Brain and Cognition 118 (2017) 118–127

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

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, & Pulvermueller, 2004; Kemmerer, Castillo, Talavage, Patterson, & Wiley, 2008; Papeo, Vallesi, Isaja, & Rumiati, 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).

http://dx.doi.org/10.1016/j.bandc.2017.08.001
Received 18 March 2016; Received in revised form 4 August 2017; Accepted 4 August 2017
Available online 19 August 2017
0278-2626/

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 Libermann 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; Orban & 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: