European Journal of Neuroscience, Vol. 17, pp. 179–184, 2003 

# SHORT COMMUNICATION
# Grasp observation influences speech production

Maurizio Gentilucci
Istituto di Fisiologia Umana, Via Volturno 39, I-43100, Parma, Italy

*Keywords:* human voice, lip kinematics, object size, observation of hand grasp, syllable pronunciation

## Abstract

Subjects pronounced either the syllable 'BA' or 'GA' while observing motor acts of hand grasp directed to objects of two sizes (experiment 1). Kinematics of lip aperture and amplitude spectrum of voice were influenced by the observation of the different grasp kinematics depending on the size of the target objects. Specifically, both lip aperture and voice peak amplitude were greater when the observed hand grasp was directed to the large object. Two control experiments ruled out that the different arm velocity when reaching objects varying in size (experiment 2), and overt visual analysis of the target-object (experiment 3), affected lip movement and voice emission. Results provide behavioural evidence in favour of the hypothesis that the system involved in observation (and preparation) of grasp movements partially shares the cortical areas involved in speech production.

## Introduction

Pronouncing a syllable can be selectively interfered by the simultaneous motor act of hand grasp aimed at target-objects of different size (Gentilucci *et al*., 2001). This finding suggests that, in humans, commands of grasping directed to both hand and mouth, similar to those discovered in monkey (Rizzolatti *et al*., 1988), can interfere with speech production. Moreover, it supports neuroimaging data showing activation of Broca's area when imitating (preparing) hand movements (Iacoboni *et al*., 1999). Broca's area is traditionally considered to be involved in speech production. Thus, behavioural data (Gentilucci *et al*., 2001) support the notion that the structures involved in the preparation of grasp share a common neural substrate with the speech system (Rizzolatti & Arbib, 1998).

In the monkey premotor cortex, where neurons involved in the control of grasp with both hand and mouth were recorded (Rizzolatti *et al*., 1988), a class of neurons ('mirror' neurons) showed the property of discharging when the monkey both observed and executed similar grasp motor acts (Gallese *et al*., 1996). In humans, observation of grasp can activate preparation of the same motor act (Fadiga *et al*., 1994). By matching the effects of the observed motor act with the aim of the same prepared motor act, the 'mirror' system can be involved in representing (and understanding) the hand gesture. On the other hand, Broca's area seems to be involved in encoding phonetic representations in terms of mouth articulation gestures (Demonet *et al*., 1992; Zatorre *et al*., 1992; Paulesu *et al*., 1993). If, in humans, representations of grasp gestures are coded also in speech areas (Buccino *et al*., 2001; Iacoboni *et al*., 1999), reading syllables aloud (and their consequent encoding in mouth articulation representations) may be influenced by the contemporaneous observation of hand grasps. Influence can depend on the different finger kinematics of the agent, which, in turn, is consequent to visual analysis of target-object features. This hypothesis was verified in three experiments. In experiment 1, subjects pronounced a syllable after observing a hand grasp motor act directed to objects of two sizes. In the control experiment 2, subjects pronounced the syllable after observing the movement of a spot of light pointing to the same objects. In the control experiment 3, subjects pronounced the syllable after paying overt attention to the sizes of the objects, which had been the targets of the movements observed in the other two experiments.

## Methods

Twenty-six right-handed (according to Edinburgh Inventory, Oldfield, 1971) subjects (13 females, 13 males, aged 22–27 years) participated in the present study to which they gave informed consent. All of them were naïve as to the purpose of the experiment.

In a dark and soundproof room, subjects sat in front of a black table, placing their forearms on the table plane (Fig. 1). Subjects were required not to move their head and trunk during the experimental session. Visual stimulus was one of two white parallelepipeds with a square base: small object (sides $3 \times 3$ cm, height 1 cm), and large object (sides $5 \times 5$ cm, height 1 cm). On their visible face, either a syllable (syllable configuration) or a configuration of small points randomly scattered in the same area occupied by the syllable (cloud configuration) was printed in black. Syllables could be either BA (/ba/) or GA (/ga/). The points of the cloud configuration were obtained by decomposing the letters XX using a PC software. Letters were 0.5 cm wide and 0.6 cm high. One parallelepiped was placed on the plane of the table along the subject's sagittal plane at a distance of 45 cm from the subject's chest. In all experiments, illumination of the room commanded by a PC was the signal to begin the trial. In the following three experiments, subjects were required to pronounce the syllable when presented with the object.

### *Experiment 1: grasping observation (Fig. 1)*

On the subject's right side, the experimenter placed his right thumb and index finger, held in the pinch position, on a white circle located on the table plane (SP, starting position). SP was 30 cm distant from the visual stimulus and 40 cm distant from the subject's chest. He reached and grasped the target (i.e. the visual stimulus) with the maximal velocity

*Correspondence*: Dr Maurizio Gentilucci, as above.
E-mail: gentiluc@unipr.it

*Received 18 April 2002, revised 27 September 2002, accepted 30 October 2002*

doi:10.1046/j.1460-9568.2003.02438.x

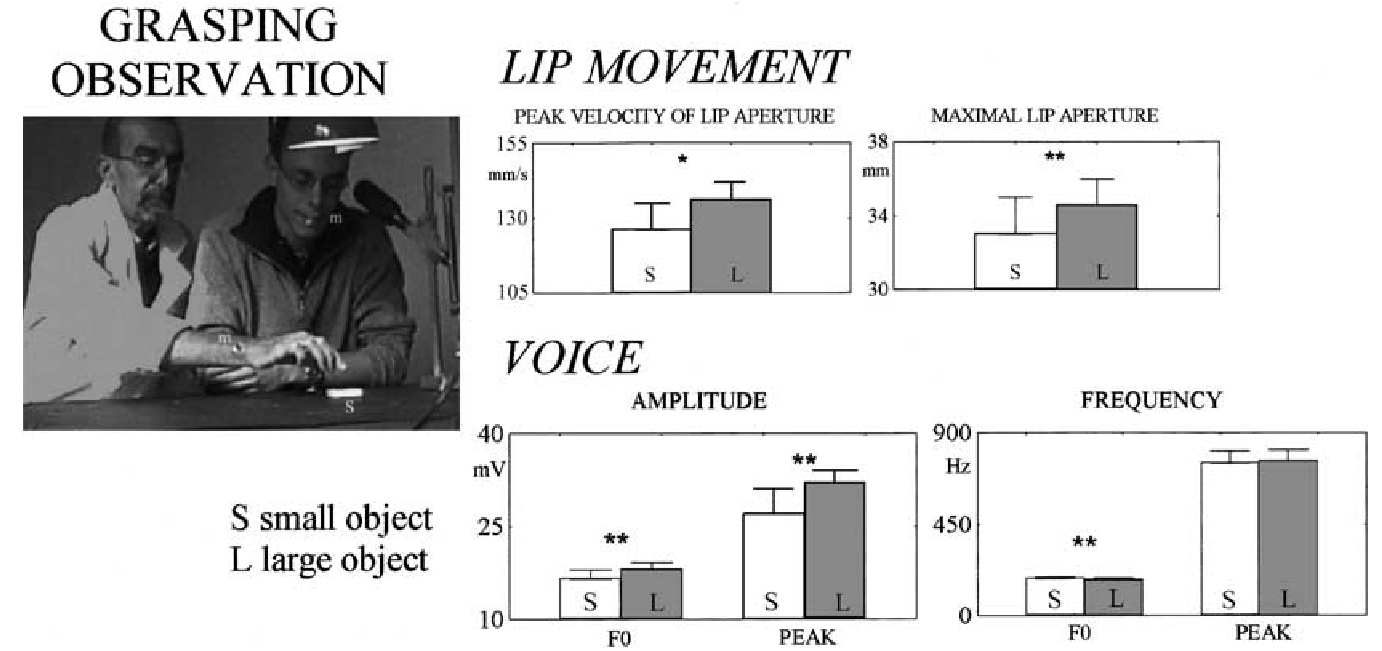


FIG. 1. The experimental apparatus and mean values of parameters of lip kinematics and voice analysed in experiment 1. Left panel: experimental apparatus. Upper right panels: mean values of parameters of subject's lip aperture. Lower right panels: mean values of amplitude and frequency of subject's voice. F0: fundamental frequency. Peak amplitude and frequency correspond to F1. Bar markers are S.E. Asterisks indicate statistical significance.

compatible with the accuracy required by the task. A precision grip was used when grasping the small object, and a finger grip when grasping the large object (Rizzolatti *et al*., 1988). Subjects were required to pronounce the syllable as soon as the experimenter touched the parallelepiped and to use a voice volume as during normal conversation. When the cloud configuration was presented, they were required to remain silent. Four females and 4 males participated in this experiment.

### *Experiment 2: moving-light-spot observation (Fig. 2)*

The experimenter moved a torch producing a circular light spot on the table plane (whose diameter was approximately 5 cm) from SP to the visual stimulus (which was 30 cm from SP like in experiment 1). A manual movement was chosen in order to reproduce as best as possible both the same pattern and accuracy of the arm movements as in experiment 1. The experimenter used a higher velocity when pointing

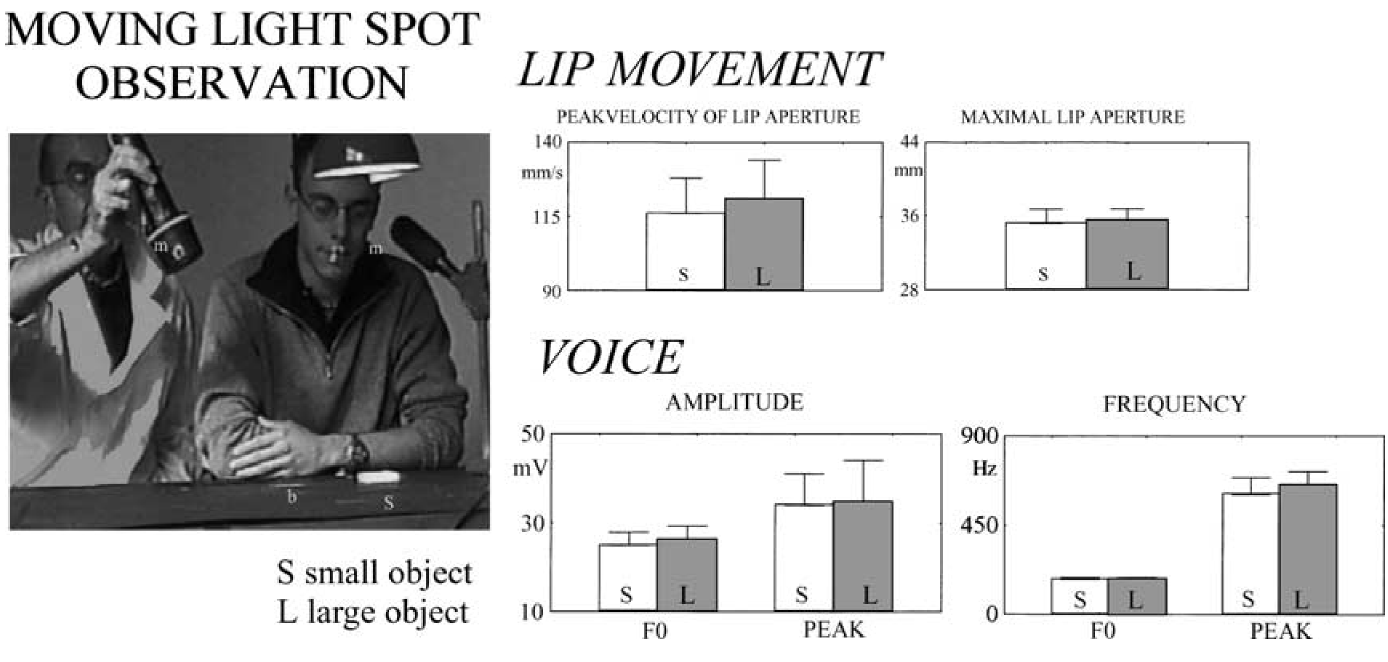


FIG. 2. The experimental apparatus and mean values of parameters of lip kinematics and voice analysed in experiment 2. Conventions as in Fig. 1.

to the large object like during reaching-grasping of the same object in experiment 1 (see below). Five females and 5 males participated in the experiment.

### *Experiment 3: overt-stimulus-size analysis*

Subjects were required to analyse overtly the visual stimulus and, in particular, its size before pronouncing the syllable. The following two procedures were designed in order to force subjects to analyse overtly the stimulus every time it was presented, even if it could be of only two sizes. First, before the experimental session, they were informed that, at the end of the experimental session, they would be asked questions concerning object features and, in particular, their size. Second, at the beginning of the session subjects did not know how many syllables and stimuli would have been presented. Six different objects were randomly presented during the experimental session. The latter procedure was the same as in the other experiments, the aim of which was to induce observation of grasping (experiment 1), and of pointing of the light spot to (experiment 2) objects of the same two sizes. In each trial, the light was turned off after syllable pronunciation, or approximately 2 s after cloud configuration presentation. Four females and 4 males participated in the experiment.

Each experiment consisted of 36 trials. In 12 trials the BA configuration was presented, in other 12 trials the GA configuration was presented, and, in the remaining trials, the cloud configuration was presented. For each configuration the target was large in 6 trials, and was small in the other 6 trials. Configurations and object sizes were randomly presented.

Movements of the subject's mouth, experimenter's arm, and torch were recorded using the 3D-optoelectronic ELITE system (B.T.S., Milan, Italy). It consists of two TV cameras detecting infrared reflecting markers at the sampling rate of 50 Hz. Movement reconstruction in 3D coordinates and computation of the kinematic parameters are described in a previous work (Gentilucci *et al*., 1992). In the present study, either 2 or 3 markers were used. In all experiments, two markers were placed on the centre of the subject's upper and lower lip, respectively. Subject's lip aperture during syllable pronunciation was studied by analysing the time course of the distance between upper and lower lip. Peak velocity of lip aperture, and maximal lip aperture were measured. In experiment 1, another marker was placed on the ulnar side of the experimenter's right wrist; it was used to analyse the reach component. The time course of the wrist position along the subject's sagittal, vertical, and transverse axes, and the profile of the module of the tangential velocity vector were computed. The analysed kinematic parameters were peak arm velocity and movement time. In experiment 2, the third marker was placed on the torch. Peak velocity and movement time were measured. The procedures for calculating beginning and end of hand, torch, and lip movements were identical to those previously described (Gentilucci *et al*., 1994).

The voice emitted by subjects during syllable pronunciation was recorded. A microphone (Studio Electret Microphone, 20–20000 Hz, 500 Ω, 5 mv/Pa/1 kHz) was placed on the table by means of a support (Fig. 2). The centre of the support was 35 cm from the subject's chest and 18 cm from the subject's sagittal axis. The microphone was connected to a PC for sound recording by a card device (16 PCI Sound Blaster, Creative Technology Ltd, Singapore). Spectrogram of each pronounced syllable was computed and amplitude spectrum of the selected period (see below) was constructed by means of the spectrogram (Fig. 1). Abscissa was frequency and ordinate was mean voltage amplitude. As the aim was to study possible mouth articulation changes during voice emission, peaks in the amplitude spectrum ranging from 250 to 3500 Hz were analysed. Peaks could correspond to one of the formants 1, 2, and 3 (F1, F2, and F3) in Italian speech (Ferrero *et al*., 1979). Maximal peak and corresponding frequency of the amplitude spectrum were measured. Amplitude and frequency of the fundamental formant (F0), which is mainly related to vocalisation (i.e. modification in the vocal cords and air flux) were analysed. The ratio between maximal peak and F0 amplitude was calculated (the resonance factor). It can be an index of the amplification (due to resonance) of the sound amplitude depending on mouth articulation modification.

The experimental design included the following two within-subject factors for the parameters of lip aperture, voice of subjects, reaching-grasping of experimenter, and torch movement: object size (small vs. large), and syllable (BA vs. GA). Separate ANOVAs were carried out on mean values of the parameters. Newman-Keul's *post hoc* test was used.

## Results

### *Experiment 1*

Examples of spectrograms and amplitude spectra of the syllables BA and GA are shown in Fig. 3. Note, in the spectrograms, the following three phases: a low frequency sound phase due to the release of the mouth occlusion (C), a transitional phase during which sounds due to vowel and consonant cannot be disjoined from each other because of mouth coarticulation (C/V, Liberman, 1970), and a final phase during which only the vowel is audible (V). The F2 pattern, in particular, during the transitional phase was different between the two syllables (Ferrero *et al*., 1979). Amplitude spectra were constructed by taking into account C/V and V phases. Figure 1 shows mean values of parameters of lip aperture kinematics and voice amplitude spectrum. Maximal lip aperture was significantly greater during observation of grasp of the large object ($F_{1,7} = 7.3$, $P < 0.05$). Similarly, peak velocity of lip aperture showed a trend to significance ($F_{1,7} = 5.0$, $P = 0.055$). Voice peak amplitude corresponded to F1 amplitude. F0 amplitude ($F_{1,7} = 5.7$, $P < 0.05$) and peak amplitude ($F_{1,7} = 12.7$, $P < 0.01$) were higher when observing the grasp of the large object, whereas frequency of F0 decreased ($F_{1,7} = 7.8$, $P < 0.01$). The resonance factor was significantly higher when observing the grasp the large object ($F_{1,7} = 5.4$, $P = 0.05$, 1.82 vs. 1.51).

Peak velocity of lip aperture was higher when pronouncing BA than GA ($F_{1,7} = 11.2$, $P < 0.01$, 162.9 vs. 117.8 mm/s) in accordance with the fact that BA is a bilabial plosive with the lips closed at the onset of syllable pronunciation, whereas GA is a velar plosive with the back tongue closed on the soft palate at syllable onset. Consequently, lips moved faster when pronouncing BA than GA.

Experimenter's arm peak velocity ($F_{1,7} = 43.8$, $P < 0.0005$, large object 524.9 mm/s, small object 457.0 mm/s) and movement time ($F_{1,7} = 28.2$, $P < 0.001$, large object 814.9 ms, small object 1062.7 ms) were significantly affected by object size.

In experiments 2 and 3, the possibility that stimulus velocity rather than observation of a moving (reaching-grasping) biological stimulus (hand), and overt visual analysis of object size influenced syllable pronunciation in experiment 1 was tested, respectively.

### *Experiment 2*

Torch kinematics was analysed in order to match it with the kinematics of experimenter's hand in experiment 1. When the torch pointed to the large object, peak velocity ($F_{1,9} = 68.6$, $P < 0.00005$, large object 349.3 mm/s, small object 115.9 mm/s) was higher and movement time ($F_{1,9} = 9.0$, $P < 0.01$, large object 505.0 ms, small object 733.9 ms) was shorter than when the torch pointed to the small object. Distance of

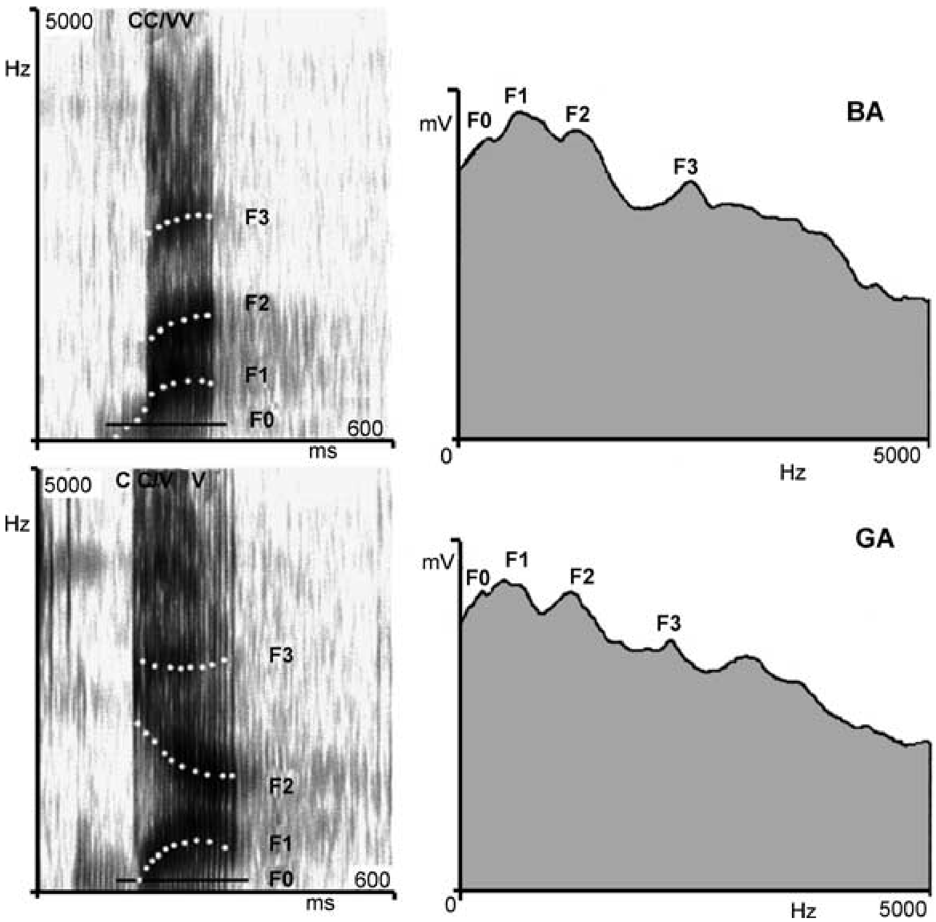


FIG. 3. Examples of spectrograms (left panels) and amplitude spectra (right panels) when pronouncing the syllables BA (upper panels) and GA (lower panels) and contemporaneously observing the grasping of a large object (subject B.I.). In the spectrograms, C, C/V and V indicate the following three phases: a low frequency sound phase due to the release of the mouth occlusion (C), a transitional phase during which the sounds of vowel and consonant cannot be disjoined from each other because of mouth coarticulation (C/V), and a final phase during which only vowel is audible (V). White dots represent amplitude peaks of F1, F2, and F3. Note the different pattern of the formants during the transitional phase according to the consonant. F0: fundamental frequency. The amplitude spectra were constructed by considering the C/V and V phases. Note the peaks of F1, F2, and F3.

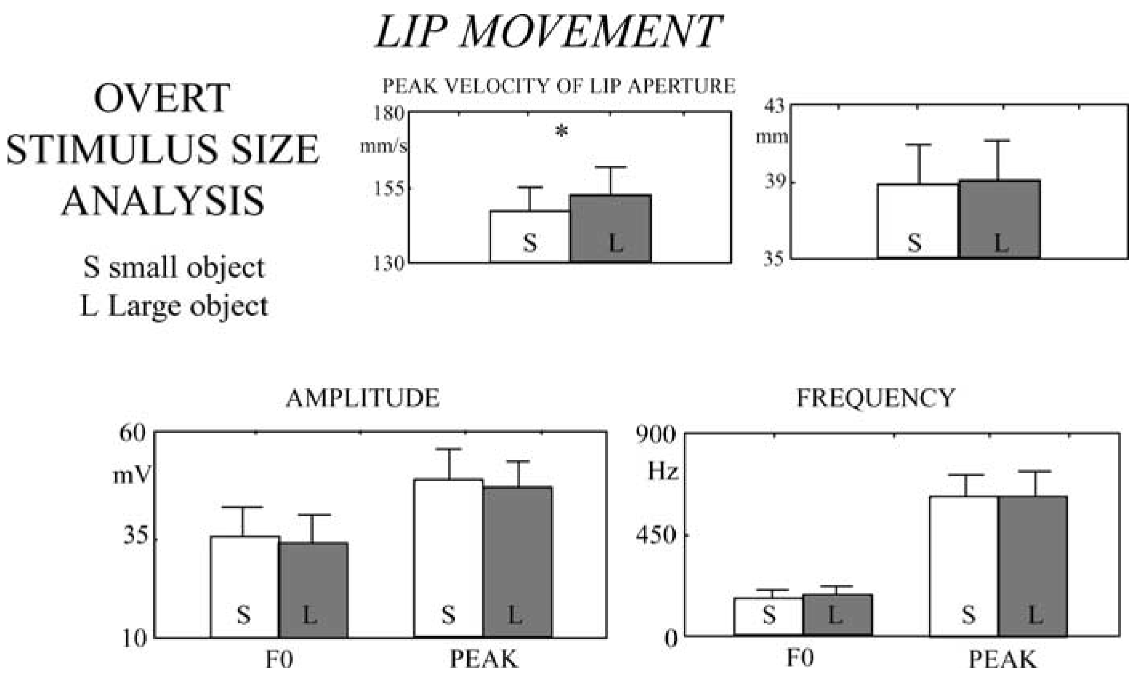


FIG. 4. Mean values of parameters of lip kinematics and voice analysed in experiment 3. Conventions as in Fig. 1.

the marker on the torch from the rotation fulcrum (experimenter's wrist) was approximately 1/3 of that of the light spot on the table. Consequently, the velocity of the light spot was three times higher than the velocity of the marker, i.e. peak velocities were approximately 1050 and 350 mm/s for movements directed to the large and the small object, respectively. Summing up, velocities of the light spot were of the same order of magnitude and differences in stimulus velocity were similar to those of the experimenter's arm measured in experiment 1. This allowed the comparison of results obtained in experiments 2 with those obtained in experiment 1.

The results concerning subjects' lip aperture kinematics and voice are summarised in Fig. 2. Factor object size affected no parameter of lip aperture and voice. Peak velocity of lip aperture ($F_{1,9} = 44.8$, $P < 0.0001$, 156.7 vs. 81.7 mm/s), and voice peak frequency ($F_{1,9} = 7.4$, $P < 0.05$, 685.6 vs. 573.3 Hz) increased when pronouncing BA.

### *Experiment 3*

Mean values of parameters of lip aperture kinematics and voice amplitude spectrum are presented in Fig. 4. Factor object size was significant only for peak velocity of lip aperture ($F_{1,7} = 5.2$, $P = 0.05$, Fig. 4). Peak velocity of lip aperture increased when observing the grasp of the large object. Peak velocity of lip aperture ($F_{1,7} = 12.3$, $P < 0.01$, 175.1 vs. 124.5 mm/s), voice peak amplitude ($F_{1,7} = 11.1$, $P < 0.01$, 52.9 vs. 42.8 mV), and peak frequency ($F_{1,7} = 5.8$, $P < 0.05$, 699.4 vs. 538.2 Hz) increased when pronouncing BA.

## Discussion

The results of experiment 1 are in favour of the hypothesis that the observation of a grasping hand can influence syllable pronunciation. However, another interpretation is possible. Subjects were instructed to pronounce the syllable as soon as the experimenter's hand touched the object, and likely they paid attention to the moving hand. As the experimenter's arm velocity varied according to object size, subjects might have synchronised inner and outer mouth velocity with the experimenter's arm velocity. This might have influenced voice emission. In conclusion, subjects' syllable pronunciation might have been affected by stimulus velocity independently of the observation of a grasping motor act. However, this interpretation can be discarded by considering the results of experiment 2 in which a light spot pointed with different velocities to the same objects as in experiment 1. These are in favour of the hypothesis that the observation of a biological stimulus performing an aimful movement is the necessary prerequisite for an interference effect on lip movement and voice emission.

Another factor might have influenced syllable pronunciation in experiment 1. Subjects might have overtly analysed the size of the objects. This explanation was verified in experiment 3 in which subjects were explicitly required to analyse the size of the same objects presented in experiment 1. A size effect was observed only on peak velocity of lip aperture. The poor effect might be due to the fact that, as only two object sizes were presented, subjects did not overtly analyse its features after few trials. However, if this was true, no overt analysis of object should have occurred also in experiment 1, during which the same stimuli as in experiment 3 were presented and, in addition, no overt object analysis was required. Summing up, the modification in lip kinematics and voice observed in experiment 1 cannot be due to overt visual analysis of the object.

Labial movement was influenced by the observation of a grasping hand. Modification in labial movement was congruent with the imitation with the mouth of the observed grasp-with-hand movement. Peak amplitude of the pronounced syllables corresponded to the frequencies of F1 in the amplitude spectrum. Indeed, F1 frequency of the Italian vowel/a/(Ferrero *et al*., 1979) is of the same order of magnitude of the reported peak frequency. Voice peak amplitude and F0 amplitude were higher when observing grasping of the larger object. Increase in F0 amplitude could correspond to higher vocalisation. Conversely, the increase in peak amplitude could be due to modification also in the mouth articulation. The latter interpretation is supported by the finding that the resonance factor was higher when observing the grasp of the large object. Taken together, the data of the present study suggest that observation of hand grasp motor acts influences movements of the inner and outer mouth during syllable pronunciation.

The following mechanism explaining the effect of grasp observation on syllable pronunciation can be hypothesised: Observation of the grasping motor act activated preparation of the same motor act in the observer (Fadiga *et al*., 1994). Preparation concerned grasping the object with both hand and mouth (Gentilucci *et al*., 2001). Preparation of grasp with the mouth influenced syllable pronunciation. The finding that grasp observation has access to the areas involved in speech production is supported by neuroimaging data showing that the observation of grasping with the hand and the mouth, but not of motor acts performed by other effectors, activates the Broca's area (Buccino *et al*., 2001).

The data of the present study suggest that phonemic representations expressed by mouth articulation gestures correlates with representations of hand gesture. Indeed, an influence of grasp observation on syllable production was observed. It may be speculated that the grasp representation is related to the phonemic representation also in terms of causality. It has been hypothesised that speech has evolved from a communication system based on hand gestures (Armstrong *et al*., 1995). This hypothesis can be supported by the properties of the 'mirror' system. Indeed, as this system seems to serve for both *recognition* and *production* of actions (Gallese *et al*., 1996), it supplies the link between sender and receiver – the necessary prerequisite of intended interaction – and may have played a determinant role in the development of motor skills for refining willed communication. The human ability to communicate beyond that of other primates may depend on the progressive evolution of the mirror system in its globality. However, it is possible that this system maintained visual access to the communicative hand gestures. What I observed in the present study might be remnants of this ancient system.

## Acknowledgements

I wish to thank L. Caselli and C. Secchi, for the help in carrying out the experiments. I thank G. Rizzolatti for discussion of data and G. Cossu, G. DiPellegrino and V. Gallese for the comments on the manuscript. The work was supported by a grant from MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca).

## References

Armstrong, D.F., Stokoe, W.C. & Wilcox, S.E. (1995) *Gesture and the Nature of Language*. Cambridge: University Press.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese. V., Seitz, R.J., Rizzolatti, G. & Freund, H.J. (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.*, **13**, 400–4004.

Demonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.C., Wise, R. & Frackowiak, R.S.J. (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain*, **115**, 1753–1768.

Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1994) Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.*, **73**, 2608–2611.

Ferrero, F., Genre, A., Boë, L.J. & Contini, M. (1979) *Nozioni di fonetica acustica*, Torino, Edizioni Omega.

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

Gentilucci, M., Benuzzi, F., Gangitano, M. & Grimaldi, S. (2001) Grasping with hand and mouth: a kinematic study on healthy subjects. *J. Neurophysiol.*, **86**, 1885–1699.

Gentilucci, M., Chieffi, S., Scarpa, M. & Castiello, U. (1992) Temporal coupling between transport and grasp components during prehension movements: effects of visual perturbation. *Behav. Brain Res.*, **47**, 71–82.

Gentilucci, M., Toni, I., Chieffi, S. & Pavesi, G. (1994) The role of proprioception in the control of prehension movements: a kinematic study in a peripherally deafferented patient and in normal subjects. *Exp. Brain Res.*, **102**, 483–494.

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

Liberman, A.M. (1970) The grammars of speech and language. *Cogn. Psychol.*, **1**, 301–323.

Oldfield, R.C. (1971) The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, **9**, 97–113.

Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993) The neural correlates of the verbal component of working memory. *Nature*, **362**, 342–345.

Rizzolatti, G. & Arbib, M.A. (1998) Language within our grasp. *TINS*, **21**, 188–194.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. (1988) Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.*, **71**, 491–507.

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