



Language for action: Motor resonance during the processing of human and robotic voices



G. Di Cesare^{a,b}, A. Errante^b, M. Marchi^c, V. Cuccio^{d,*}

^a Istituto Italiano di Tecnologia (IIT), Department of Robotics, Brain and Cognitive Sciences (RBCS), Genova, Italy

^b University of Parma, Department of Neuroscience, via Volturno 39/E, 43100 Parma, Italy

^c University of Milan, Department of Computer Science, Via Comelico 39/41, 20135 Milan, Italy

^d University of Parma, Department of Humanities, Social Sciences and Cultural Industries, via Massimo D'Azeglio 85, 43125 Parma, Italy

ARTICLE INFO

Keywords:

Motor system
Language processing
Motor simulation
Action verbs
Abstract verbs
Robot

ABSTRACT

In this fMRI study we evaluated whether the auditory processing of action verbs pronounced by a human or a robotic voice in the imperative mood differently modulates the activation of the mirror neuron system (MNs). The study produced three results. First, the activation pattern found during listening to action verbs was very similar in both the robot and human conditions. Second, the processing of action verbs compared to abstract verbs determined the activation of the fronto-parietal circuit classically involved during the action goal understanding. Third, and most importantly, listening to action verbs compared to abstract verbs produced activation of the anterior part of the supramarginal gyrus (aSMG) regardless of the condition (human and robot) and in the absence of any object name. The supramarginal gyrus is a region considered to underpin hand-object interaction and associated to the processing of affordances. These results suggest that listening to action verbs may trigger the recruitment of motor representations characterizing affordances and action execution, coherently with the predictive nature of motor simulation that not only allows us to re-enact motor knowledge to understand others' actions but also prepares us for the actions we might need to carry out.

1. Introduction

Recent proposals in cognitive science and neuroscience claim that cognition is embodied. In this view, cognition is considered to be grounded in action and perception, upon sensory and motor brain mechanisms (Jeannerod, 1994; Jeannerod, 2006; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995; Pulvermüller & Fadiga, 2010; Rizzolatti & Sinigaglia, 2010). This embodied approach to cognition contrasts with the classical cognitivist account according to which the mind is a mechanism for manipulating abstract and amodal symbols (Fodor, 1983; Pylyshyn, 1984).

Embodied approaches have also been applied to language (Barsalou, 2008; Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005), which, in this perspective, is considered too as grounded in action-perception systems. With regard to semantic processing, it has been claimed that the comprehension of action verbs somatotopically recruits the premotor cortex (e.g. Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006) and the Mirror Neuron system (MNs) (e.g. Tettamanti et al., 2005). In particular, Tettamanti et al. (2005) investigated the brain activity while presenting sentences expressing actions performed with the mouth, the hand or the foot. Specifically, hand actions and related words were

found activated in the left precentral gyrus, the posterior intraparietal sulcus and the left posterior inferior temporal area. In contrast, leg activity has been identified in the left dorsal premotor and left intraparietal sulcus, but located more dorsally and rostrally in relation to the parietal hand activities. In turn, abstract sentences compared to action-related sentences were specifically associated with an effect in the posterior cingulate cortex. Other studies on action-related language come to similar conclusions (e.g. Aziz-Zadeh et al., 2006; Boulenger, Hauk, & Pulvermüller, 2009; Buccino et al., 2005; Glenberg & Kaschak, 2002; Glenberg et al., 2008; Hauk & Pulvermüller, 2004; Hauk, Johnsrude, & Pulvermüller, 2004; Kemmerer, Castillo, Talavage, Patterson, & Wiley, 2008; Papeo, Vallesi, Isaja, & Rumiat, 2009; Pulvermüller, 1999; Pulvermüller et al., 2005; Sato, Mengarelli, Riggio, Gallese, & Buccino, 2008; Tettamanti et al., 2005).

Findings on the activation of the motor system during the comprehension of action-related language are today copious (for reviews and critical discussions: Barsalou, 2010; Fischer & Zwaan, 2008; Glenberg, Witt, & Metcalfe, 2013; Jirak, Menz, Buccino, Borghi, & Binkofski, 2010; Pulvermüller, Moseley, Egorova, Shebani, & Boulenger, 2014). However, the specific contribution provided by this mechanism to the processing of language is still under discussion. While authors

* Corresponding author.

E-mail address: vale.cuccio@libero.it (V. Cuccio).

committed to embodied explanations of language and cognition consider the mechanism of simulation to be constitutive of the understanding of language (e.g. Barsalou, 1999; Gallese, 2008; Glenberg, 2010; Pulvermüller, 2012), authors embracing more disembodied explanations do not consider the mechanism of simulation neither a necessary nor a sufficient condition for the comprehension of language (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Kemmerer & Gonzalez-Castillo, 2010; Mahon & Caramazza, 2008; van Elk, Slors, & Bekkering, 2010). In this latter approach, motor simulation occurring during language comprehension is mainly considered as a by-product phenomenon (Mahon & Caramazza, 2008) that does not constitutively contribute to the construction of linguistic meaning (Caramazza, Anzellotti, Strnad, & Lingnau, 2014; Mahon & Caramazza, 2008).

In addition to semantic processing, motor simulation has been suggested to also contribute to speech perception. Several years ago, Alvin Liberman and colleagues proposed a theory of speech perception according to which speech sounds are understood not only as sounds, but as articulatory gestures necessary to speak (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Harris, Hoffman, & Griffith, 1957). This theory, called motor theory of speech perception, suggested the existence of a link between action and perception and pointed to a simulation process that people use to perceive other people talking. The discovery of mirror neurons (MN) in the macaque monkey brain (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) revived the hypothesis that motor structures may be concerned with perceptual processes (Rizzolatti, Cattaneo, Fabbri-Destro, & Rozzi, 2014) and gave new momentum to the motor theory of speech perception. The linking element between monkey and human is the ventral premotor area (area F5, monkey) known to be the homolog of human Broca's region involved in speech processing (Rizzolatti et al., 2014). This leads to the assumption, that homolog to F5, also Broca's region contains mirror neurons (Buccino et al., 2005). As a consequence, Broca's region is no longer regarded as a pure language area, but also as a region linking action and language (Binkofski & Buccino, 2004). As a further development of studies on the MNs, recently several studies have shown that this system is involved not only during the observation of familiar motor actions performed by human agents, but also during observation of motor actions performed by robotic agents (Chaminade et al., 2010; Cross et al., 2012; Miura et al., 2010; Shimada, 2010; Tai, Scherfler, Brooks, Sawamoto, & Castiello, 2004). In this line, Tai et al. (2004) in a Positron Emission Tomography (PET) study presented videos in which either a human or a robot arm grasped an object. They showed that observing grasping actions performed by the human elicited a significant neural response in the MNs, while the observation of the same action by the robot didn't show the same activation. However, these results have not been consistently replicated. Gazzola, Rizzolatti, Wicker, and Keysers (2007) in a fMRI study, compared videos of simple and complex movements performed either by a human or a robot to investigate the neural activation elicited by the observation of human and robotic actions and found the activation of the MNs in both the human and robotic condition.

The role and modulation of the MNs in the human-robot interaction, thus, is still under discussion and some important issues have not yet been addressed in previous studies. One of these issues is certainly related to a possible different modulation of the MNs determined by human speech processing compared to robot speech processing. With regard to this point, the aim of the present study was to assess possible differences in the simulation of articulatory gestures during speech perception in the human and the robot conditions. According to Liberman's motor theory of speech perception (Liberman et al., 1967) a difference should not be found. But previous studies on motor simulation during human-robot interaction suggest that this difference can be plausibly found. Hence, the robot condition, on the one hand, allowed us to assess whether the mechanism of simulation is sensitive to non-

human voices and, thus, it allowed us to deepen our knowledge of the boundary conditions where linguistic processing triggers motor simulation and, on the other, it provided a further testbed for Liberman's theory of speech perception.

In addition to the topic of the simulation of articulatory gestures, due to speech processing, in the human-robot interaction, in our study, we also addressed the issue of the involvement of the MNs during semantic processing when we listen to action verbs compared to abstract verbs pronounced by a robotic and a human voice. In particular, we evaluated whether the auditory processing of action verbs, presented in the imperative mood, could trigger the internal recruitment of motor representations classically involved in action execution. For this purpose, participants were asked to listen to action verbs in imperative mood (e.g., "touch!") and abstract verbs not associated with specific motor programs (e.g., "think!").

Our study produced three main results: (1) the activation pattern found during listening to action verbs was very similar in both the robot and human conditions; (2) the processing of action verbs compared to abstract verbs determined the activation of the fronto-parietal circuit (MNs), classically involved during action understanding; (3) listening to action verbs, regardless of the condition (human and robot), activated the anterior part of the supramarginal gyrus (aSMG) a region considered to underpin hand-object interaction (Caruana & Cuccio, 2015; Urban & Caruana, 2014) and associated to the processing of affordances.

2. Material and methods

2.1. Participants

Twenty-two healthy right-handed volunteers [13 females (Mean age = 25.4 yrs, SD = 3.57 yrs, range = 22–32) and 9 males (Mean age = 24.2 yrs, SD = 1.85 yrs, range = 22–27)] participated in the Experiment. All participants had normal or corrected-to-normal visual and normal hearing. They gave their written informed consent to the experimental procedure, which was approved by the Local Ethics Committee (University of Parma).

2.2. Experimental design

A sparse block design (Gazzola, Aziz-Zadeh, & Keysers, 2006; van Atteveldt, Formisano, Goebel, & Blomert, 2004) was used in the experiment. The scan cycle (TR) was composed by 37 sequential slices (slice thickness = 3 plus inter-slice gap = 0.5 mm) covering the whole brain collected in 2000 ms (acquisition time) followed by a silence period lasting 2000 ms (TR = 4000 ms). Experimental stimulus was presented during the silence period. Audio stimuli were presented in blocks of three consecutive stimuli of the same condition [Human Action Verbs, Human Abstract Verbs, Robotic Action Verbs, Robotic Abstract Verbs, Silence]. In 16% of cases, intermixed with experimental blocks, catch trial blocks were presented. During the catch trials, participants had to indicate the category of the last presented stimulus by pressing a bottom (human voice, robotic voice). An inter block period of 3 TR without audio stimuli were present between two consecutive blocks (cleaning phase). The experiment was composed of 4 functional runs with a total of 20 blocks (60 single trials) for each condition presented in a randomized order. Each functional run lasted about 9 min. Before the experiment, participants performed a training test to assess the audio stimuli recognition.

2.3. Stimuli

Native Italian participants were presented with audio stimuli consisting in Italian action and abstract verbs. More specifically, a male actor and a female actress pronounced 4 different action verbs and 4 different abstract verbs in imperative mood [Italian action verbs:

Table 1
Lemma frequency, letters length, absolute frequency logarithm, for each verb.

Stimuli	Letter length	Absolute Freq.	Absolute Freq. Log.	Total Variance
<i>Action-related verbs</i>				
“dammi” (give)	5	3745	3.57	0.98
“prendi” (take)	6	2446	3.39	0.96
“tocca” (touch)	5	598	2.78	0.95
“strappa” (tear)	7	163	2.21	0.90
<i>Abstract verbs</i>				
“ama” (love)	3	818	2.91	0.83
“sogna” (dream)	5	251	2.40	0.90
“pensa” (think)	5	2602	3.42	0.94
“conosci” (know)	7	1362	3.13	0.96

“dammi” (give), “prendi” (take), “tocca” (touch), “strappa” (tear); Italian abstract verbs: “ama” (love), “sogna” (dream), “pensa” (think), “conosci” (know)].

In order to obtain the word frequency for each verb, a lexical database for Italian words was used (www.ge.ilc.cnr.it/lessico.php). The parameter of lemma frequency was obtained for both conditions (Action Verbs: Mean frequency = 1738, SD = 1664.2; Abstract Verbs: Mean frequency = 1258.2, SD = 1004.1). Moreover, verbs were also matched across conditions for letters length (Action Verbs: Mean length = 5.75 letters, SD = 0.95; Abstract Verbs: Mean length = 5 letters, SD = 1.63) (for details see Table 1). In 50% of cases verbs were pronounced by the female actress and the other 50% by the male actor. All verbs were pronounced in a neutral way. Additionally, for each verb, we also presented a not ecological form of audio stimuli consisting in a robotic voice pronouncing the same verbs.

All audio stimuli were recorded using a cardioid condenser microphone (RODE NT1) placed at 30 cm from the speaker and digitized with an A/D converter module with phantom power supply (M-AUDIO M-TRACK). The audio stimuli were then processed with FL Studio 11 software. The verbs pronounced by the robotic voice were obtained by a vocal synthesizer (TextAloud software) and then processed with FL Studio 11 software. The robotic action verbs were equated for loudness to the human action verbs.

A total of 16 experimental stimuli [(4 action verbs + 4 abstract verbs) × 2 actors] and 8 control stimuli (4 robotic action verbs + 4 robotic abstract verbs) were presented. Each experimental trial lasted 2000 ms.

The physical characteristics of the all presented audio stimuli were assessed using MATLAB (The Mathworks, Natick, MA). For each verb (audio stimulus), pronounced respectively by the male actor, female actress and robot, we estimate the sound wave amplitude, the sound intensity, and the frequency spectrum. Fig. 1 and Fig. 2, show physical characteristics related to “strappa” (tear) action verb and “sogna” (dream) abstract verb (see supplementary Figs. S1–S6 for the other action and abstract verbs).

2.4. Paradigm and task

Participants lay in the scanner in a dimly lit environment. The stimuli were presented via digital audio system with 30 dB noise-attenuating headset with 40 Hz to 40 kHz frequency response (VisuaSTIM). The software E-Prime 2 Professional (Psychology Software Tools, Inc., Pittsburgh, USA, <http://www.pstnet.com>) was used both for stimuli presentation and the recording of participants' answers. Before the experiment, participants performed a training test showing that they were able to recognize the verbs (mean accuracy = 98%).

During stimuli presentation, participants were requested to fixate a white cross on a black screen and passively listen the audio stimuli. In the 16% of cases, the participants were asked to respond to a question

related to the category of the stimuli (“Have you listened a human voice, or a robotic voice?”). The catch trials was randomly presented and lasted 2000 ms.

2.5. fMRI data acquisition

Anatomical T1-weighted and functional T2*-weighted MR images were acquired with a 3 Tesla General Electrics scanner equipped with an 8-channel receiver head-coil. Functional images were acquired using a T2*-weighted gradient-echo, echo-planar (EPI) pulse sequence acceleration factor 2, 37 sequential transverse slices (slice thickness = 3 plus inter-slice gap = 0.5 mm) covering the whole brain, with a TR time of 4000 ms (TE = 30 ms, flip-angle = 90°, FOV = 205 × 205 mm², in-plane resolution 2.5 × 2.5 mm²). The scanning sequence comprised 165 ascending sequential volumes. Additionally, a T1 weighted structural image was acquired for each participant (acceleration factor 2, 156 sagittal slices, matrix 256 × 256, isotropic resolution 1 × 1 × 1 mm³, TI = 450 ms, TR = 8100 ms, TE = 3.2 ms, flip angle 12°).

2.6. Statistical analysis

Data analysis was performed with SPM8 (Statistical Parametric Mapping software; The Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk>) running on MATLAB R2013 (The Mathworks, Inc., Natick, MA). The first three volumes of each run were discarded to allow T1 equilibration effects. For each participant, all volumes were spatially realigned to the first volume of the first session and un-warped to correct for between-scan motion, and a mean image from the realigned volumes was created. Time slicing correction was applied accordingly to sparse imaging acquisition (acquisition time, TA = 2000 ms). All volumes were normalized to SPM EPI template (MNI space) and spatially smoothed with a 6 mm full-width half-maximum isotropic Gaussian kernel for the group analysis.

Data were analyzed using a random-effects model (Friston, Holmes, & Worsley, 1999), implemented in a two-level procedure. In the first level, single-subject fMRI BOLD signal was modeled in a General Linear Model (GLM) by a design-matrix comprising the onsets, the durations of each event according to the experimental task for each functional run. The model was composed by six regressors as follows: *Human Action Verbs (Hum_Action)*, *Human Abstract Verbs (Hum_Abstract)*, *Robotic Action Verbs (Rob_Action)*, *Robotic Abstract Verbs (Rob_Abstract)*, *Silence and Response*.

Within each block, the audio stimuli were presented in blocks of the same audio stimuli of the same condition (Human Action Verbs; Human Abstract Verbs; Robotic Action Verbs; Robotic Abstract Verbs). Within each block, the audio stimuli were modeled as a single event lasting 2000 ms. The silence and response were also modeled as single event lasting 2000 ms.

In the second level analysis (group-analysis), corresponding contrast images of the first level for each participant were entered into a flexible ANOVA with sphericity-correction for repeated measures (Friston et al., 2002). This model was composed of 4 regressors by contrasting the activation pattern obtained for each condition with Silence (*Hum_Action vs. Silence*, *Hum_Abstract vs. Silence*, *Rob_Action vs. Silence*, *Rob_Abstract vs. Silence*; see Fig. 3, $P_{FWE} < 0.05$ corrected at voxel level). Within this model, we also assessed the activations resulting from the direct contrasts between conditions (*Hum_Action vs. Hum_Abstract*, *Rob_Action vs. Rob_Abstract*, *Hum_Action vs. Rob_Action*, *Hum_Abstract vs. Rob_Abstract*, and all reverse contrasts; see Fig. 4, $P_{FWE} < 0.05$ corrected at cluster level). The location of the activation foci was determined in the stereotaxic space of MNI coordinates system. Those cerebral regions for which maps are provided were also localized with reference to cytoarchitectonical probabilistic maps of the human brain, using the SPM-Anatomy toolbox v 1.7 (Eickhoff et al., 2005).

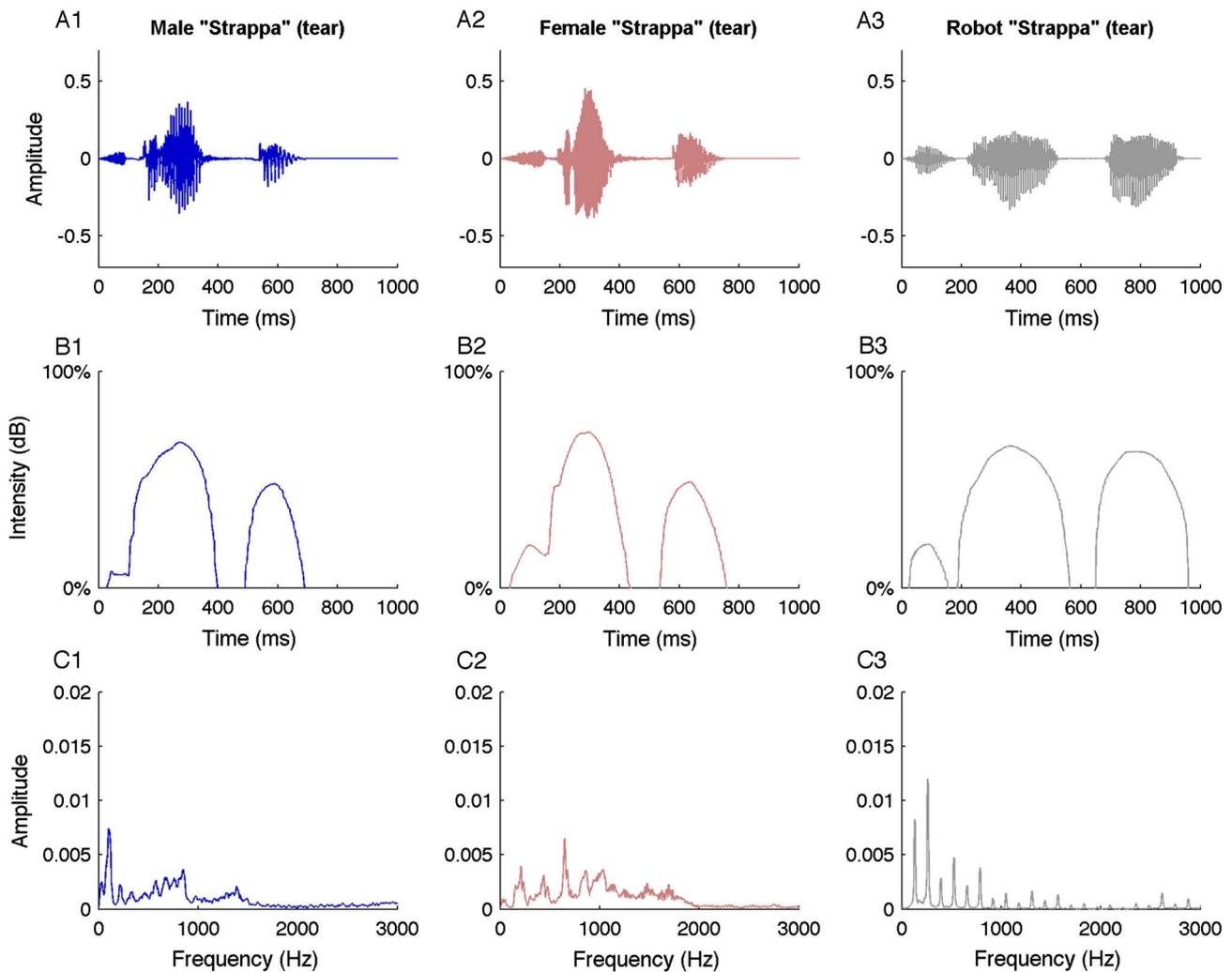


Fig. 1. Physical characteristics for the verb “strappa” (tear). The graph A shows the audio waves amplitude for all categories [A1. male (blue color), A2. female (pink color), A3. robot (grey color)]. The graph B shows the related intensity (mobile average 100 ms of the power of signal). Finally the graph C shows the frequency spectrum (magnitude of the Discret FFT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To assess possible differences present between the Human and Robot condition, three regions of interest (ROIs) were created respectively in the left inferior frontal gyrus (Brodmann 45 and 44) and left anterior supramarginal gyrus (MarsBaR ROI Toolbox for SPM, release 0.42). The ROIs were defined centering a sphere (radius 2.5 mm) in area BA 45 (ROI 1: $x = -44$, $y = -34$, $z = 10$), area BA 44 (ROI 2: $x = -50$, $y = 8$, $z = 16$) and in the aSMG (ROI 3: $x = -62$, $y = -28$, $z = 34$) respectively. Then, for each subject the signal change was extracted using the SPM Rex Toolbox (<http://web.mit.edu/swg/rex>). The Rex tool permitted to extract the signal change values in the ROIs for each participant on the basis of contrast images (second-level analysis: *Hum_Action vs. Silence*, *Hum_Abstract vs. Silence*, *Rob_Action vs. Silence* and *Rob_Abstract vs. Silence*) based on the previous flexible ANOVA model.

3. Results

3.1. Overall effect of listening to verbs pronounced by human and robot voices

The hearing of the human action verbs revealed a signal increase in auditory areas of the superior temporal gyrus, left inferior parietal lobe, left dorsal and ventral premotor cortex, left prefrontal cortex and posterior part of the left inferior frontal gyrus (Fig. 3A). A very similar

activation pattern was observed for the Robot condition (Fig. 3B, for statistical and coordinates see Table 2A and C). The hearing of the abstract verbs in the Human condition revealed a signal increase in auditory areas of the superior temporal gyrus and middle temporal gyrus bilaterally, left supramarginal gyrus, left ventral premotor cortex, left inferior frontal gyrus (BA 44, 45 Fig. 3A). The same activation pattern was observed during the hearing of the abstract verbs pronounced by the robotic voice (Fig. 3B, for statistical and coordinates see Table 2B and D).

3.2. Contrasts between action vs. Abstract verbs

Fig. 4A shows enhanced activations resulting from the direct contrast *Human Action vs. Human Abstract verbs* in the left superior and inferior parietal lobe (SPL, IPL), left supramarginal gyrus (SMG), left middle posterior temporal gyrus (MTG), left inferior frontal gyrus (BA 44, 45; IFG), left dorsal premotor cortex (PMd, for statistical and coordinates see Table 3A). A similar activation pattern was observed for the contrast *Robotic Action vs. Robotic Abstract verbs*, except for the lack of the activation in the left inferior frontal gyrus (for statistical and coordinates see Table 3B). The opposite contrasts revealed no significant activation patterns.

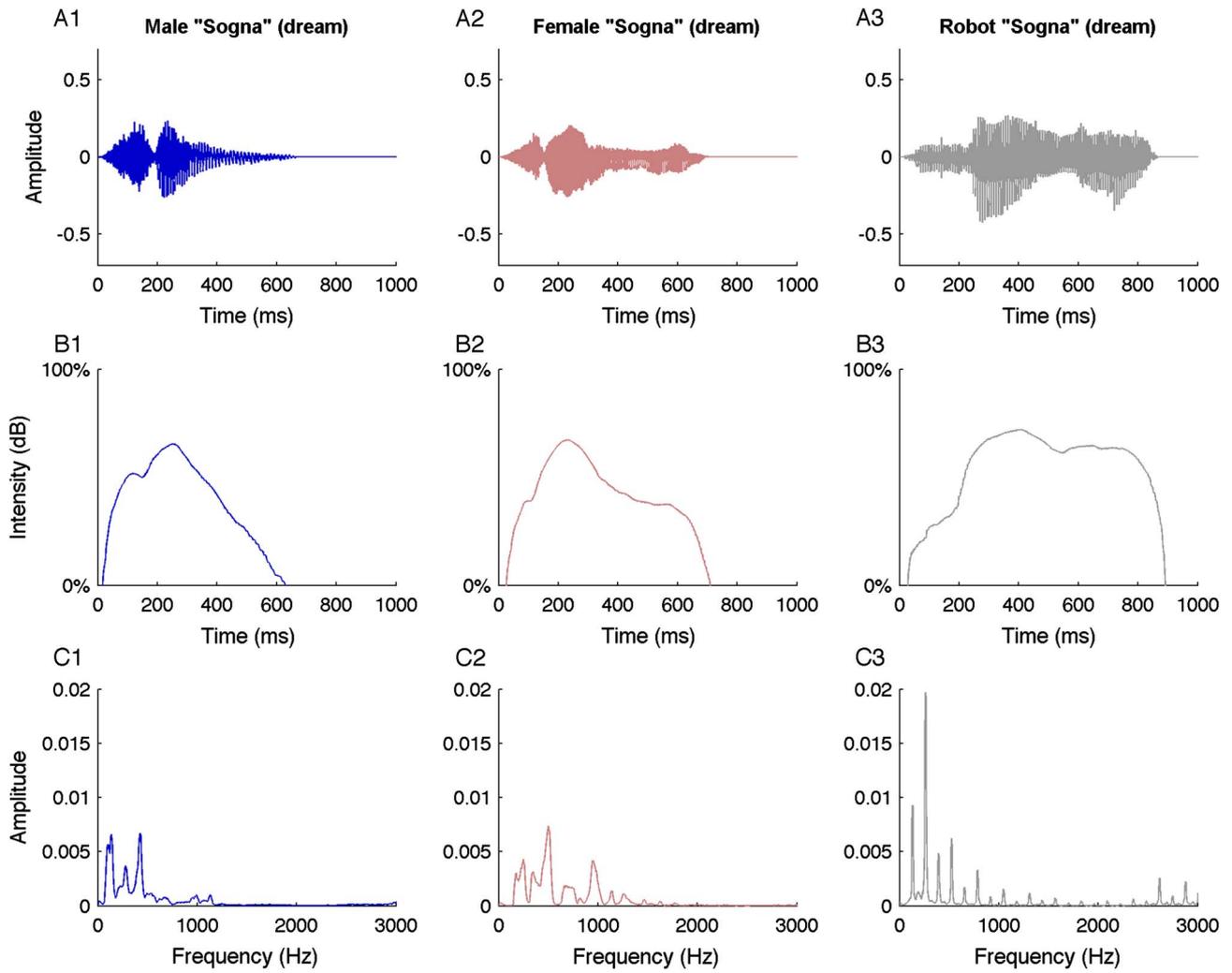


Fig. 2. Physical characteristics for the verb “sogna” (dream). The graph A shows the audio waves amplitude for all categories [A1. male (blue color), A2. female (pink color), A3. robot (grey color)]. The graph B shows the related intensity (mobile average 100 ms of the power of signal). Finally the graph C shows the frequency spectrum (magnitude of the Discret FFT). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

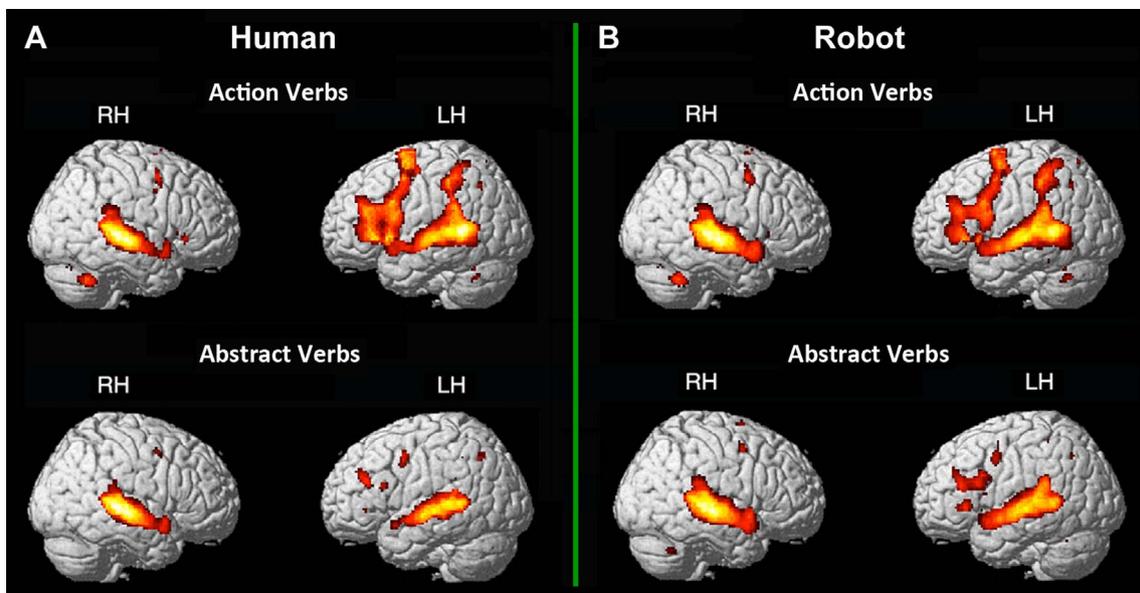


Fig. 3. Brain activations obtained in Human (A) and Robot (B) conditions. These activations are rendered into a standard Montreal Neurological Institute brain template ($P_{FWE} < 0.05$ at voxel level). RH, right hemisphere; LH, left hemisphere.

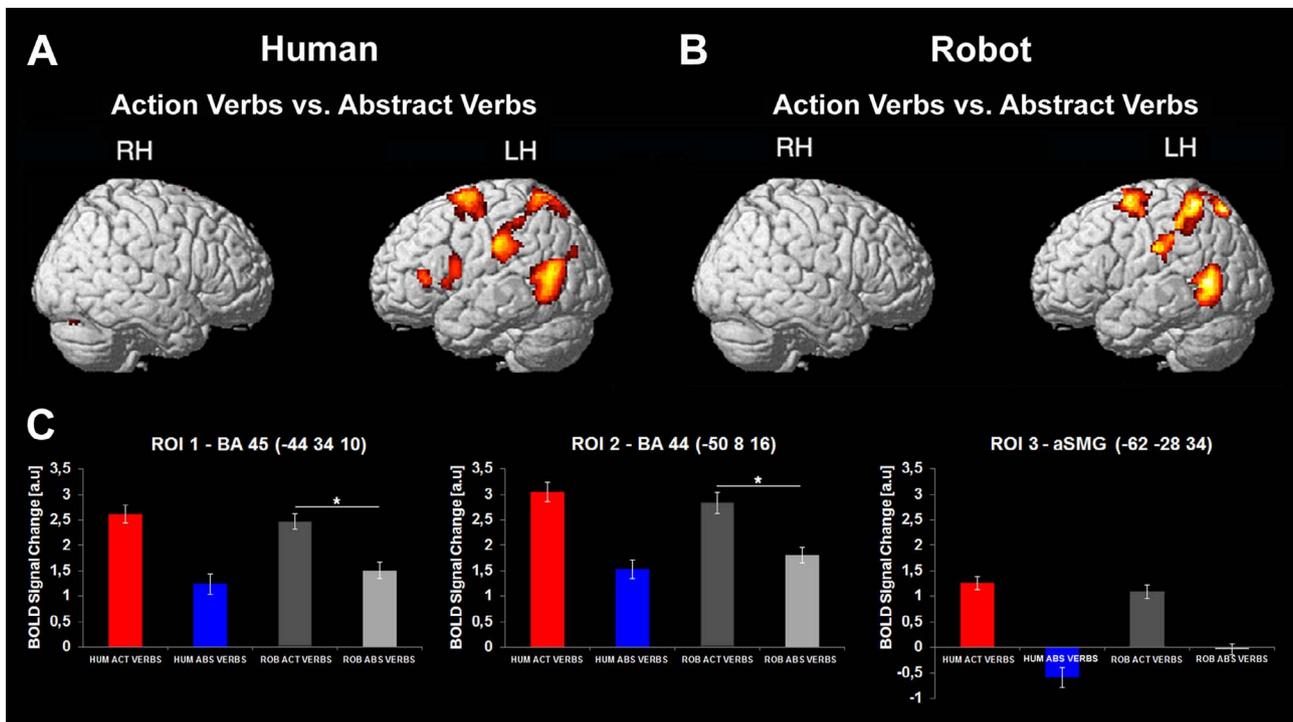


Fig. 4. Brain activations selective for action verbs processing. (A) These activations resulting from the direct contrast *Hum_Action* vs. *Hum_Abstract verbs* and *Rob_Action* vs. *Rob_Abstract verbs*. These activations are rendered into a standard Montreal Neurological Institute brain template ($P_{FWE} < 0.05$ at cluster level). RH, right hemisphere; LH, left hemisphere. (B) Bold Signal Changes extracted ROI 1 (BA 45), ROI 2 (BA 44) and ROI 3 (aSMG) resulting from the following contrast images: *Hum_Action* vs. *Silence*, *Hum_Abstract* vs. *Silence*, *Rob_Action* vs. *Silence* and *Rob_Abstract* vs. *Silence*. The horizontal line above the columns indicate the comparisons among *Robotic Action Verbs* vs. *Robotic Abstract Verbs*. The bars indicate the standard error of the mean (SEM). Asterisks indicate significant differences set at $p < 0.05$ (*).

3.3. Contrasts between human vs. robot

The differences present in Human and Robot conditions was highlighted only in the contrast *Robotic Action* vs. *Human Action* producing a larger activation in the superior temporal gyrus bilaterally (for statistical and coordinates see Table 3C).

3.4. Testing for processing of action verbs: ROIs analysis

To better understand whether, the activation obtained in the left inferior frontal gyrus (BA 45 and 44) during the contrast *Hum_Action* vs. *Hum_Abstract verbs* was only specific for the human condition, we analyzed the BOLD signal change in ROI 1 (ROI 1: $x = -44$, $y = -34$, $z = 10$) and ROI 2 (ROI 2: $x = -50$, $y = 8$, $z = 16$) obtained during listening to action and abstract verbs in both human and robot conditions (see Fig. 4B). *T*-test revealed that as observed in the human condition, also for the robot condition there was a significant difference in the BOLD signal change between action verbs and abstract verbs in both ROI 1 and ROI 2 ($P < 0.05$).

To better clarify the role of the left aSMG in the processing of action and abstract verbs, we analyzed the BOLD signal change in ROI 3. This analysis demonstrated that, regardless of the human and robot conditions, listening to action verbs produced enhanced activation of the aSMG, while listening to abstract verbs produced a deactivation of the same area (Fig. 4B, right side).

Note that, in all ROIs analyses, to avoid the problem of circularity (Kriegeskorte et al., 2009) no statistical analysis was carried out for the meaningful comparison (ROI 1 and 2: *hum act verbs* vs. *hum abs verbs*; ROI 3: *Hum_Action* vs. *Hum_Abstract verbs*, *Rob_Action* vs. *Rob_Abstract verbs*) already showed in Fig. 4A.

4. Discussion

The current fMRI study was designed to explore possible differences

in the modulation of the mechanism of simulation determined by the processing of human and robot speech. In particular, the first aim of the present study was to assess a possible difference in the simulation of articulatory gestures during speech perception in the human and the robot condition. To this purpose, we asked participants to listen to action verbs and abstract verbs, pronounced in the imperative mood by a human and a robotic voice. The results showed a very similar activation pattern during listening to action and abstract verbs in the two conditions. However, a difference between the robot and the human condition was observed in the superior/middle temporal gyrus (STG, MTG), typically associated with the processing of biological stimuli. Indeed, although these regions are involved in the processing of biological stimuli, their activations may represent the encoding of different acoustical frequencies, characterizing the robotic voice, in the ventral language pathway.

A second important issue addressed in this study concerns the possible involvement of the MNs during semantic processing when we listen to action verbs compared to abstract verbs pronounced by both a robotic and a human voice. The results showed that the processing of action verbs compared to abstract verbs pronounced by a human and a robotic voice produced the activation of the fronto-parietal circuit (MNs), classically involved in actions understanding (Rizzolatti et al., 2014). The activation of the MNs have been also found in previous studies in the processing of action-related sentences in comparison with their abstract control condition (Aziz-Zadeh et al., 2006; Tettamanti et al., 2005).

Another interesting result highlighted in the present study is the activation of the anterior part of SMG, classically involved in both object-use observation (Peeters et al., 2009) and object-manipulation (Brandi, Wohlschläger, Sorg, & Hermsdörfer, 2014), by using auditory stimuli. These results will be separately discussed in the rest of this section.

Table 2

Cerebral activity obtained during the hearing of **A. Human Action Verbs vs. Silence**; **B. Human Abstract Verbs vs. Silence**; **C. Robotic Action Verbs vs. Silence**; **D. Robotic Abstract Verbs vs. Silence**. Local maxima, as shown in Fig. 3, are given in MNI standard brain coordinates at voxel-level $P_{FWE} < 0.05$ [ATB: most probable anatomical region in the Anatomy Toolbox 1.7, Eickhoff et al., 2005].

Anatomical region	Left Hemisphere					Right Hemisphere				
	x	y	z	Z-score	ATB	x	y	z	Z-score	ATB
Stimulus Type vs. Silence (PFWE-COR VXL LEVEL = < 0.05)										
<i>A. Human Action vs. Silence</i>										
Middle Temporal Gyrus	–56	–48	6	Inf		50	–38	4	Inf	
Superior temporal Gyrus						58	–22	–2	Inf	
Inferior Parietal Lobule	–32	–62	42	5.26	61% IPS					
Superior Parietal Lobule	–16	–68	64	4.75	66% SPL					
IFG (Pars Triangularis)	–48	34	2	6.56						
Precentral Gyrus	–36	–2	64	6.41		48	4	50	5.18	
Insula						36	22	2	5.12	
Cerebellum	–30	–58	–30	6.16	79% Lobule VI	48	–60	–36	6.51	94% Lobule VIIa
Putamen	–22	8	–2	5.53						
<i>B. Human Abstract vs. Silence</i>										
Middle Temporal Gyrus	–64	–30	6	Inf		50	–38	4	Inf	
Superior temporal Gyrus						60	–20	–2	Inf	
SupraMarginal Gyrus	–52	–42	26	4.71	48% IPL					
Posterior-Medial Frontal Gyrus	–6	10	54	6.09		6	12	54	5.90	
Middle Frontal Gyrus	–42	38	30	5.68						
Precentral Gyrus	–40	0	44	5.45		50	4	48	4.84	
IFG (Pars Triangularis)	–48	18	20	5.06						
Insula						34	22	2	4.70	
<i>C. Robotic Action vs. Silence</i>										
Middle Temporal Gyrus	–64	–32	6	Inf		62	–20	0	Inf	
Inferior Temporal Gyrus	–52	–44	–20	4.75						
Inferior Parietal Lobule	–32	–62	42	5.25	61% IPS					
Superior Parietal Lobule	–16	–68	64	4.94	66% Area 7A					
IFG (Pars Triangularis)	–48	34	2	6.37						
Precentral Gyrus	–4	2	64	6.55		50	4	50	5.45	
Insula						34	22	0	5.03	
Cerebellum	–30	–58	–30	6.08	79% Lobule VI	38	–56	–34	6.27	Lobule VIIa
Putamen	–20	10	2	4.77						
<i>D. Robotic Abstract vs. Silence</i>										
Middle Temporal Gyrus	–64	–32	6	Inf	42% Area TE 3	50	–38	4	Inf	
Superior temporal Gyrus	–52	–2	–8	Inf		62	–20	–2	Inf	
Posterior-Medial Frontal Gyrus	–6	10	54	6.50		2	8	60	6.29	
Superior Parietal Lobule	–32	–62	44	4.93	61% IPS					
Inferior Parietal Lobule	–48	–42	56	4.78						
Precentral Gyrus	–38	2	44	5.43		54	2	50	5.04	
IFG (pars Triangularis)	–40	10	24	5.82						
IFG (pars Opercularis)	–56	12	12	4.71	64% Area 44					
Cerebellum	–30	–58	–30	5.50	79% Lobule VI	36	–56	–36	5.24	94% Lobule VIIa

4.1. Auditory processing of verbs pronounced by human and robotic agents

Our findings showed that the processing of action and abstract verbs, regardless of the agent (human and robot speakers) produced the activation of the left inferior frontal gyrus (IFG, Fig. 3). Previous studies reported the activation of the IFG during human speech processing (Bookheimer, 2002; Galantucci, Fowler, & Turvey, 2006; Pulvermüller et al., 2006). This finding has been interpreted as to show that articulatory features are accessed in speech perception and, thus, to support the hypothesis that speech perception determines a motor simulation of articulatory gestures. Our results supports this view showing that speech processing determines motor simulation even when we process robotic speaker's voice. This finding certainly deepens our knowledge of the possible boundary conditions where the mechanism of simulation is activated. However, the activation pattern showed in the opposite contrast, robot vs. human condition, revealed the role of the superior/middle temporal gyrus (STG, MTG). These regions are classically involved in the processing of biological stimuli. These activations may represent the processing of different acoustical frequencies, characterizing the robotic voice, in the ventral language pathway. Indeed, being the robot voice characterized by non-biological frequencies, it is plausible that the enhanced activation observed in the language ventral

stream could be dependent on the participants being less familiar with the frequencies of the robotic voice. The activation of this areas likely reflects the detection, by participants, of a non-biological stimulus.

With regard to the human-robot interaction, previous studies have shown that brain regions associated with the MNs show stronger responses to human than robotic agents (Chaminade et al., 2010; Miura et al., 2010; Shimada, 2010; Tai et al., 2004) or other non-human agents (Engel, Burke, Fiehler, Bien, & Rösler, 2008). Stronger responses to humans than animals (Buccino et al., 2004) and to familiar than unfamiliar actions (Calvo Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Calvo Merino, Glaser, Grezes, Passingham, & Haggard, 2006) have also been reported. In this framework, observing actions with familiar kinematic features, which are within one's motor repertoire, results in greater MNs activation than observing less familiar actions. However, previous experiments on motor simulation in the human-robot interaction have not addressed the issue of the processing of verbs pronounced by a human voice (biological condition) with respect to the same verbs pronounced by a robotic voice (not biological condition). Our study extended previous results concerning the human-robot interaction showing that speech processing produced a motor simulation of articulatory gestures even when we process robotic speaker's voice and provided a piece of

Table 3

Cerebral activity obtained during the direct contrast of **A. Human Action vs. Human Abstract Verbs**; **B. Robotic Action vs. Robotic Abstract Verbs**; **C. Robotic Action vs. Human Action Verbs**. Local maxima, as shown in Fig. 4, are given in MNI standard brain coordinates at cluster-level $P_{FWE} < 0.05$ [ATB: most probable anatomical region in the Anatomy Toolbox 1.7, Eickhoff et al., 2005].

Anatomical region	Left Hemisphere					Right Hemisphere				
	x	y	z	Z-score	ATB	x	y	z	Z-score	ATB
Contrasts between conditions (PFWE-COR CLUSTER LEVEL = <0.05)										
<i>A. Human Action vs. Human Abstract</i>										
Middle Temporal Gyrus	−54	−62	6	5.40						
SupraMarginal Gyrus	−62	−28	34	5.44	39% Area Pft					
Precuneus	−12	−60	60	4.94	53% SPL					
Postcentral Gyrus	−34	−34	40	4.43						
Superior Frontal Gyrus	−20	2	72	4.93						
Precentral Gyrus	−26	−6	48	3.93						
IFG (Pars triangularis)	−40	32	10	4.43						
IFG (Pars opercularis)	−50	8	14	4.22						
Cerebellum						22	−74	−26	4.34	54% Lobule VII
Putamen	−20	10	−6	4.11						
<i>B. Robotic Action vs. Robotic Abstract</i>										
Inferior Temporal Gyrus	−58	−58	−6	5.93						
Middle Temporal Gyrus	−50	−60	6	5.53						
Superior Parietal Lobule	−38	−48	60	4.87	49% SPL					
Postcentral Gyrus	−46	−24	28	4.38						
Superior Frontal Gyrus	−24	−4	70	4.57						
Precentral Gyrus	−28	−8	58	4.55						
<i>C. Robotic Action vs. Human Action</i>										
Superior Temporal Gyrus	−48	0	−14	5.19		46	−8	−14	4.88	
Medial temporal Pole						50	2	−16	5.73	
Middle Temporal Gyrus	−60	−16	0	3.51	39% Area TE 3					

evidence in support of Liberman's motor theory of speech perception (Liberman et al., 1967).

4.2. Motor simulation associated with action-related verbs

In this study, participants were presented with action vs. abstract verbs, pronounced in imperative mood. Regardless of the agent (human and robot), we found a significantly different activation pattern in the contrast action verbs vs. abstract verbs, including the fronto-parietal circuit (left inferior parietal lobe, left premotor cortex). These findings are in agreement with Tettamanti et al. (2005), which described the activation of the premotor cortex during the processing of action sentences as compared to their abstract control condition. However, as Borghi (2013) has highlighted, the definition of the role played by the mechanism of simulation in language comprehension is still controversial. The topic is currently under discussion. In this debate (van Elk et al., 2010 for a review), two functions have been attributed to the mechanism of simulation. Simulation has been considered as a form of re-enactment of our past sensorimotor experiences and as a prediction of our future actions. These two functions are not mutually exclusive. According to Barsalou (2008), simulations as re-enactment produce predictions. "The perception of something familiar in the environment, body or introspection activates a simulation or situated conceptualization that contains it. Components of the simulation or situated conceptualization not yet experienced constitute predictions about events likely to occur, actions likely to be effective and introspections likely to result" (Barsalou, 2008). The predictive character of the mechanism of simulation has been widely investigated in tasks involving processing of visual objects, the observation/imitation of actions carried out with objects, the processing of linguistic descriptions of these actions and the processing of objects' names (Borghi & Riggio, 2015 for a review). For example, the processing of sentences that describe actions carried out with objects activate the neural circuits that underpin our interactions with these objects and, thus, prompts us to interact with them. As Borghi (2013) argue, the comprehension of a sentence such as "Grasp the brush" will likely lead us to prepare a power grip that would be necessary if we really had to use this object.

Interestingly, considering that no object name was explicitly mentioned in our stimuli, the processing of action verbs compared to the processing of abstract verbs, regardless of the condition (human and robot), determined the activation of the aSMG considered to underpin hand-object interaction and usually associated to the processing of affordances. This finding will be discussed in the next section.

4.3. The functional role of the anterior SMG

Interestingly, in spite of the absence of visual objects or object names in our stimuli, the auditory processing of action verbs pronounced in the imperative mood compared to abstract verbs activated the anterior part of the SMG both in human and robot conditions. The activation of aSMG is considered to underpin the hand/object interaction and is usually associated to the processing of affordances (Brandi et al., 2014; Caspers, Zilles, Laird, & Eickhoff, 2010; Peeters et al., 2009; for a review see Borghi & Riggio, 2015). In addition, the processing of action verbs determined the activation of the dorsal premotor area and the superior parietal lobule that are considered to be the neural substrate of the arm reaching movements (Di Dio et al., 2013; Filimon, Nelson, Hagler, & Sereno, 2007). Recent meta-analysis based on action observation (Caspers et al., 2010; Molenberghs, Cunnington, & Mattingley, 2012) showed that the observation of object-related hand actions is consistently associated with activations of the IFG (BA 44), PMC (BA 6), aSMG (Pft), SPL (area 7A), the PMTG and V5 bilaterally, as well as with activations in SI (BA 2) and the anterior IPS (area hIP3). Differently, these studies also reported that the observation of intransitive actions results produced activations in the temporo-occipital areas. Furthermore, in the meta-analysis of Caspers et al. (2010), contrast analysis between *observation of hand-object action vs. hand non-object action* revealed a strong activation within left IFG (BA 44), PMC (BA 6), and aSMG. The activation of parietal regions such as the IPL, the aSMG and the SPL were also found during the Motor Imagery task (MI, Jeannerod et al., 1995). MI represents a specific aspect of motor preparation. It activates areas corresponding to those that are involved in action execution except for the primary motor cortex. In this respect, recent studies (for a review, see Héту et al., 2013) revealed that the

fronto-parietal network, including the aSMG, is involved in motor acts imagination.

These findings are in line with a recent study by Di Cesare et al. (2017), which clearly demonstrated that hand-object interaction (take a bottle, touch a point, etc.) produced enhanced activations of aSMG during both the action observation and action imagination tasks. Thus, in the present study, the presence of voxels found selectively activated in the aSMG during listening to action verbs strongly suggests the involvement of this area in planning of motor acts. This activation supports the idea that listening to action verbs may trigger the internal recruitment of motor representations. Coherently with its predictive nature, motor simulation prepares us for the actions we might need to carry out and, when the verbs denote transitive actions, as in our study, it makes us ready to also interact with the objects that would likely be involved in these actions, even though these are not explicitly mentioned.

5. Conclusion

The present study employed audio stimuli consisting in Italian action and abstract verbs presented in the imperative mood and pronounced by both a human and a robotic voice. We investigated whether the auditory processing of these stimuli, when pronounced by a human or a robotic voice, differently recruits a portion of the mirror system usually involved in the simulation of articulatory gestures. No differences were detected in the two conditions (human and robotic voice). This suggests that simulation of articulatory gestures during linguistic processing takes place even though we process a not ecological robotic voice. This result deepens our knowledge of the role of the mechanism of simulation in the human-robot interaction and can certainly have implications for fields of study related to the social and clinical use of robots. For example, research on the clinical application of robots in the diagnosis and treatment of Autistic Spectrum Disorders is today significantly growing (Diehl, Schmitt, Villano, & Crowell, 2012 for a review). However, much work needs to be done. The potentiality of the clinical use of robots will be more effectively exploited only when we will have reached a deeper comprehension of the nature of the human-robot interaction. This study offers a piece of evidence in this direction. It shows the flexibility of the mechanism of simulation and its sensitivity when we process auditory not ecological linguistic stimuli.

A second conclusion that can be drawn from this study concerns the context and task dependency of the activation of motor simulation. On the basis of previous works, we can say that the activation of the motor network that underpins the hand/object interaction was determined by the effectiveness of our stimuli (imperative verbs, oral modality; see Borghi, 2013 and Borghi & Riggio, 2015 for a review) to engage participants as agents. The processing of action-related imperative verbs presented in the verbal modality prepared participants of our study to produce a reaching-manipulation action. This reflects the predictive function of the mechanism of simulation. However, it is of paramount importance to highlight that, contrary to previous studies, in our experiment neither visual objects nor object names were presented. This leads us to a third conclusion concerning the inferential nature of language (Sperber & Wilson, 1995) and the possibility that implicit meaning can modulate the activation of the mechanism of simulation (Cuccio et al., 2014). Indeed, our findings show that the processing of action related imperative verbs activates affordances and, thus, prepares us to interact with an object even though no object name is presented in our linguistic stimuli (the object is very likely inferred by the participants). To the best of our knowledge, this study is the first showing the activation of affordances by using stimuli where neither visual objects nor object names were explicitly presented.

Acknowledgments

This study was supported by the Advanced Research Grant

COGSYSTEMS to Professor Giacomo Rizzolatti. Funding to pay the publication fees for this article was provided by the Department of Robotics, Brain and Cognitive Sciences, Istituto Italiano di Tecnologia (IIT).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bandc.2017.08.001>.

References

- Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*(18), 1818–1823.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–609 discussion 610–660.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 1–21.
- Barsalou, L. W. (2010). Grounded cognition: Past, present, and future. *Topics in Cognitive Science*, *2*(4), 716–724.
- Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: The case of action verbs. *Journal of Neuroscience*, *28*(44), 11347–11353.
- Binkofski, F., & Buccino, G. (2004). Motor functions of the Broca's region. *Brain and Language*, *89*(2), 362–369.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review Neuroscience*, *25*, 151–188.
- Borghi, A. M. (2013). Action language comprehension, affordances and goals. In Yan Coello, & Angela Bartolo (Eds.), *Language and action in cognitive neuroscience. Contemporary topics in cognitive neuroscience series* (pp. 125–143). Psychology Press ISBN: 978-1-84872-082-4.
- Borghi, A. M., & Riggio, L. (2015). Stable and variable affordances are both automatic and flexible. *Front Hum Neurosci*. *19*(9), 351. <http://dx.doi.org/10.3389/fnhum.2015.00351> eCollection 2015.
- Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex*, *19*(8), 1905–1914.
- Brandi, M. L., Wohlschläger, A., Sorg, C., & Hermsdörfer, J. (2014). The neural correlates of planning and executing actual tool use. *Journal of Neuroscience*, *34*(39), 13183–13194.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., ... Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by non-conspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, *16*(1), 114–126.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action-related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Brain Research, Cognitive Brain Research*, *24*(3), 355–363.
- Calvo Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243–1249.
- Calvo Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(22), 21.
- Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied cognition and mirror neurons: A critical assessment. *Annual Review of Neuroscience*, *37*, 1–15.
- Caruana, F., & Cuccio, V. (2015). Types of abduction in tool behaviour. *Phenomenology and the Cognitive Sciences* (pp. 1–19). Springer.
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, *50*(3), 1148–1167.
- Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., ... Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE*, *5*(7), e11577.
- Cross, E. S., Liepelt, R., de C. H., Antonia, F., Parkinson, J., Ramsey, R., ... & Prinz, W. (2012). Robotic movement preferentially engages the action observation network. *Human Brain Mapping*, *33*(9), 2238–2254.
- Cuccio, V., Ferri, F., Ambrosecchia, M., Carapezza, M., Lo Piparo, F., Fogassi, L., & Gallese, V. (2014). How the context matters. Literal and figurative meaning in the embodied language paradigm. *PLoSOne*, *9*(12), e115381.
- Di Cesare, G., Marchi, M., Errante, A., Fasano, F., & Rizzolatti, G. (2017). Mirroring the social aspects of speech and actions: The role of the insula. *Cerebral Cortex*, *1*, 1–10.
- Di Dio, C., Di Cesare, G., Higuchi, S., Roberts, N., Vogt, S., & Rizzolatti, G. (2013). The neural correlates of velocity processing during the observation of a biological effector in the parietal and premotor cortex. *Neuroimage*, *64*, 425–436.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180.
- Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders*, *6*(1), 249–262.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*(4), 1325–1335.

- Engel, A., Burke, M., Fiehler, K., Bien, S., & Rösler, F. (2008). What activates the human mirror neuron system during observation of artificial movements: Bottom-up visual features or top-down intentions? *Neuropsychologia*, *46*(7), 2033–2042.
- Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: Mirror neurons for execution, observation, and imagery. *Neuroimage*, *37*(4), 1315–1328.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology (Hove)*, *61*(6), 825–850.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.
- Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage*, *16*, 484–512.
- Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *Neuroimage*, *10*, 1–5.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. Review. Erratum in: *Psychon Bull Rev*. Aug;13(4):742.
- Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, *3*, 317–333.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in premotor cortex. *Brain*, *119*, 593–609.
- Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*, 1824–1829.
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674–1684.
- Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 586–596. <http://dx.doi.org/10.1002/wcs.55>.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology (Hove)*, *61*(6), 905–919.
- Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, *8*, 573–585.
- Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.
- Hauk, O., & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201.
- Hétu, S., Grégoire, M., Saimpont, A., Coll, M. P., Eugène, F., Michon, P. E., & Jackson, P. L. (2013). The neural network of motor imagery: An ALE meta-analysis. *Neuroscience and Biobehavioral Reviews*, *37*(5), 930–949.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(02), 187–202.
- Jeannerod, M. (2006). *Motor cognition: What actions tell the self (No. 42)*. Oxford University Press.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, *18*(7), 314–320.
- Jirak, D., Menz, M. M., Buccino, G., Borghi, A. M., & Binkofski, F. (2010). Grasping language – A short story on embodiment. *Consciousness and Cognition*, *19*(3), 711–720.
- Kemmerer, D., Castillo, J. G., Talavage, T., Patterson, S., & Wiley, C. (2008). Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. *Brain & Language*, *107*, 16–43.
- Kemmerer, D., & Gonzalez-Castillo, J. (2010). The two-level theory of verb meaning: An approach to integrating the semantics of action with the mirror neuron system. *Brain and Language*, *112*(1), 54–76.
- Kriegerkorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540.
- 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., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1), 59–70.
- Miura, N., Sugiura, M., Takahashi, M., Sassa, Y., Miyamoto, A., Sato, S., ... Kawashima, R. (2010). Effect of motion smoothness on brain activity while observing a dance: An fMRI study using a humanoid robot. *Social Neuroscience*, *5*(1), 40–58.
- Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*, *36*(1), 341–349.
- Orban, G. A., & Caruana, F. (2014). The neural basis of human tool use. *Frontiers in Psychology*, *9*(5), 310. <http://dx.doi.org/10.3389/fpsyg.2014.00310> eCollection 2014.
- Papeo, L., Vallesi, A., Isaja, A., & Rumiati, R. I. (2009). Effects of TMS on different stages of motor and non-motor verb processing in the primary motor cortex. *PLoS ONE*, *4*(2), e4508.
- Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2009). The representation of tool use in humans and monkeys: Common and uniquely human features. *Journal of Neuroscience*, *29*(29), 11523–11539.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Science*, *22*(2), 253–279 discussion 280–336.
- Pulvermüller, F. (2012). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics*, *25*(5), 423–459.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360.
- Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–7870.
- Pulvermüller, F., Moseley, R. L., Egorova, N., Shebani, Z., & Boulenger, V. (2014). Motor cognition-motor semantics: Action perception theory of cognition and communication. *Neuropsychologia*, *55*, 71–84.
- Pyllyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, *94*, 655–706.
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274.
- Sato, M., Mengarelli, M., Riggio, L., Gallese, V., & Buccino, G. (2008). Task related modulation of the motor system during language processing. *Brain and Language*, *105*(2), 83–90.
- Shimada, S. (2010). Deactivation in the sensorimotor area during observation of a human agent performing robotic actions. *Brain and Cognition*, *72*, 394–399.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (Second Edition). Oxford/Cambridge: Blackwell Publishers.
- Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is “mirror” only for biological actions. *Current Biology*, *14*, 117–120.
- Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., ... Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*, 271–282.
- van Elk, M., Slors, M., & Bekkering, H. (2010). Embodied language understanding requires an enactivist paradigm of cognition. *Frontiers in Cognition*, 1–9.