

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

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.

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.

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.

ACKNOWLEDGMENTS

We wish to thank Elena Borra and Hélène Loevenbruck for their help with this study. This research was supported by MIUR (Ministero Italiano dell'Istruzione, dell'Universita e della Ricerca) and CNRS (Centre National de la Recherche Scientifique).

FIGURES

Fig. 1. Stimulus sample. A part of video sequence for /ba/. All the stimuli were temporally aligned at their consonantal onset, occurring in the 14th frame of the movie. TMS pulses were delivered either at 100 ms or 200 ms from the consonant release of the acoustically presented syllable.

Figure 2. Mean MEP sizes observed for the congruent audiovisual /ba/, /ga/ and sound-vision pair /ba/-/ga/ and /ga/-/ba/ stimuli expressed as a percentage of the mean MEP size observed for the congruent audiovisual /da/ stimulus (x-axis through the 100% level). Error bars represent standard errors of the mean.

Figure 3. Perceptual scores reported during the behavioral experimental session. Error bars represent standard errors of the mean.