Computer Engineering
ISSN: 10003428

Volume 15 | Issue 3 | Year 2026
https://journalofcomputerengineering.com/

Mirroring the Social Aspects of Speech and Actions: An Exploration of the Insula's Role

Alessandro Bianchi, Elena Rossi, Lorenzo Ferrari, Francesca Zampieri, Andrea Bellazzi

Alessandro Bianchi and Elena Rossi, Department of Neuroscience, University of Modena and Emilia-Romagna, Modena, Italy; Lorenzo Ferrari, Department of Computer Science, University of Bologna, Bologna, Italy; Francesca Zampieri, Department of Psychology, University of Trento, Trento, Italy; Andrea Bellazzi, Department of Mathematics, University of Pavia, Pavia, Italy

Abstract

Action and speech may take different forms, being expressed, for example, gently or rudely. These aspects of social communication, named vitality forms, have been little studied in neuroscience. In the present functional magnetic resonance imaging study, we investigated the role of insula in processing action and speech vitality forms. In speech runs, participants were asked to listen or imaging themselves to pronounce action verbs gently or rudely. In action runs, they were asked to observe or imaging themselves to perform actions gently or rudely. The results showed that, relative to controls, there was an activation of the dorso-central insula in both tasks of speech and action runs. The insula sector specific for action vitality form was located slightly more dorsally than that of speech with a large overlap of their activations. The psycho-physiological interaction analysis showed that the insular sector involved in action vitality forms processing is connected with the left hemisphere areas controlling arm actions, whereas the sector involved in speech vitality forms processing is linked with right hemisphere areas related to speech prosody. We conclude that the central part of the insula is a key region for vitality forms processing regardless of the modality by which they are conveyed or expressed.

Key words: fMRI, insula, mirror mechanism, social interactions, vitality forms

Introduction

What individuals feel toward others can be appreciated by observing how they act and by listening to the tone of their voice. These action and speech aspects, which are fundamental for social life, have been named “vitality affects” or “vitality forms” by Stern (1985, 2010).

In a previous study, we demonstrated that the dorso-central insula is involved in encoding the vitality forms of listened action verbs (Di Cesare et al. 2016). These findings were in line with previous data showing that the insula plays an important role in encoding the vitality forms of an action performed by another individual (Di Cesare et al. 2015). Taken together, these studies strongly suggest that the dorso-central insula is a key region involved in the action and speech vitality forms processing.

There are, however, some important issues that have been not addressed in previous experiments. The first concerns the possible role of the mirror mechanism in mediating the vitality forms of speech listening. The mirror mechanism is a basic brain mechanism that transforms sensory representation of others’ behavior into one’s own motor representation of that behavior. This mechanism is based on the activity of a distinct class of neurons that discharge both when individuals perform a given motor act and when individuals observe another person performing a motor act with a similar goal. The mirror mechanism was originally discovered in the ventral premotor cortex of the macaque monkey (area F5). Subsequently, mirror mechanism has been found in humans in the premotor and parietal cortex (see Rizzolatti et al. 2014) as well as in anterior cingulate cortex (Caruana et al. 2015) and the anterior insula (Wicker et al. 2003).

As far as the possible presence of the mirror mechanism for the speech vitality forms, this mechanism could be hypothesized if the same area is active during both vitality form listening and speech production. However, because it is impossible to test participants in the scanner while they speak for the presence of artifacts due to concomitant head movements, in the present functional magnetic resonance imaging (fMRI) experiment, we asked them to imagine themselves pronouncing action verbs with different styles (gentle or rude). Note that motor imagery is considered to be equivalent to a motor preparation (Jeannerod 1995) and is therefore an appropriate way to test the presence of the mirror mechanism. An overlap of the voxels active during speech imagination, that is, during speech motor representation and speech vitality form listening would be therefore a strong argument in favor of the presence of the mirror mechanism for speech vitality forms encoding.

A second important issue to be clarified is whether there is a common region in the dorso-central insula active during “listening” and “imaging” to pronounce action verbs with different vitality forms (gentle or rude) and “observing” and “imaging” the identical actions executed with those forms. To address this issue, in the same fMRI experiment, participants were tested in speech and action sessions, and the insula activations were assessed.

The general schema of the experiment was the following. To create a socially realistic context, we presented in both listening and action observation tasks, video clips showing an actor/actress expressing vocally or gesturally different action vitality forms directed toward the participants (Fig. 1). Specifically, in the speech session, participants had to listen action verbs pronounced by the actors (listening task) and imagine themselves to pronounce the same verbs directed toward the actors in gentle or rude way (speech imagination task). In the action session, participants had to observe the actions performed by the actors (observation task) and imagine themselves to perform the same action toward the actors in gentle or rude way (action imagination task). The results showed that during vitality forms perception (auditory, visual) and during vitality forms imagination, there is a common specific activation of the dorso-central insula.

Finally, in order to assess which are the cortical areas connected with the insula in the speech and action vitality processing, we used psycho-physiological interaction (PPI) analysis. The results showed that in spite of the fact that the dorsal sector of the insula was activated by both vocal and action vitality forms, this sector is connected with 2 different cortical circuits. The one for speech is located in the right hemisphere, while that for action in the left one. The possible role of these 2 circuits in modulating action and speech cortical areas will be discussed on the basis of anatomical and functional data.

Materials and Methods

Participants

Eighteen healthy right-handed volunteers (9 females [mean age = 26.2 years, SD = 2.7 years] and 9 males [mean age = 25.1 years, SD = 2.9 years]) 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 (Parma).

Experimental Design, Stimuli, and Tasks

The experiment was organized in a mixed sparse design. The experiment consisted of a speech session and an action session. Each session was organized in a mixed event-related sparse blocked design based on a 2 × 2 factorial design with "task" (speech session: "listening" and "imaging to speak"; action session: "observation" and "imaging to perform the action") and "condition" (Vitality forms; Control) as factors. The presentation order of the runs was balanced across participants. Nine participants started with the speech session (runs 1, 2), followed by the action session (runs 3, 4) and 9 participants started with action session (runs 1, 2), followed by the speech session (runs 3, 4).

Speech Session

The aim of the speech session was to assess the possible existence of the mirror mechanism in the central insula by comparing insula activations when participants listened to and imagined to pronounce action verbs in a rude or gentle way. For this purpose, 4 conditions were presented: "vitality forms listening" (VF listening), "control listening" (Ctrl listening), "vitality forms speech imagination" (VF speech imagination), "control speech imagination" (Ctrl speech imagination). Stimuli were presented in blocks of 4 consecutive stimuli of the same condition. Each block condition was introduced by an instruction panel reminding participants the task.

In the listening task (LST), participants were presented with video clips showing an object positioned on a table (bottle, jar, or a smartphone) and an actor/actress on the background pronouncing 3 Italian actions verbs (Italian used verbs: "prendi" [take], "chiudi" [close], "tocca" [touch]) in the imperative mood using a gentle or rude vitality forms (VF listening, Fig. 1A1). In the control condition, video clips showed 3 colored objects (orange, green, and blue) positioned on the table and the actors spelled 3 different pseudo words without any vitality form: d-i-m-a, m-a-p-a, and i-r-p-a (Ctrl listening, Fig. 1A2). According to the instruction panel, "listen to vitality forms" or "listen," participants were required to pay attention to the audio vitality forms or to the spelling (Supplementary Fig. S1).

In the speech imagination task (Speech IMA), video clips showed the same scenarios with a silent actor/actress in the background. According to the screen color, participants were required to imagine to pronounce internally 3 action verbs (take, close, and touch) in a rude (red color) or gentle (blue color) way (VF speech imagination, Fig. 1A1). In the control condition, according to the object color (orange, green, and blue), participants also imagined to spell (Ctrl speech imagination, Fig. 1A2) 3 different pseudo words (d-i-m-a; m-a-p-a; and i-r-p-a) using an Italian pronunciation. A total of 12 experimental audio stimuli (3 action verbs × 2 vitality forms × 2 actors) and 6 control audio stimuli (3 spelling words × 2 actors) were presented in the speech session. Each trial lasted 2s.

Action Session

The aim of the action session was to assess whether observing and imaging to perform the same actions (take a bottle, close a jar, and touch a smartphone) performed with different vitality forms (gentle and rude) would produce the activation of the same sector of the insula. For this purpose, in the action session, as in the speech session, 4 conditions were presented: "vitality forms observation" (VF observation), "control observation" (Ctrl observation), "vitality forms action imagination" (VF action imagination), "control action imagination" (Ctrl action imagination). In the observation task (OBS), participants were presented with video clips showing an actor/actress performed 3 different actions (take a bottle, close a jar, and touch a smartphone; Fig. 1B1) in gentle or rude way. Additional video clips showed the hand of the actor/actress placing a small ball of 3 different color (green, yellow, and blue) randomly into the right or left box (control condition, Ctrl; Fig. 1B2). In all video clips, the actor started the action from a fixed initial position and reached a fixed final one (see trajectory curves, Supplementary Fig. S2). According to the instruction panel, "observe vitality forms" or "observe," participants were required to pay attention to the action vitality forms or to the action goal (Supplementary Fig. S1).

In action imagination task (Action IMA), participants were presented with video clips showing an object on the table (bottle, jar, and smartphone) and an actor/actress on the background. According to the screen color, participants were required to imagine to perform (VF Imagination; Fig. 1B1) the same observed actions in a rude (red color) or gentle way (blue color). In the control condition, video clips showed 3 differently colored balls and 2 boxes with an actor/actress on the background. The participants were required to imagine to place the balls into the left or right box without any particular vitality forms (Ctrl imagination, Fig. 1B2). A total of 12 experimental video stimuli (3 action × 2 vitality forms × 2 actors) and 12 control stimuli (1 action × 3 colors × 2 directions × 2 actors) were presented in the action session. Each trial lasted 2s.

Physical Properties of Vitality Forms

The audio stimuli were recorded using a cardioid condenser microphone (RODE NT1) placed 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 video stimuli were recording using 2 identical camcorders (Panasonic HCX 900), one positioned to the actor's right side (lateral viewpoint) and the other in front (frontal viewpoint). After stimuli recording (audio and video), the physical characteristics of all presented stimuli were analyzed using MATLAB (The Mathworks).

For the audio stimuli (Rude, Gentle, and Ctrl), we estimated the audio mean intensity power [$dE/dt = d(\text{wave amplitude})^2/dt$; Fig. 2A1]. In addition, for each stimulus, we calculated the

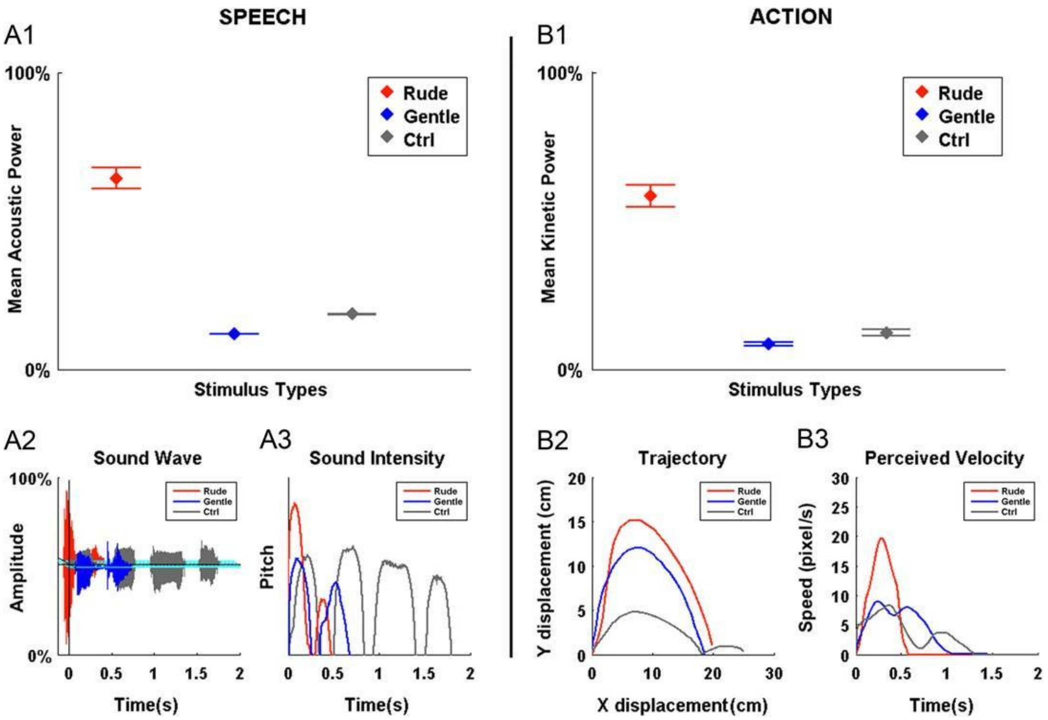

Figure 2. Physical characteristics of the audio and video stimuli. Graphs depict the mean intensity power of audio stimuli (A) and the mean kinetic power of the video stimuli (B) (rude, red line; gentle, blue line; ctrl, gray line); the bars indicate the standard error of the mean (SEM). Graphs A1−A2 show, respectively, the sound wave amplitude and the related intensity for the action verb "touch" pronounced by the male actor and relative control stimulus (spelling of the pseudo word: M-A-P-A). Graphs B1−B2 show the trajectory and velocity profiles of one action performed by the male actor (touch a smartphone) and relative control stimulus (placing a ball in a right or left box, see also Supplementary Fig. S4 for the complete stimuli description).

sound wave amplitude and the sound intensity (Fig. 2A2, A3; see also Supplementary Fig. S4).

For the video stimuli, using the 2D point kinematics, we analyzed all actors' movement profiles present in the video stimuli. For this purpose, a specific point of the actor's hand was marked for all video clips. The origin of the x/y axes was fixed in the rest position of the thumb and its position was tracked in the space every 20 ms until the end of action. Considering the hand effector as constant, for the video stimuli, we estimated the mean kinetic power [$dE_k = d(1/2\ mv^2)$] required to perform rude, gentle, and control actions (Fig. 2B1). Additionally, for each action, we calculated the module of the perceived velocity ($|v| = v_x^2 + v_y^2$) and the trajectory (Fig. 2B2, B3; see also Supplementary Fig. S4).

$\sqrt{\quad}$

Experimental Paradigm

Participants laid in the scanner in a dimly lit environment. The stimuli were presented via digital audio/video system (VisuaSTIM) with 30 dB noise-attenuating headset with 40 Hz to 40 kHz frequency response and with a 500.000 px × 0.25 square inch resolution with horizontal eye field of 30°. The digital transmission of the signal to the scanner was via optic fiber. The software EPrime 2 Professional (Psychology Software Tools, Inc., http://www.pstnet.com) was used both for stimulus presentation and the recording of participants' answers. A sparse block design (van Atteveldt et al. 2004; Gazzola et al. 2006) was used in the experiment. The scan cycle (TR) covering the whole brain was collected in 2 s (acquisition time) followed by a silence period lasting 2 s (TR = 4 s). The experimental stimuli were presented during the silence period. Audio and video stimuli were presented in blocks of 4 consecutive stimuli of the same condition (speech session: vitality listening, control listening, vitality speech imagination, control speech imagination; action session: vitality observation, control observation, vitality action imagination, control action imagination). An inter block period of 12 s (3 TR) without audio/video stimuli was present between 2 consecutive blocks. In the 18% of cases, in the inter block period, experimental catch trials were randomly presented and participants had to indicate the correct vitality form listened/ observed/imagined or the color of the presented object by pressing a button on a response box placed inside the scanner (see Supplementary Fig. S1). The experiment was composed of 4 functional runs (2 runs for the speech session and 2 runs for the action session) with a total of 8 blocks (32 single trials) for each condition, presented in a randomized order. Each functional run lasted about 9 min.

fMRI Data Acquisition

Anatomical T_1-weighted and functional T_{2*}-weighted MR images were acquired with a 3-T General Electrics scanner equipped with an 8-channel receiver head-coil. Functional images were acquired using a T_{2*}-weighted gradient-echo, echo-planar imaging (EPI) pulse sequence acceleration factor asset 2, 37 sequential transverse slices (slice thickness = 3 plus interslice gap = 0.5 mm) covering the whole brain, with a TR time of 4 s (flip angle = 90°, FOV = 205 × 205 mm^2, in-plane resolution 2.5 × 2.5 mm^2). Each scanning run comprised 135 ascending sequential volumes. Additionally, a T_1-weighted structural image was acquired for each participant (acceleration factor arc 2, 156 sagittal slices, matrix 256 × 256, isotropic resolution 1 × 1 × 1 mm^3, TI = 450 ms, TR = 8100 ms, TE = 3.2 ms, flip angle 12°).

Univariate Statistical Analysis

Data analysis was performed with SPM8 (Statistical Parametric Mapping software; The Wellcome Department of Imaging Neuroscience; http://www.fil.ion.ucl.ac.uk) running on MATLAB R2013 (The Mathworks, Inc.). The first 3 volumes of each run were discarded to allow T_1 equilibration effects. For each participant, all volumes were spatially realigned to the first volume of the first session and unwarped to correct for between-scan motion, and a mean image from the realigned volumes was created. Time slicing correction was applied according to sparse imaging acquisition (acquisition time, TA = 2s). All volumes were normalized to the SPM EPI template (Montreal Neurological Institute, MNI space) and resampled to a final spatial resolution of 2 × 2 × 2mm^3 and smoothed with an isotropic Gaussian kernel with a full-width half-maximum of 6mm. For all participants, we checked possible head motion and found it never exceeded one voxel (3mm).

Two statistical analyses were carried out. The first analysis was executed using the peak-level inferences (Friston et al. 2006) in order to assess the voxels selectivity of the left insula in vitality forms processing. For this purpose, we built a region of interest (ROI) at the level of the left insula (size of 1634 voxels) drawing a line between the border of the insula and the parietal, frontal, and temporal opercula cortices, which were all excluded from the ROI. To be sure that each drawn point belonged to the insula, for each slice, we checked the coordinates of the border points with SPM Anatomy Toolbox (Version 2.2c).

In addition, in order to assess the general activation patterns, a second analysis was executed in the whole brain.

Data were analyzed using a random-effects model (Friston et al. 1999), implemented in a 2-level procedure. In the first level, single-subject fMRI "blood oxygen level-dependent" (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. This analysis employed event-related convolution models using the hemodynamic response function (HRF) provided by SPM8.

Two different GLM models were used. The first GLM model was used for the speech session and was composed of 6 regressors as follows: "Vitality Listening", "Ctrl Listening", "Vitality Speech Imagination", "Ctrl Speech Imagination", "Instruction," and "Response". The second GLM model was used for the action session and was composed of other 6 regressors: "Vitality Observation", "Ctrl Observation", "Vitality Action Imagination", "Ctrl Action Imagination", Instruction, and Response.

The experiment was organized in a mixed sparse design. Audio and video stimuli were presented in blocks of 4 consecutive stimuli of the same condition (speech session: vitality listening, control listening, vitality speech imagination, control imagination; action session: vitality action observation, control observation, vitality action imagination, control imagination). Within one block, each single stimulus was modeled as a single event lasting 2 s. The response was modeled as a single event lasting 14 s (2 s for the response + 12 s post response). 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 considering the activation patterns obtained for the direct contrast between conditions ("Vitality Listening vs. Ctrl Listening", "Vitality Speech Imagination vs. Ctrl Speech Imagination", "Vitality Observation vs. Ctrl Observation", "Vitality Action Imagination vs. Ctrl Action Imagination"). The location of the activation foci was determined in the stereotaxic space of MNI coordinates system.

To highlight voxels activated both in the contrast "Vitality Listening versus Ctrl Listening" and in the contrast "Vitality Speech Imagination versus Ctrl Speech Imagination" ("Speech VF cluster"), a conjunction analysis was performed. In addition, a second conjunction analysis was performed to underline voxels activated both in the contrast "Vitality Observation versus Ctrl Observation" and in the contrast "Vitality Action Imagination versus Ctrl Action Imagination" ("Action VF cluster"). Finally, voxels involved in the recognition and imagination of the vitality forms were highlighted using a conjunction analysis of all 4 experimental conditions (Vitality Listening, Vitality Speech Imagination, Vitality Observation, and Vitality Action Imagination) relative to controls. All fMRI results (including the PPI analyses, see below) of this study were considered significant at $P < 0.05$ after correction for multiple comparisons using the family-wise error (FWE) rate at the peak level (insular analysis) and cluster level (whole brain analysis), respectively.

Using the SPM Rex Toolbox (http://web.mit.edu/swg/rex), for each participant we extracted the BOLD signal change in 2 ROIs of the left insula. In particular, ROI1 corresponded to the Speech VF cluster (blue color; Fig. 3B1), while ROI2 corresponded to the Action VF cluster (red color; Fig. 3B1). The Rex tool permitted to extract in these 2 ROIs the signal change values associated with rude/gentle vitality forms and relative control calculated for each participant on the basis of contrast images based on the previous flexible ANOVA model (second-level analysis; see Supplementary Material). These data were analyzed with a GLM analysis and Newman Keuls was used as a post hoc test.

Testing for Task Attention: Behavioral Analysis

To test the responses accuracy of participants, we carried out a further analysis based on the responses given by the participants during the scanning sessions when presented with the catch trials, that is, those trials in which the participants were required to give an explicit response on the presented stimuli. More specifically, during the speech and action tasks, participants had to indicate if the vitality form listened/observed/imagined was rude or gentle (vitality form condition) or if the color of the presented object was blue/green/orange (control condition) by pressing a button on a response box placed inside the scanner. For each participant, 8 responses were recorded in the speech session and 8 responses were recorded in the action session (16 responses in total). The mean responses accuracy of participants was 82.4% (SD ±10.6%).

Psycho-Physiological Interaction

To identify possible brain regions that increased their relationship with the left dorso-central insula in the processing of the speech and action vitality forms, 2 PPI analyses were carried out (Friston et al. 1997). In particular, these analyses allowed us to highlight brain regions whose activity depends on an interaction between speech and action vitality processing (Psychological task) and the BOLD signal change of the left dorso-central insula during the contrast "Vitality versus Control" (Physiological response). The PPI analysis employed 3 different steps: extraction of the BOLD signal in a seed ROI; PPI analysis; PPI GLM analysis.

Step 1. For each participant, 2 seed regions (local maxima) were localized in the left dorso-central insula (middle and

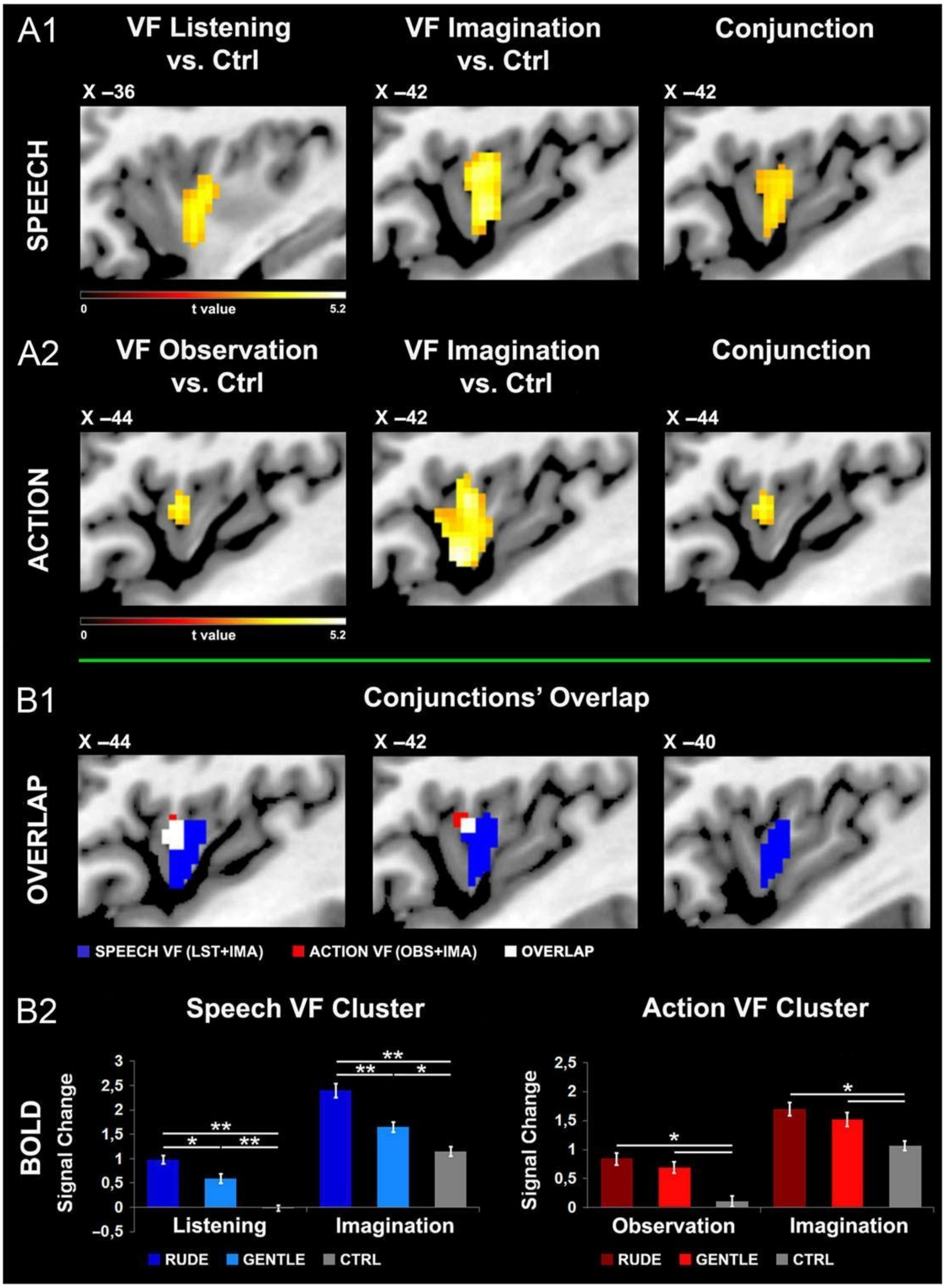

Figure 3. Insular activations obtained in speech and action vitality processing. (A1) Parasagittal sections showing the left insular activations during the speech session in the contrast "Vitality versus Ctrl" during LST (local maxima: X =−36, Y = 6, Z =−4) and Speech IMA tasks (local maxima: X =−42, Y = 6, Z = 6). (A2) The central insula was also active during the action session in the contrasts Vitality versus Ctrl in OBS (local maxima: X =−44, Y = 10, Z = 4) and Action IMA tasks (local maxima: X =−42, Y = 10, Z = 6). The conjunction analysis revealed in the speech session a common activation of the dorso-central sector of insula in the LST and Speech IMA tasks (A1, right side). A very similar insular sector was also found activated in the action conjunction analysis in OBS and Action IMA tasks (A2, right side). For all insular analyses, significant threshold has been set at $P_{FWE} < 0.05$ voxel level. (B1) Overlapping of the 2 conjunction analyses. The insular sectors found active in the conjunction analysis of speech and action sessions are, respectively, shown in blue color (161 voxels) and red color (21 voxels). White color highlighted the area (15 voxels) activated among all tasks (LST, speech IMA, OBS, action IMA) in the expression and recognition of the speech and action vitality forms. (B2) BOLD signal changes extracted in ROI1 (Speech VF cluster, blue color, B1) and ROI2 (Action VF cluster, red color, B1) of the left insula resulting from the speech tasks (LST, IMA) and action tasks (OBS, IMA). The horizontal line above the columns indicates the comparisons among rude vitality form, gentle vitality form, and control. The bars indicate the SEM. Asterisks indicate significant differences set at $P < 0.05$ (*) and $P < 0.001$ (**).

posterior short gyri). In particular, one seed region was localized on the basis of first-level GLM results obtained using the contrast Vitality Speech (Vitality Listening + Vitality Speech Imagination) versus baseline in the speech session (speech seed region) and one on the basis of data obtained in the action session using the contrast Vitality Action (Vitality Observation + Action Imagination) versus baseline (action seed region, see Supplementary Materials). After the seed regions localization, using the SPM8 eigenvariate function, we extracted for each seed region and for each participant, the time course of activity in a volume of interest (VOI) centering a 3-mm radius sphere around the local maxima identified before. More specifically, for each participant the first VOI was created around the local maxima obtained in the speech session (mean coordinates: x = −42, y = 7, z = −3; SD ± 3 mm). The second VOI was created around the local maxima obtained in the action session (mean coordinates: x = −41, y = 10, z = 3; SD ± 3 mm).

Step 2. The PPI analysis employed 3 regressors as follows: one regressor represented the deconvolved activation time course in the speech/action seed region (Y vector, physiological variable), one regressor represented the contrast of interest ([“Vitality Speech vs. Ctrl Speech”; “Vitality Action vs. Ctrl Action”] P vector, psychological variable), and one regressor representing the interaction of the previous 2 vectors (PPI, the interaction term). More specifically, 2 PPI analyses were carried out separately for each participant. The first PPI analysis assessed the neural activity in the speech seed region by multiplying the deconvolved BOLD signal with psychological vector for the onset times of the “Vitality speech” (1) and “Ctrl speech” (−1) trials (Vitality Speech vs. Ctrl Speech). The second PPI analysis assessed the neural activity in the action seed region by multiplying the deconvolved BOLD signal with psychological vector for the onset times of the “Vitality Action” (1) and “Ctrl Action” (−1) trials (Vitality Action vs. Ctrl Action).

Step 3. After the convolution of HRF, for each participant, the 3 regressors (PPI term, Y vector, and P vector) were entered run by run into a first-level GLM to determine individual brain regions showing psycho-physiological interaction with the speech/action seed regions. Finally, using 2 different second-level GLM, we identified the brain regions showing PPI connectivity with the speech seed region and the action seed region, respectively.

Results

Insular Activation During the Speech Session: Vitality form Versus Ctrl Condition

The main aim of our study was to assess the cortical activation in the insula selective for the speech vitality forms processing. The results of the insular analysis revealed that the contrast “speech vitality form” relative to the “speech control” condition produced activation of the dorso-central sector of the insula in the listening task (cluster size: 229 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −36, Y = 6, Z = −4, $P_{FWE} < 0.05$ voxel level). For the same contrast, the dorso-central sector of

Table 1 Cerebral activity during the direct contrasts: (A) “Listening versus Ctrl”; (B) “Speech Imagination versus Ctrl”; (C) “Observation versus Ctrl”; and (D) “Action Imagination versus Ctrl”

Anatomical region	Left hemispher				Right hemisphere			
		y	z	Z-score	x	y	z	Z-score
SPEECH								
(A) Listening versus Ctrl								
Middle temporal gyrus	−48	−64	16	5.58				
Insula	−36	2	2	3.95				
IFG (pars orbitalis)	−46	34	−8	5.69				
IFG (pars triangularis)	−52	20	2	5.55	54	28	4	5.01
Superior medial gyrus	−8	54	30	4.71				
(B) Imagination versus Ctrl								
Middle temporal gyrus	−52	−64	14	6.91	60	−60	8	5.14
Medial temporal pole					48	10	−32	4.93
Insula	−40	4	8	4.05				
IFG (pars orbitalis)					56	28	−8	5.65
IFG (pars triangularis)					52	32	−2	5.90
Supramarginal gyrus	−58	−38	30	5.52	68	−22	30	4.99
Middle cingulate cortex	−6	−30	48	4.63				
Insula	−44	6	−4	4.27				
Cerebellum	−34	−54	−24	5.20				
ACTION								
(C) Observation versus Ctrl								
Middle temporal gyrus	−56	−62	4	4.33	50	−80	0	5.16
IFG (pars triangularis)	−46	24	10	4.65	58	28	8	5.27

IFG (pars opercularis)	−44	10	4	4.22	60	20	26	3.93
Fusiform gyrus					38	−44	−24	4.87
(D) Imagination versus Ctrl								
IFG (pars triangularis)	−44	28	12	5.02	52	30	−2	4.20
Insula	−36	6	−4	4.26				
Posterior medial frontal gyrus	−12	10	68	4.82	8	14	62	4.56
Middle occipital gyrus	−34	−96	2	4.78				

Local maxima, as shown in Supplementary Figure S3, are given in MNI standard brain coordinates. Significant threshold was set at $P_{FWE} < 0.05$ (cluster level).

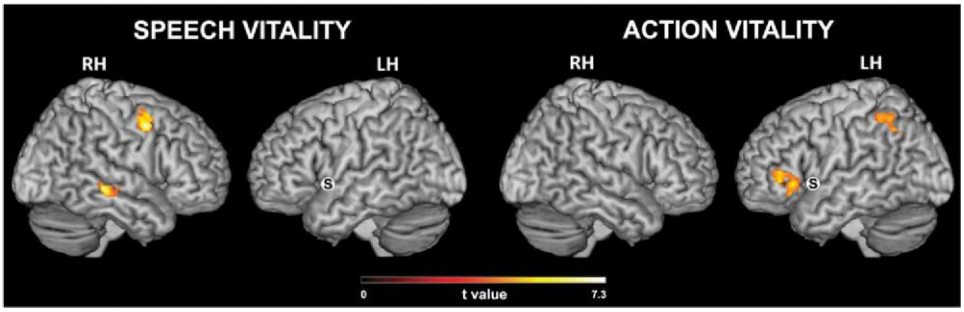

Figure 4. Brain areas connectivity. Figure shows brain areas increasing functional connectivity with the left dorso-central insula (S), respectively, in speech vitality processing (Vitality Speech vs. Ctrl Speech; left side) and action vitality processing (Vitality Action vs. Ctrl Action; right side). These activations are rendered into a standard MNI brain template ($P_{FWE} < 0.05$ cluster level). LH, left hemisphere; RH, right hemisphere.

insula was found also active in the "speech imagination" task (cluster size: 220 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −42, Y = 6, Z = 6, $P_{FWE} < 0.05$ voxel level; Fig. 3A1). In addition, the whole brain analysis confirmed the same insula activation and showed in both listening and speech imagination tasks enhanced bilateral activations in the left middle temporal gyrus and the inferior frontal gyrus ($P_{FWE} < 0.05$ cluster level; Supplementary Fig. S3A1; for statistical data and MNI coordinates, see Table 1).

Insular Activation During the Action Session: Vitality form Versus Ctrl Condition

A second fundamental aim of our study was to assess the cortical activation in the insula selective for the action vitality forms processing. The results of the insular analysis revealed that the contrast "action vitality form" relative to "action control" condition produced activation of the dorso-central sector of the insula in the observation task (cluster size: 21 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −44, Y = 10, Z = 4, $P_{FWE} < 0.05$ voxel level). For the same contrast, the dorso-central sector of insula was found also active in the "action imagination" task (cluster size: 293 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −42, Y = 10, Z = 6, $P_{FWE} < 0.05$ voxel level; Fig. 3A2). The whole brain analysis confirmed the same insula activation. In addition, this analysis showed for the observation task, enhanced bilateral activations in the middle temporal gyrus and inferior frontal gyrus, and, for the action imagination task, bilateral activations in the left occipital lobe and inferior frontal gyrus bilaterally ($P_{FWE} < 0.05$ cluster level; Supplementary Fig. S3A2, for statistical data and MNI coordinates, see Table 1).

Conjunction Analyses

Results of the conjunction analysis between speech listening and imagination tasks revealed a common activation of the dorso-central sector of insula (cluster size: 161 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −42, Y = 4, Z = 4, $P_{FWE} < 0.05$ voxel level; Fig. 3A1). The same sector of the insula was also found activated in the conjunction analysis between action observation and imagination tasks (cluster size: 21 voxels, $P_{FWE} < 0.05$ cluster level; local maxima: X = −44, Y = 10, Z = 4, $P_{FWE} < 0.05$ voxel level; Fig. 3A2). Most interestingly, the overlap between the 2 conjunction analyses revealed a larger and more ventrally activation for speech relative to that found for action (Fig. 3B1). Finally, the conjunction analysis among all 4 tasks (listening, speech imagination, observation, action imagination) highlighted a region involved in the recognition and expression of the speech and action vitality forms (local maxima: X = −44, Y = 8, Z = 4, $P_{FWE} < 0.05$ voxel level; Fig. 3B1, white color).

Testing for the Vitality Effect: ROIs Analysis

Figure 3B2 illustrates the mean BOLD signal changes extracted in ROI1 (Speech VF cluster, blue color, B1) and ROI2 (Action VF cluster, red color, B1) during the speech tasks (LST, IMA) and action tasks (OBS, IMA). A significant difference was present among the 3 conditions (rude vitality form, gentle vitality form, and control) in ROI1 during the speech tasks (LST, $F_{2,34} = 19.6$, $P < 0.001$, partial-$\eta^2 = 0.5$, $\delta = 1$; IMA, $F_{2,34} = 17{,}1$, $P < 0.001$, partial-$\eta^2 = 0.5$, $\delta = 0.9$) and in ROI2 during the action tasks (OBS, $F_{2,34} = 4.7$, $P < 0.01$, partial-$\eta^2 = 0.2$, $\delta = 0.7$; IMA, $F_{2,34} = 5.5$, $P < 0.01$, partial-$\eta^2 = 0.2$, $\delta = 0.8$). Post hoc analysis revealed that for ROI1, Rude and Gentle conditions produced higher signal change relative to the Ctrl for LST and IMA speech tasks (LST: Rude > Ctrl $P < 0.001$, Gentle > Ctrl $P < 0.001$; IMA: Rude > Ctrl $P < 0.001$, Gentle > Ctrl $P < 0.05$). A significant difference was present between Rude and Gentle conditions in LST task ($P < 0.05$) and IMA speech task ($P < 0.001$). Additionally, post hoc analysis revealed that for ROI2, Rude and Gentle conditions produced higher signal change relative to the Ctrl for OBS and IMA action tasks (OBS: Rude > Ctrl $P < 0.05$, Gentle > Ctrl $P < 0.05$; IMA: Rude > Ctrl $P < 0.05$, Gentle > Ctrl $P < 0.05$). No difference was present between Rude and Gentle conditions ($P > 0.05$).

PPI Analysis

To identify possible brain areas connected with the left dorsocentral insula, we conducted 2 PPI analyses, one in the speech session and one in the action session. For the speech session, the effect of the contrast "Vitality Speech versus Ctrl Speech", pooling together listening and speech imagination tasks, revealed a significant connectivity of the left dorso-central insula with the superior temporal gyrus (x = 44, y = −36, z = 6) and precentral gyrus in the right hemisphere (x = 52, y = 4, z = 44; $P_{FWE} < 0.05$ cluster level; Fig. 4, left side).

Differently, for the action session, the effect of the contrast "Vitality Action versus Ctrl Action", pooling together observation and action imagination tasks, revealed a significant connectivity of the left dorso-central insula with the intra parietal sulcus (x = −36, y = −54, z = 36), inferior parietal lobe (x = −42, y = −54, z = 52) and inferior frontal gyrus in the left hemisphere (x = −54, y = 26, z =−2; $P_{FWE} < 0.05$ cluster level; Fig. 4, right side).

Discussion

Action and speech may take different forms. A hand shaking can be gentle or vigorous. Similarly, the voice tone could be imperative or pleading. These aspects of social communication have been virtually neglected in neuroscience. In contrast, several psychological studies have stressed their importance in social life (Brazelton et al. 1975; Trevarthen 2006). In particular, the studies of Stern (1985, 2010) have shown that these aspects are rooted in our most basic social interactions and shape our experience of our self and others. Stern coined the term of "vitality affects" or "vitality forms" to describe these action and speech forms (1985, 2010).

In a previous study, we demonstrated that listening to action verbs pronounced with different vitality forms activate selectively, relative to controls, the dorso-central insula (Di Cesare et al. 2016). The first aim of the present study was to establish whether imaging to pronounce action verbs with specific vitality forms (rude or gentle) would activate the same sector of the insula as listening to sentences with the same vitality forms.

The results showed that this is the case: the dorso-central insula is indeed activated, relative to controls, both when participants listen to action verbs pronounced with different vitality forms and when they imagine themselves internally generate the same action verbs with the same vitality forms. Note that although the participants were instructed to not vocalize during the imagination task, they have, of course, to simulate internally the oro-laryngeal gestures that they would have overtly used to pronounce the same verbs.

Internal simulation of actions has been named motor imagery by Jeannerod (1995). It represents a specific aspect of motor preparation. Motor imagery activates areas corresponding to those that are involved in action execution except for the primary motor cortex. Thus, the presence of voxels selectively activated for both listening and imaging speech vitality forms strongly suggests the existence of the mirror mechanism for speech vitality forms in the dorso-central insula. Unlike the mirror mechanism located in the parietal and frontal areas, which plays a role in action goal understanding, the speech mirror mechanism located in the insula allows one to express own affective states and to understand those of others.

The second aim of the present study was to assess whether the sector of insula involved in the processing of "speech" vitality forms was also activated during the processing of the "action" vitality forms of the same actions. More specifically, we aimed to understand whether listening to an action verb like "open" pronounced by the actors with different vitality forms and observing them to open a jar with different vitality forms would activate the same voxels of the insula.

The results showed that both action vitality forms (observation and imagination) and speech vitality forms (listening and imagination) determine the activation of the dorso-central insula. The location of the action vitality forms coincided with that described in a previous paper (Di Cesare et al. 2015). It is interesting to note that this correspondence was present regardless of the fact that in the videos used in the present study the observed action was executed by an actor sitting opposite to the participants (180°), while in the videos of the previous study, the observed action were seen as performed by the participant (0°). This demonstrates that this sector of insula encodes vitality forms in the case of both egocentric and allocentric actions.

The findings that the dorso-central sector of the insula is involved in the encoding action vitality forms are in line with data on the functional organization of the insula in monkeys and humans. Using intracortical electrical stimulation, it has been shown that in the monkey, the insula is subdivided in different sectors endowed with specific functional properties. The stimulation of the rostral sector of the insula determines positive ingestive behavior dorsally, and negative ingestive behavior (disgust) ventrally (Jezzini et al. 2012). In contrast, the stimulation of the dorso-central sector of insula elicits movements of the upper limb or of the body. In humans, Kurth et al. (2010) reported a similar insular organization based on a metaanalysis of large number of functional neuroimaging studies. In particular, they reported that human insula consists of 4 distinct functional fields: the cognitive field, the sensorimotor, the olfactory-gustatory and the socio-emotional. The sensorimotor field appears to closely correspond to the sector involved in vitality forms processing.

Although we found a considerable overlap between the action and speech vitality form locations, the speech vitality forms also produced activations located more ventrally than action vitality forms. This distinction between action and speech vitality form encoding was confirmed by the PPI analysis of our data. This analysis showed that there are 2 distinct sets of areas connected with the insula sectors involved in speech and action vitality forms. Specifically, the action related sector is connected with the "left" intra parietal sulcus, the left inferior parietal lobe and the posterior part of the left inferior frontal gyrus. In contrast, the speech related sector is connected with the right hemisphere and specifically with the "right" posterior part of the superior temporal sulcus (STS) and with the right precentral gyrus.

As far as vitality form of action sector of the dorsal–central insula is concerned, anatomical data in the monkey showed that the this sector has rich connections with the parietal and frontal areas that form the circuit controlling the organization of arm movements (Jeannerod et al. 1995; Nelissen and Vanduffel 2011), and namely areas anterior intraparietal (Borra et al. 2008), F5 (Gerbella et al. 2011), and prefrontal area 12r (Borra et al. 2011). It is important to note that a homologous parieto-frontal circuit underlies arm movement organization also in humans (Iacoboni 2009; Caspers et al. 2010; Grosbras et al. 2012; Molenberghs et al. 2012; Rizzolatti et al. 2014). Furthermore, Almashaikhi et al. (2014) recently showed that the electrical stimulation of the middle and posterior short gyri of the insula in patients with drug-resistant epilepsy determines evoked potentials in the parietal lobe and inferior frontal gyrus. These findings confirm in humans the connectivity of the central insula with the cortical areas involved in the control of the voluntary movements first described in the monkey.

As far as vitality forms of the speech insular sector are concerned, our data confirm a dominance of the right hemisphere, in line with the prevailing model of prosody (Witteman et al. 2011). They are also in full agreement with the recent brain imaging data provided by Sammler et al. (2015), who showed that the right posterior STS and the right precentral gyrus are 2 areas involved in the prosody processing. The PPI data of our study show that the central dorsal part of insula is functionally connected with these areas and possibly is responsible for their modulation linked of the internal state of the speaker.

These findings and those previously discussed on action vitality form encoding, strongly support the idea that, during action execution, the central insula modulates the "left parietofrontal action circuit" allowing the agent to produce rude or gentle action vitality forms, while during speech production, the insula modulate the right premotor cortex controlling th

References

Almashaikhi T, Rheims S, Jung J, Ostrowsky-Coste K, Montavont A, De Bellescize J, Arzimanoglou A, Kosal PK, Guenot M, Bertrand O, et al. 2014. Functional connectivity of insular efferences. Hum Brain Mapp. 35:5279–5294.

Borra E, Gerbella M, Rozzi S, Luppino G. 2011. Anatomical evidence for the involvement of the macaque ventrolateral prefrontal area 12r in controlling goal-directed actions. J Neurosci. 31:12351–12363.

Borra E, Belmalih A, Calzavara R, Gerbella M, Murata A, Rozzi S, Luppino G. 2008. Cortical connections of the macaque anterior intraparietal (AIP) area. Cereb Cortex. 18:1094–1111.

Brazelton TB, Tronick E, Adamson L, Als H, Wise S. 1975. Early mother-infant reciprocity. Ciba Found Symp. 33:137–154.

Caruana F, Avanzini P, Gozzo F, Francione S, Cardinale F, Rizzolatti G. 2015. Mirth and laughter elicited by electrical stimulation of the human anterior cingulate cortex. Cortex. 71:323–331.

Caspers S, Zilles K, Laird AR, Eickhoff SB. 2010. ALE metaanalysis of action observation and imitation in the human brain. Neuroimage. 50:1148–1167.

Di Cesare G, Fasano F, Errante A, Marchi M, Rizzolatti G. 2016. Understanding the internal states of others by listening to action verbs. Neuropsychologia. 89:172–179.

Di Cesare G, Di Dio C, Marchi M, Rizzolatti G. 2015. Expressing and understanding our internal states and those of others. Proc Natl Acad Sci U S A. 112(33):10331–10335.

Friston KJ, Ashburner J, Kiebel S, Nichols T, Penny W. 2006. Topological inference. In: Statistical parametric mapping. London (UK): Academic Press. p. 237–244.

Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. 2002. Classical and Bayesian inference in neuroimaging: applications. Neuroimage. 16:484–512.

Friston KJ, Holmes AP, Worsley KJ. 1999. How many subjects constitute a study? Neuroimage. 10:1–5.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage. 6:218–229.

Gazzola V, Aziz-Zadeh L, Keysers C. 2006. Empathy and the somatotopic auditory mirror system in humans. Curr Biol. 16:1824–1829.

Gerbella M, Belmalih A, Borra E, Rozzi S, Luppino G. 2011. Connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. Brain Struct Funct. 216: 43–65.

Grosbras MH, Beaton S, Eickhoff SB. 2012. Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. Hum Brain Mapp. 33:431–454.

Iacoboni M. 2009. Imitation, empathy, and mirror neurons. Annu Rev Psychol. 60:653–670.

Jeannerod M. 1995. Mental imagery in the motor context. Neuropsychologia. 33(11):1419–1432.

Jeannerod M, Arbib MA, Rizzolatti G, Sakata H. 1995. Grasping objects: the cortical mechanisms of visuomotor transformation. Trends Neurosci. 18:314–320.

Jezzini A, Caruana F, Stoianov I, Gallese V, Rizzolatti G. 2012. The functional organization of the insula and of inner perisylvian regions: an intracortical microstimulation study. Proc Natl Acad Sci U S A. 109:10077–10082.

Kurth F, Zilles K, Fox PT, Laird AR, Eickhoff SB. 2010. A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. Brain Struct Funct. 214:519–534.

Molenberghs P, Cunnigton R, Mattingley JB. 2012. Brain regions with mirror properties: a meta-analysis of 125 fMRI studies. Neurosci Behav Rev. 36:341–349.

Nelissen K, Vanduffel W. 2011. Grasping-related functional magnetic resonance imaging brain responses in the macaque monkey. J Neurosci. 31:8220–8229.

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. Physiol Rev. 94:655–706.

Sammler D, Grosbras MH, Anwander A, Bestelmeyer PE, Belin P. 2015. Dorsal and ventral pathways for prosody. Curr Biol. 25 (23):3079–3085. doi10.1016/j.cub.2015.10.009. Stern DN. 1985. The interpersonal world of the infant. New York: Basic Books.

Stern DN. 2010. Forms of vitality exploring dynamic experience in psychology, arts, psychotherapy, and development. UK: Oxford University Press.

Trevarthen C. 2006. Collaborative regulations of vitality in early childhood: stress in intimate relationships and postnatal psychopathology. In: Cicchetti D, Cohen DJ, editors. Developmental psychopathology. Volume 2. 2nd ed. New York: Wiley.

van Atteveldt N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. Neuron. 43:271–282.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. 2003. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. Neuron. 40: 655–664.

Witteman J, van Ijzendoorn MH, van de Velde D, van Heuven VJ, Schiller NO. 2011. The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. Neuropsychologia. 49: 3722–3738.