Speech and gesture share the same communication system

Paolo Bernardis, Maurizio Gentilucci *
Dipartimento di Neuroscienze, Università di Parma, via Volturno 39, 43100 Parma, Italy
Received 4 February 2005; received in revised form 12 April 2005; accepted 10 May 2005

Abstract
Humans speak and produce symbolic gestures. Do these two forms of communication interact, and how? First, we tested whether the two communication signals influenced each other when emitted simultaneously. Participants either pronounced words, or executed symbolic gestures, or emitted the two communication signals simultaneously. Relative to the unimodal conditions, multimodal voice spectra were enhanced by gestures, whereas multimodal gesture parameters were reduced by words. In other words, gesture reinforced word, whereas word inhibited gesture. In contrast, aimless arm movements and pseudo-words had no comparable effects. Next, we tested whether observing word pronunciation during gesture execution affected verbal responses in the same way as emitting the two signals. Participants responded verbally to either spoken words, or to gestures, or to the simultaneous presentation of the two signals. We observed the same reinforcement in the voice spectra as during simultaneous emission. These results suggest that spoken word and symbolic gesture are coded as single signal by a unique communication system. This signal represents the intention to engage a closer interaction with a hypothetical interlocutor and it may have a meaning different from when word and gesture are encoded singly.

© 2005 Elsevier Ltd. All rights reserved.
Keywords: Symbolic gesture; Speech; Arm kinematics; Voice spectra; Social intention

1. Introduction

Gesture is a universal feature of human communication. Every culture speakers produce gestures, although the extent and typology of the produced gesture vary. For some types of gestures, execution is frequently associated with speech production (see Goldin-Meadow, 1999; see McNeill, 1992). In particular, speakers frequently pronounce words while they also execute symbolic gestures, which express the same meaning as the word. Consider expressing approbation. While one pronounces “ok”, one often forms a circle with the forefinger and thumb in contact at their tips, while the rest of the fingers extend outward.

There are two competing views of the relationship between gesture and speech. The first posits that gesture and speech are separate communication systems (Hadar, Wenkert-Olenik, Krauss, & Soroker, 1998; Krauss & Hadar, 1999; Levelt, Richardson, & La Heij, 1985). According to this view, gesture functions as an auxiliary support when verbal expression is temporally disrupted or word retrieval is difficult. The second view (McNeill, 1992) posits that gesture and speech form a single system of communication, because they are linked to the same thought processes even if the expression modality differs.

Neuropsychological and neurophysiological evidence is in accord with the idea that speech and gestures share common neural substrates. It is well known that left cortical lesions producing aphasia are frequently associated to apraxia and, especially to ideomotor apraxia, a disorder of purposeful actions (Heath, Roy, Black, & Westwood, 2001). Brain-imaging studies show that language areas, such as Broca’s area, are activated by imitation (Grezes, Armony, Rowe, & Passingham, 2003; Tanaka & Imi, 2002), by observation of gestures (Decety et al., 1997; Grezes et al., 2003; MacSweeney et al., 2002, 2004) or, more generally, by observation of upper arm actions (Buccino et al., 2004).

From an evolutionary point of view, Stokoe and colleagues (Amstrong, Stokoe, & Wilcox, 1995; see also Corballis, 2002; Hewes, 1973) hypothesized that spoken...
language derives from an ancient system which used arm gestures in order to communicate. Interestingly, the 18th-century Neapolitan philosopher, Giam battista Vico, in *La Scienza Nuova* (1744/1953) also formulated a theory of the origin of language proposing that humans were originally mute and communicated by gestures, not by speech. In the same vein, Condillac (1746) proposed that language evolved from manual gestures, and not from animal calls. Recently Corballis (2002) proposed that spoken language developed as the repertoire of gestures was gradually transferred from arm to mouth. This may have happened because of the existence of double motor commands simultaneously sent to both arm and mouth (Gentilucci, Benussi, Gangitano, & Grimaldi, 2001). In addition there is a strict relationship between early language development in children and several aspects of manual activity, such as communicative and symbolic gestures (for a review see Bates & Dick, 2002; Volterra, Caselli, Capirci, & Pizzuto, 2005). Canonical babbling in 6–8 months aged children is accompanied by rhythmic hand movements (Masataka, 2001). Word comprehension in children between 8 and 10 months and word productions between 11 and 13 months are accompanied by deictic and recognition gestures, respectively (Bates & Snyder, 1987; Volterra, Bates, Benigni, Bretherton, & Camaioni, 1979).

From a behavioral point of view, if pronouncing words and executing symbolic gestures with the same meaning are controlled by a single communication system, these two tasks in an experimental setting should interact. Previous studies (Levelt et al., 1985; McNeill, 1992) showed that word pronunciation and gesture execution are temporally coordinated. More generally, Kimura (1973) showed that speaking is frequently associated with free movements of the hand opposite to the speech hemisphere. Conversely, we showed that hand actions guided by an object influence voice spectra of syllables pronounced simultaneously with the actions (Gentilucci, Santunione, Roy, & Stefanini, 2004; Gentilucci, Stefanini, Roy, & Santunione, 2004). In experiments 1 and 2 we aimed to extend the conclusions of previous studies and to determine whether gesture execution and word pronunciation are functionally related. If the hypothesis of a single communication system is true, the command to execute a meaningful gesture should modify the voice spectra of a word having the same meaning, but not of a meaningless word (i.e. a pseudo-word). Conversely, if the hypothesis about a single communication system is true, this system should be involved in the execution and in the interpretation of messages expressed by words or gestures, or by both. Indeed, a link is necessary between actor and observer as the link between sender and receiver of each message is. According to Rizzolatti and Arbib (1998), the “mirror system” may provide the necessary bridge from ‘doing’ to ‘communicating’. Consequently, the observation of both word and gesture should affect the voicing of a verbal response just as the simultaneous emission of the two communication signals can do. This hypothesis moved from our previous findings (Gentilucci, 2003a; Gentilucci, Santunione, et al., 2004; Gentilucci, Stefanini, et al., 2004) showing that the observation of transitive hand actions such as grasping and bringing to the mouth an object affects voicing of syllables pronounced during action observation. The effects were the same as those found during their execution (Gentilucci, Santunione, et al., 2004; Gentilucci, Stefanini, et al., 2004).

In experiments 3 and 4, we compared the voice spectra of the same words used in the first part of the study, repeated after presentation of either spoken word or gesture with those repeated after simultaneous presentation of both the two communication signals.

2. Experiments 1 and 2

We verified whether gesture execution and word pronunciation influenced each other when simultaneously emitted. We expected that the execution of a meaningful gesture modified the voice spectra of a word having the same meaning, but not of a meaningless word (i.e. a pseudo-word). Conversely, we expected that the pronunciation of a word modified the kinematics of the corresponding meaningful gesture, but not of a meaningless gesture.

2.1. Methods

2.1.1. Participants

Twenty-eight Italian male adults (age 21–24 years) were classified as right-handed according to Edinburgh Inventory (Oldfield, 1971). They were divided in two groups of 14 individuals participating in one of the two experiments. Only male participants were included to reduce variability in vocal formants (Ferrero, Magno Caldognetto, & Cosi, 1996; Pickett, 1999). The Ethics Committee of the Medical Faculty of the University of Parma approved the study. All participants were naive to the purpose of the study.

2.1.2. Apparatus and stimuli

Participants sat on a chair in front of a table. Before each trial, their right wrist was placed on a fixed starting position. The stimuli were the three words CIAO (/tʃaʊ/), NO (/nɔ/), STOP (/stɒp/) printed on a PC monitor. In addition, the string of letters XXX in experiment 1 or the pseudo-word LAO (/lɔ/ in experiment 2 were presented. LAO was chosen because it contains the same vowels as the other presented words. The stimuli were presented in medium grey (approximately 6 cd/m²) against a black background (1 cd/m²) on a 19 in. SONY LCD monitor controlled by a PC. The monitor was set at a spatial resolution of 1024 pixels x 768 pixels and at a temporal resolution of 75 Hz. It was placed in front of the table, at a distance of 2 m from the participant. Each trial started with a BEEP followed by a black screen (300 ms), after which the stimulus was presented (duration of 1000 ms).
2.1.3. Procedure
In four successive blocks of trials the participants were required to emit different responses at the end of stimulus (printed word, pseudo-word or string XXX) presentation. The response 1 was the gesture having the same meaning as the printed word (meaningful gesture, Fig. 1). The response 2 was pronunciation of the word or the pseudo-word. The response 3 was pronunciation of the word and simultaneous execution of the corresponding gesture. The response 4 in experiment 1 was pronunciation of the word and execution of a meaningless arm movement (meaningless gesture). This consisted of up-down oscillatory movements involving wrist, elbow, and shoulder joints just as the meaningful gestures (Fig. 1). This meaningless gesture was executed also when the string XXX was presented. The response 4 of experiment 2 was execution of the gesture corresponding to the printed word and pronunciation of the pseudo-word LAO. To exclude that silent repetition of the read word influenced gesture execution, the response 1 of experiments 1 and 2 and the response 4 of experiment 2 was emitted after counting aloud from 1 to 5. We preferred to run two experiments instead of a longer one to insure that the task did not become automatic and participants emitted the responses in a spontaneous way like in everyday life. Summing up, both experiments 1 and 2 consisted of four experimental conditions in each of which a different response was required to stimuli presented with the same modality.

The participants were presented with each condition in one block of trials during the experimental session. The blocks of trials were quasi-counterbalanced across participants. Each stimulus was quasi-randomly presented five times during each block of trials. Consequently, each of the four blocks of trials consisted of 15 trials (3 stimuli × 5 repetitions) except the blocks corresponding to condition 1 of experiment 1 and condition 2 of experiment 2, which consisted of 20 trials (4 stimuli × 5 repetitions). Participants were left free to use as much time as needed to complete each trial.

2.1.4. Data recording
The participants wore a light-weight dynamic headset microphone (Shure, model WH20). The frequency response of the microphone ranged from 50 to 15,000 Hz. The microphone was connected to a second PC by a sound card (16 PCI Sound Blaster; CREATIVE Technology Ltd, Sin-
than 0.4 mm (Gentilucci, Toni, Chieffi, & Pavesi, 1994). The spatial error of the ELITE system is 0.4 mm, tip movement) with respect to the end of stimulus presentation. Since the spatial error of the ELITE system is 0.4 mm, a measure of how much the movement was stereotyped), and, less gesture; S.D. of the number of velocity peaks (which gave oscillations of the CIAO and NO gestures and the meaningless gesture; number of velocity peaks, i.e. the maximal peak velocity during the CIAO and NO gestures and the meaningless gesture; maximal peak velocity along any Cartesian axis in each of the eight forthcoming frames. The frame corresponding to movement beginning was selected after verifying that the displacement of the marker on the index finger tip increased along any Cartesian axis in each of the eight forthcoming frames. The frame corresponding to movement beginning was selected after verifying that the displacement of the marker on the index finger tip increased along any Cartesian axis in each of the eight forthcoming frames. The frame corresponding to movement beginning was selected after verifying that the marker went down along the vertical axis in each of the four forthcoming frames.

2.1.5. Data analysis

Three series of analyses were carried out. The first series aimed to characterize the voice spectra of the words and the kinematics of the gestures when executed singly. Only data of the two baseline conditions, i.e. when either the word was singly pronounced or the gesture was singly executed, were included. In the ANOVAs the between-subjects factor was experiment (experiment 1 versus experiment 2) and the within-subjects factor was communication signal (i.e. either word or gesture, CIAO versus NO versus STOP). For S.D. of number of velocity peaks the within-subjects factor was gesture (CIAO versus NO versus meaningless gesture). This analysis aimed to discover whether the meaningless gesture was less stereotyped than the NO and CIAO gestures.

The second series of analyses aimed to discover differences between the baseline condition and each of the other three conditions (i.e. word and gesture, word and meaningless gesture, pseudo-word and gesture). We performed the following comparisons:

(a) The first comparison was between the condition involving production of simultaneous word and congruent gesture and the condition involving only the emission of word or meaningful gesture. In the ANOVAs the between-subjects factor was experiment (experiment 1 versus experiment 2) and the within-subjects factors were condition (either word or gesture versus word and meaningful gesture) and communication signal (CIAO versus NO versus STOP).

(b) The second comparison was between the condition involving the only pronunciation of a word, and the one involving both a meaningless gesture and a word (experiment 1). In the ANOVAs the within-subjects factors were condition (no gesture versus meaningless gesture) and word (CIAO versus NO versus STOP) for voice parameters, whereas for meaningless gesture parameters it was condition (no word versus word).

(c) The third comparison was between the condition involving only gesture and the one involving the pronunciation of the pseudo-word LAO during the execution of meaningful gestures (experiment 2). In the ANOVAs the within-subjects factors were condition (no word versus pseudo-word) and gesture (CIAO versus NO versus STOP) for gesture parameters, whereas for voice parameters of the pseudo-word LAO it was condition (no gesture versus meaningful gesture).

The third series of analyses concerned the temporal relationship between gesture and speech beginnings. When comparing words with congruent gestures, the between-subjects factor was experiment (experiment 1 versus experiment 2) and the within-subjects factors were effector (mouth versus hand) and communication signal (CIAO versus NO versus STOP). In both the comparisons of meaningless gesture with words and of meaningful gestures with pseudo-word the within-subjects factors were effector (mouth versus
hand) and communication signal (CIAO versus NO versus STOP).
In all analyses paired comparisons were performed using
the Newman–Keuls procedure. The significance level was
fixed at $P < 0.05$.

2.2. Results

2.2.1. Emission of single communication signals
(baseline condition)
We firstly analyzed the words pronounced without per-
forming gestures and, conversely, the corresponding-in-
meaning gestures executed without pronouncing words (first
series of analyses).

Fig. 2 presents examples of spectrograms when pronounc-
ing the words CIAO, NO, and STOP. For all the three words
the time course of F1 showed an increasing phase followed
by a decreasing phase, whereas the time course of F2 was
characterized only by a decreasing phase. However, both
mean F1 and F2 of the three words varied. F1 of NO was
significantly higher than F1 of CIAO, and F1 of CIAO was
significantly higher than F1 of STOP (Fig. 3, Table 1). F2
gradually decreased moving from CIAO to NO and STOP
(Fig. 3, Table 1). Moreover, intensity, pitch and vowel dura-
tion differed among the three words (Fig. 3, Table 1).

Fig. 1 shows representative examples of the index finger
path when participants executed the three gestures. Note
that the amplitudes of the hand oscillations were larger for CIAO
than for NO. The ANOVAs showed that gesture times of
CIAO and NO were, as expected, longer than that of STOP
(Fig. 4, Table 1). The maximal height reached during the
NO gesture was lower than that of CIAO and STOP (Fig. 4,
Table 1). Finally, hand oscillation amplitude, maximal peak
velocity, and number of velocity peaks of the NO gesture
decreased compared with those of CIAO (Fig. 4, Table 1).
S.D. of number of velocity peaks did not differ between
meaningful and meaningless gestures (Table 1, meaningful
gestures 1.10, meaningless gesture 1.31). In other words, the
meaningless movement was stereotyped as well as the mean-
ingful gestures were.

Table 1
Voice spectra parameters and kinematics parameters in experiments 1 and 2. Results of the statistical comparisons among the words and gestures in the baseline condition

<table>
<thead>
<tr>
<th>Voice spectra parameters</th>
<th>Kinematics parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Gesture time</td>
</tr>
<tr>
<td>$F(2,52) = 83.3, P &lt; 0.00001$</td>
<td>$F(2,52) = 94.8, P &lt; 0.00001$</td>
</tr>
<tr>
<td>F2</td>
<td>Maximal arm height</td>
</tr>
<tr>
<td>$F(2,52) = 22.8, P &lt; 0.00001$</td>
<td>$F(2,52) = 6.2, P &lt; 0.00001$</td>
</tr>
<tr>
<td>Intensity</td>
<td>Hand oscillation amplitude</td>
</tr>
<tr>
<td>$F(2,52) = 26.0, P &lt; 0.00001$</td>
<td>$F(1,26) = 17.8, P &lt; 0.0003$</td>
</tr>
<tr>
<td>Pitch</td>
<td>Maximal peak velocity</td>
</tr>
<tr>
<td>$F(2,52) = 6.2, P &lt; 0.004$</td>
<td>$F(1,26) = 16.3, P &lt; 0.0004$</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>Number of velocity peaks</td>
</tr>
<tr>
<td>$F(2,52) = 614.0, P &lt; 0.00001$</td>
<td>$F(1,26) = 6.6, P &lt; 0.01$</td>
</tr>
<tr>
<td>S.D. of number of velocity peaks</td>
<td>$F(2,24) = 0.80, P = 0.46$, n.a.</td>
</tr>
</tbody>
</table>

S.D. of number of velocity peaks refers to comparisons among the gestures CIAO and NO, and the meaningless gesture.
2.2.2. Comparison between words pronounced without and with gestures

In the second series of analyses, we studied the effects of meaningful and meaningless gestures on the pronunciation of words and pseudo-words. In the comparison with the pronunciation of words alone, F2 increased when CIAO, NO, and STOP were pronounced while executing a gesture of the same meaning (Fig. 5, Table 2). In contrast, a meaningless gesture did not affect F2 of the three words (Table 2). Similarly, the meaningful gestures did not affect pronunciation of the meaningless word LAO (Table 2). The meaningful and meaningless gestures induced an increase in F1 of the word STOP (Fig. 5, Table 2). In contrast, F1 of the pseudo-word was not affected by the meaningful gestures (Table 2).

Finally, the pitch of the three words increased when they were pronounced simultaneously with meaningful gestures but not when pronounced simultaneously with meaningless gestures (Fig. 5, Table 2). The pitch of LAO was not affected by the gestures (Table 2). Both the intensity and duration of the vowels were not affected by the execution of meaningful and meaningless gestures.

2.2.3. Comparison between gestures executed with and without pronunciation of words

The second series of analyses allowed us to study the effects of words and pseudo-words on the execution of meaningful and meaningless gestures. In the comparison with the sole execution, the duration of meaningful gestures was short-

<table>
<thead>
<tr>
<th>Word</th>
<th>No gesture vs. meaningful gesture</th>
<th>No gesture vs. meaningless gesture</th>
<th>Pseudo-word</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>(F(1,26) = 6.8, P &lt; 0.002)</td>
<td>(F(2,26) = 3.4, P = 0.04)</td>
<td>(F(1,39) = 0.00, P = 0.98, \text{n.s.})</td>
</tr>
<tr>
<td>F2</td>
<td>(F(1,26) = 11.6, P &lt; 0.002)</td>
<td>(F(1,35) = 1.9, P = 0.18, \text{n.s.})</td>
<td>(F(1,39) = 0.04, P = 0.99, \text{n.s.})</td>
</tr>
<tr>
<td>Pitch</td>
<td>(F(1,26) = 11.6, P &lt; 0.002)</td>
<td>(F(1,35) = 3.15, P = 0.10, \text{n.s.})</td>
<td>(F(1,39) = 1.5, P = 0.23, \text{n.s.})</td>
</tr>
</tbody>
</table>

The analyses on intensity and vowel duration are not reported because not significant in all the three conditions.
Fig. 4. Mean values of arm kinematics parameters of gestures executed without simultaneous pronunciation of words or pseudo-word in experiments 1 and 2. Bars are S.E.

Ened by the pronunciation of words of the same meaning (Fig. 6, Table 3), but not of meaningless word (Table 3). The duration of the meaningless gesture was not significantly affected by the words (Table 3). The maximal height reached by the hand was lowered only by the pronunciation of the words (Fig. 6, Table 3), which, however, did not affect the maximal height of the meaningless gesture (Table 3). The same results were found for hand oscillation amplitude and maximal peak velocity during the hand oscillations of CIAO and NO (Fig. 6, Table 3), although the word effect on the latter parameter was more evident in experiment 1 than in experiment 2 (Fig. 6, interaction between experiment and condition, Table 3). The number of velocity peaks of both the meaningful (Fig. 6, Table 3) and the meaningless gestures (Table 3) decreased when the word was simultaneously pronounced with the gesture. In contrast, the kinematics of the meaningful gestures was not affected by the pseudo-word. Finally, S.D. of number of velocity peaks of meaningful and meaningless gestures was not affected by word and pseudo-word pronunciation.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Meaningful gesture</th>
<th>Meaningless gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No word vs. word</td>
<td>No word vs. pseudo-word</td>
</tr>
<tr>
<td>Gesture time</td>
<td>$F(1,26) = 19.9$, $P &lt; 0.001$</td>
<td>$F(1,13) = 1.02$, $P = 0.34$, n.s.</td>
</tr>
<tr>
<td>Maximal height</td>
<td>$F(1,26) = 4.9$, $P &lt; 0.03$</td>
<td>$F(1,13) = 1.5$, $P = 0.24$, n.s.</td>
</tr>
<tr>
<td>Hand oscillation amplitude</td>
<td>$F(1,26) = 23.7$, $P &lt; 0.0001$</td>
<td>$F(1,13) = 0.7$, $P = 0.42$, n.s.</td>
</tr>
<tr>
<td>Maximal peak velocity</td>
<td>$F(1,26) = 9.9$, $P &lt; 0.004$</td>
<td>$F(1,13) = 2.10$, $P = 0.18$, n.s.</td>
</tr>
<tr>
<td>Number of velocity peaks</td>
<td>$F(1,26) = 6.3$, $P &lt; 0.002$</td>
<td>$F(1,13) = 4.4$, $P = 0.26$, n.s.</td>
</tr>
</tbody>
</table>

The analyses on S.D. of number of velocity peaks are not reported because not significant in all the three conditions.
2.2.4. Temporal relations between the beginning of speech and gesture

The third series of analyses showed that the pronunciation of words and pseudo-word always started after the beginning of the meaningful and meaningless gestures. In addition, whereas the beginning of speech did not vary for the three words, the NO gesture started in advance compared with the other two meaningful gestures (Fig. 7, interaction between effector and communication signal, $F(2,52) = 4.2$, $P < 0.02$, $P < 0.05$). The interval between the beginning of word and meaningless gestures decreased (546.8 versus 375.2 ms, $F(1,12) = 7.3$, $P < 0.03$, $P < 0.05$), whereas the corresponding interval between pseudo-words and meaningful gestures increased (517.8 versus 693.1 ms, $F(1,12) = 13.2$, $P < 0.003$), in the comparison with the interval between word and meaningful gesture.

3. Experiments 3 and 4

We presented participants with the same communication signals (words and/or gestures) emitted in experiments 1 and 2 and studied the effect of the observation of these stimuli on the pronunciation of the corresponding words. We expected the same effects on the voice spectra as those found for the emission of the communication signals.

3.1. Methods

3.1.1. Participants

Twenty male adults (age 21–24 years) were classified as right-handed according to Edinburgh Inventory (Oldfield, 1971), and divided in two groups. The first group of 12 individuals and the second group of 8 individuals were assigned to experiments 3 and 4, respectively. All participants were naïve to the purpose of the study.

3.1.2. Apparatus and stimuli

Participants sat on a chair in front of a table. The stimuli (words, or symbolic gestures, or both) were the three communication signals: CIAO (/kəʊ/), NO (/nɔ/), and STOP (/stɔp/). The modality of stimulus presentation of the communication signals varied in each of four successive blocks of trials. The first modality (1) was the word printed on a PC display. The modality 2 was a short video-clip showing the half body of an actress (face: 3 × 4.5 degrees of visual angle) pronouncing the words, or executing the symbolic gesture (modality 3), or simultaneously pronouncing the word and executing the gesture (modality 4). In experiment 4 the three communication signals were presented in three successive blocks of trials according to the following three modalities of stimulus presentation: a video-clip showing the actress (modality 1) pronouncing only the word or (modality 2) simultane-
ously executing the gesture and pronouncing the word with the voice recorded when pronouncing only the word. In this condition we synchronized the actress’s voice with her labial movements using the FINAL CUT PRO software (Apple, California, USA). The last modality of stimulus presentation was necessary because in experiment 3 the actress’s voice parameters of words pronounced simultaneously with gesture execution differed from those when the only word was pronounced (see below). Consequently, possible changes in the participant’s voice could be due to imitation of the actress’s voice. The printed words and the video-clips were presented on the same monitor with the same settings as in experiments 1 and 2. The distance of the monitor from the participant was 57 cm. The trial started with a BEEP followed by a black screen (300 ms), after which the stimulus was presented. The stimulus duration was 3000 ms.

3.1.3. Procedure

The participants were required to pronounce the word corresponding to the communication signal presented with different modalities (i.e. printed word, spoken word, meaningful gesture, and spoken word and meaningful gesture) at the end of stimulus presentation. Consequently, experiments 3 and 4 consisted of four and three experimental conditions, respectively, in which the same response was required to the stimuli presented, respectively, in four and three successive blocks of trials. The blocks of trials were quasi-counterbalanced across
Table 4

<table>
<thead>
<tr>
<th></th>
<th>Pitch (Hz)</th>
<th>Intensity (dB)</th>
<th>F1 (Hz)</th>
<th>F2 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIAO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>210.5</td>
<td>61.7</td>
<td>845.7</td>
<td>1392.7</td>
</tr>
<tr>
<td>Word and gesture</td>
<td>219.1</td>
<td>59.2</td>
<td>901.1</td>
<td>1422.9</td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>204.3</td>
<td>62.9</td>
<td>964.9</td>
<td>1388.8</td>
</tr>
<tr>
<td>Word and gesture</td>
<td>206.3</td>
<td>60.6</td>
<td>967.4</td>
<td>1387.1</td>
</tr>
<tr>
<td>STOP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>240.8</td>
<td>60.5</td>
<td>749.0</td>
<td>1385.3</td>
</tr>
<tr>
<td>Word and gesture</td>
<td>224.6</td>
<td>59.9</td>
<td>809.6</td>
<td>1422.5</td>
</tr>
</tbody>
</table>

participants. Each stimulus was quasi-randomly presented five times. Consequently, each block of experiments 3 and 4 consisted of 15 trials. Participants were left free to use as much time as needed to complete each trial.

3.1.4. Data recording

The participants wore the same microphone and the data were acquired and analyzed using the same technical equipment as in experiments 1 and 2. The time course of formant (F) 1 and 2, pitch, intensity and duration of vowels were analyzed. We measured also the actress’s voice parameters. They are presented in Table 4.

3.1.5. Data analysis

In experiments 3 and 4, the voice parameters were analyzed by ANOVAs in which the within-subjects factors were as follows: Condition (experiment 3: printed word versus spoken word versus gesture versus gesture and spoken word; experiment 4: spoken word versus gesture and original spoken word versus gesture and non-original spoken word) and communication signal (CIAO versus NO versus STOP). In all analyses paired comparisons were performed using the Newman–Keuls procedure. The significance level was fixed at $P < 0.05$.

Since more parameters were statistically analyzed in experiments 1 and 2 than in experiments 3 and 4, more statistical power was required. This explains the greater sample of participants in experiments 1 and 2.

3.2. Results

3.2.1. Experiment 3

Fig. 8 shows mean F2 of the word vowels recorded in the four conditions. In the ANOVA, the factor condition was significant ($F(3,21) = 9.0, P < 0.0005$). F2 significantly increased in the conditions of spoken word and gesture compared with printed word. Further, it significantly increased in the condition of simultaneous spoken word and gesture in the comparison with all the other conditions. No other voice parameter reached significance.

3.2.2. Experiment 4

All the participants except one reported that they were unaware that the voice differed between the two conditions of bimodal stimulus presentation. Fig. 8 shows that F2 of the word significantly increased in the two conditions of bimodal presentation as compared with the condition of unimodal presentation ($F(2,22) = 4.7, P < 0.05$). No other significance was found in the ANOVAs.

4. Discussion

4.1. General characteristics of word voice spectra and gesture kinematics

Some parameters of the vowel spectra and arm kinematics of the three words and gestures analyzed in the present study may be related to aspects of the meaning of the communication signals. Both F1 and F2 of the word CIAO were higher than those of the word STOP. In other words, the vowels of the word CIAO were opened and anterior as compared to the vowel of the word STOP. F2 of the word NO was intermediate between the other two words. Although the sample of words tested in the present study is small, we observed that the type of interaction coded by the meaning of the words could be related to the mouth configuration during vowel pronunciation. When pronouncing the word CIAO, which involves closer interactions, the anterior internal mouth closure and the tongue is displaced forward (Leoni & Maturi, 2002). The reverse, i.e. anterior internal mouth closure and tongue retraction occurs when pronouncing the word STOP, which codes avoidance. This is accordance with the obser-
vication that in non-humans, both mouth aperture and tongue protrusion are typical of approaching relationships. For an example lips-smacking and protruding lips precede grooming actions among monkeys (Van Hooff, 1962, 1967). The old theory by Paget (1930), called “schematopoeia” proposed a parallel between sound and meaning in a variety of languages. According to Paget, vowels are opened in words coding something that is large, whereas vowels are closed in words coding something that is small. It is possible that also the type of interaction coded by the word influences the degree of aperture and anteriority of the vowel, i.e. opened and anterior vowel in some words favoring interaction, whereas closing and posterior vowel for words coding avoidance. In addition to formants, also intensity and pitch were different according to word meaning. The two parameters increased when pronouncing STOP. Since this word expresses a command, it obviously requires higher intensity and pitch. This was automatically performed by the participants who nevertheless were not required to emphasize the voice tone according to its meaning during pronunciation of the words.

The three gestures also differed in the arm kinematics. The maximal arm height of the NO gesture was lower than those of the CIAO and STOP gestures. A possible explanation of this result is that the fulcrum of the three gestures was probably placed at the same height (Fig. 5), to send communication signals from the same spatial location in order to facilitate receipt of the signal. The different amplitude, velocity and number of hand oscillations during the CIAO, and NO executions may be explained by both the type of interaction coded by and the emotional state related to the hand gestures. CIAO, which favors interaction (and, consequently, verbal communication) and is frequently related to happiness induces increasing oscillations and velocity of the gesturing hand in order to attract the observer. NO, which is more related to avoidance and to negative emotions has an opposite effect. Summing up, some features of voice spectra and hand kinematics may reflect the social intention coded by the two communication signals.

4.2. Mutual interactions between word and gesture emission

Word and corresponding-in-meaning symbolic gesture influenced each other when they were emitted simultaneously. F2 in the voice spectra was higher when the word was pronounced together with the gesture. This effect might be due to vocalization simultaneous to arm movement. Indeed, it is well known that any movement of body effector can modify voice emission, due to proprioceptive stimulation. However, no modification in F2 was observed when executing a meaningless arm movement, which involved the same joints as the three meaningful gestures. Conversely, F2 of a pseudo-word was not affected by the meaningful gestures. Thus, it is more plausible that the meaning of the gesture influenced F2 of the words. Higher F2 is related to forward displacement of the tongue (Ferrero, Genre, Boe, & Contini, 1979; Leoni & Maturi, 2002). Consequently, the command to execute the gesture was probably sent also to the tongue. The finding that tongue protrusion instead of retraction was observed in response to the double command to mouth and arm may be explained as a non-specific effect, i.e. the command to move forward both arm and tongue. However, another explanation is that if the fact that the effect on voice was effective only for meaningful gestures may be that, independently of the word meaning, mouth articulation might be configured (i.e. the tongue was protruded) by the speaker in the expectancy of a response. In other words, this effect expresses the intention to interact closely with other individuals (see above). This explanation is supported by the results of experiments 3 and 4 showing that the verbal response to a visible individual pronouncing the word and simultaneously performing the gesture showed the same increase in F2 as during simultaneous execution of the two communication signals. F1 of the word STOP increased when pronounced simultaneously with both the meaningful and the meaningless gestures. The effect observed only on the word STOP can be explained by the initial word consonants, which require quicker release of mouth occlusion. This was enhanced by the arm lifting producing greater opening of the internal mouth. Indeed, F1 mainly depends on anterior mouth opening. The effect on F1 was also found when the meaningless gesture was executed. Both the hand configuration and movement of the meaningless gesture were similar to that of the STOP gesture. This may explain the effect of the meaningless gesture on the word STOP. F1 is mainly related to anterior mouth aperture, whereas F2 is mainly related to tongue protrusion (Ferrero et al., 1979). Gesture execution enhanced both the two mouth movements, and consequently voice parameters. In this regard, gesture reinforced word.

The kinematics of the gestures was influenced by the simultaneous pronunciation of words of the same meaning. The general effect was a shortening of the various movement phases. This result can be due to overlap between the cerebral mechanisms associated with speech and manual activity. Indeed, Kinsbourne and Cook (1971) reported that the concurrent speech and manual performance disrupted the latter. Another possibility is the trend to execute gestures in the same time as duration of word pronunciation. However, these interpretations do not explain the lack of effect of the pseudo-word on gestures. Another possibility is that the amount of the information was reduced (and consequently, duration of the various gesture phases) because it was partially supplied by the word information of briefer duration. However, the word effect was not highly specific for meaningful gestures like was the effect of meaningful gestures on speech. Indeed, the number of velocity peaks of the meaningless gesture was reduced by pronunciation of words. Words, such as CIAO and NO, may have an effect even on oscillatory movements resembling the corresponding meaningful gestures. In other words, speakers can readily associate the meaning
of words with the simultaneously executed arm movements (Gentilucci, Benussi, Bertolani, Daprati, & Gangitano, 2000; Gentilucci & Gangitano, 1998). Summing up, the fact that the word information of briefer duration reduced the communication time of the gesture information may be interpreted as an inhibition effect.

The comparison of the reciprocal effects between words pronunciation and gesture execution showed that gesture reinforced speech, whereas word inhibited gesture. Gestures are usually performed in the presence of an interlocutor, whereas words can be used to communicate also with invisible or distant interlocutors. Consequently, gesture reinforces speech in order to attract the attention of a hypothetical interlocutor and to engage immediately an interaction with him. On the other hand, speech shortens the execution of the corresponding gesture probably acting by a negative feedback once the word of shorter duration has been emitted.

Although the beginning of voice emission always followed the beginning of gesture, a different temporal relationship was observed for words emitted with meaningful gestures in the comparison with both pseudo-words emitted with meaningful gestures and words emitted with meaningless gestures. A specific temporal coupling probably exists between meaningful gesture and the corresponding word, which is lost when the gesture or the string of phonemes are meaningless. This further supports the idea of a single system, which governs the temporal aspects of arm and mouth kinematics (Levelt et al., 1985).

4.4. Conclusion

The data of the present study suggest that two types of communication signal, word and gesture, are related at the levels of execution and processing. We propose that they are controlled by a single system of communication. Specifically, words spoken while executing a gesture, may become one unitary with the aforesaid gesture. Thus, from a social point of view, they take on a different meaning from when coded in isolation. We have no experimental data to show which cortical region is involved in joining gestures to words. However, neuroimaging data suggest that Broca’s area may be involved in this function. Indeed, this area seems to be involved in encoding phonological representations in terms of mouth articulation gestures (Paulesu, Frith, & Frackowiak, 1993; Zatorre, Evans, Meyer, & Gjedde, 1992) and imitation of arm gestures (Tanaka & Inui, 2002). The suggestion that word and gesture participate to a unique communication system has also implication in explaining language development in terms of evolution. Indeed, various authors (Armstrong et al., 1995; Corballis, 2002; Hewes, 1973) have proposed that speech evolved from a primitive communication system based on gestures. Although the gestures analyzed in the present study were symbolic ones (i.e. apparently not iconic), their effects on speech were similar to those we found when syllables were simultaneously pronounced with observation/execution of actions directed to an object (Gentilucci, 2003a; Gentilucci et al., 2001; Gentilucci, Santunione, et al., 2004; Gentilucci, Stefanini, et al., 2004). Consequently, the system involved in speech production is related to the production of concrete actions aimed to interact with objects as well as of abstract gestures acquiring the meaning of words. A system relating actions to syllables might have evolved to a more sophisticated system relating symbolic gestures to words.

Acknowledgments

We thank Claudio Secchi for the help in carrying out experiments 1 and 2, and Nicola Bruno for the comments on the manuscript. The study was supported by grants from MIUR (Ministero dell’Istruzione, dell’Universit`a e della Ricerca) to M.G.

References


