Contents lists available at ScienceDirect

# Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

# Understanding the internal states of others by listening to action verbs

G. Di Cesare [a], F. Fasano [a], A. Errante [a], M. Marchi [b], G. Rizzolatti [a,c,d,]*

[a] University of Parma, Department of Neuroscience, via Volturno 39/E, 43100 Parma, Italy
[b] University of Milan, Department of Computer Science, Via Comelico 39/41, 20135 Milan, Italy
[c] Istituto di Neuroscienze, Consiglio nazionale delle Ricerche, Parma, Italy
[d] IIT (Italian Institute of Technology) Brain Center for Social and Motor Cognition, via Volturno 39/E, 43100 Parma, Italy

## ABSTRACT

The internal state of others can be understood observing their actions or listening to their voice. While the neural bases of action style (vitality forms) have been investigated, there is no information on how we recognize others' internal state by listening to their speech. Here, using fMRI technique, we investigated the neural correlates of auditory vitality forms while participants listened to action verbs in three different conditions: human voice pronouncing the verbs in a rude and gentle way, robot voice pronouncing the same verbs without vitality forms, and a scrambled version of the same verbs pronounced by human voice. In agreement with previous studies on vitality forms encoding, we found specific activation of the central part of insula during listening to human voice conveying specific vitality forms. In addition, when listening both to human and robot voices there was an activation of the posterior part of the left inferior frontal gyrus and of the parieto-premotor circuit typically described to be activated during observation and execution of arm actions. Finally, the superior temporal gyrus was activated bilaterally in all three conditions. We conclude that, the central part of insula is a key region for vitality forms processing allowing the understanding of the vitality forms regardless of the modality by which they are conveyed.

© 2016 Published by Elsevier Ltd.

## 1. Introduction

Actions and emotions might take different forms. A caress, for example, can be *gentle* or *rushed*, a hand shaking can be *delicate* or *vigorous*; similarly, also emotions, anger for example, can *explode* or be *cold*. Attention to the "forms" of the observed gestures provides information about agents' affective states, moods, attitudes and even their character traits. Action forms have been variously conceived by different researchers (Trevarthen, 1998). Stern coined the term of "vitality affects" or "vitality forms" to highlight that they are rooted in our most basic social interactions, thus deeply shaping our experience of ourselves and of others (Stern, 1985, 2010). On Stern's view, actions and emotions cannot be experienced without one or another vitality form. Indeed, *what* we are doing or feeling gives *content* to our experience, while vital affects provide these contents with a *form*, which characterizes *how* actions and emotions are displayed and experienced.

In a previous study we demonstrated that the observation of arm actions performed with two different vitality forms, rude, that is abruptly done, and gentle, determined the activation of the

*dorso-central* sector of the insula (Di Cesare et al., 2013) corresponding to middle and posterior insula short gyri. In addition, the activation of the same sector was also found during the imagination and the execution of the vitality forms, suggesting the possibility that the insula could be endowed with a mirror mechanism (Di Cesare et al., 2015).

In the present study we addressed the issue if listening to different action verbs pronounced with different vitality forms (rude and gentle) would activate the insula and more specifically its central sector. We were interested in investigating the insula responses to vitality forms presented acoustically because this modality is particular important in conveying the vitality forms and represents a primordial way to relate to and understand others.

In this respect, it has also been reported that listening to words pronounced with different vitality forms plays a crucial role in the attunement between mother and child. Mothers' musical intonations, rhythmic gestures, and affective movements are shared with infants before the acquisition of language (Trevarthen, 1998). After birth, infants manifest an intense interest in the prosodic intonation of maternal talk, and they try to synchronize their expressions with those of the mother (Marwick and Murray, 2009).

In the present study, in order to convey vitality forms, we used action verbs pronounced in imperative form and directed to the

* Corresponding author at: Department of Neuroscience, University of Parma, Via Volturno 39/E, 43100 Parma, Italy.

participants to actively involve the listener in vitality forms perceiving. As a control, we presented the same stimuli pronounced by a robot voice that did not convey any obvious vitality form.

The main result of our study was the demonstration that the listening to words pronounced with different vitality forms determines a specific activation of the dorso-central sector of the insula. These data indicate that this sector of the insula, which, as mentioned above, is involved in the perception of vitality forms through the visual modality, is also active during listening to different vitality forms. Taken together, these findings appear to indicate that the central insula is the key node for the vitality forms encoding regardless of the stimulus modality.

## 2. Materials and methods

### 2.1. Participants

Sixteen healthy right-handed volunteers [8 females (mean age=27, s.d.=4.11; range=22–32) and 8 males (mean age=25.3; s.d.=3.54, range=20–32)] 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).

### 2.2. Experimental design

A sparse block design (van Atteveldt et al., 2004; Gazzola et al., 2006) was used in the experiment. The scan cycle (TR) was composed of 37 sequential slices (slice thickness=3 plus inter-slice gap=0.5 mm) covering the whole brain 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 stimuli were presented in blocks of three consecutive stimuli of the same condition (Vitality forms: *Vitality Rude, Vitality Gentle;* Scrambled Vitality forms: *Scrambled Rude, Scrambled Gentle; Robot; Silence*). Intermixed with experimental blocks were the catch trial blocks, in which randomly, participants had to indicate the category of the presented stimulus by pressing a button (human voice, robot voice, sound). An inter block period of 3 TR without audio stimuli was present between two consecutive blocks. The experiment was composed of 3 functional runs with a total of 12 blocks (36 single trials) for each condition, presented in a randomized order. Each functional run lasted about 11 min. Before the experiment, participants performed a training session to assess the audio stimuli recognition.

### 2.3. Stimuli

Native Italian participants were presented with audio stimuli consisting of Italian action verbs. More specifically, a male actor and a female actress pronounced 4 different action verbs in imperative mood [Italian used verbs: "dammi" (give), "prendi" (take), "tocca" (touch), "strappa" (tear)]. Half of the action verbs was presented by the female actress and half by the male actor. All the action verbs were pronounced using two different vitality forms: rude and gentle (*Vitality forms* condition). Additionally, for each action verb, we also presented two audio stimuli controls: a robot voice (Control 1, *Robot* condition) pronouncing the same action verbs as the actors and a scrambled version of the vitality forms audio stimuli (Control 2, *Scrambled VF* condition).

The voice of a robot, that is a voice emitted by an artificial agent devoid, by definition, of the capacity to produce vitality forms, allowed us to control the effect of action verbs pronounced by a human voice. The robot voice pronounced the same action verbs

maintaining the meaning but not conveying a vitality form. The action verbs pronounced by the robot voice were obtained by a vocal synthesizer (TextAloud software) and then processed with FL Studio 11 software. Then, the robot action verbs were equated for loudness in order to match the mean value loudness of the corresponding vitality form audio stimuli (rude and gentle).

In addition, a scrambled version of the vitality forms was used in order to control the physical properties of the stimuli and related responses (e.g. arousal effect). The scrambled version of each vitality form audio stimulus was obtained using the Discrete Fourier transform. More specifically, we decomposed each audio stimulus into its frequency components. For some component above 100 Hz we randomly changed the phase applying a rotation of 180°. This rotation preserved the same magnitude frequency spectrum of the original vitality forms action verbs and generated the scrambled effect.

All vitality forms audio stimuli were recorded using a cardiod 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. Vitality forms audio stimuli (rude and gentle) maintained their ecological loudness as well as their corresponding scrambled form audio stimuli.

A total of 16 experimental stimuli (4 action verbs × 2 vitality forms × 2 actors) and 20 control stimuli (16 scrambled: 4 action verbs × 2 vitality form × 2 actors; 4 robot action verbs) were presented. Each experimental trial lasted 2 s.

The physical characteristics of all presented audio stimuli were assessed using MATLAB (The Mathworks, Natick, MA). For each audio action verb, we estimate the sound wave amplitude, the pitch, and the frequency spectrum for all three conditions (*Vitality forms, Robot, Scrambled VF*). Fig. 1 shows the physical characteristics related to the "*dammi*" action verb pronounced by the female actress (see Supplementary Figs. S1–S7 for the other action verbs pronounced both by the male actor and female actress).

### 2.4. Paradigm and task

Participants laid 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 stimulus presentation and the recording of participants' answers. Before the experiment, participants performed a training session showing they were able to recognize the action verbs (98% on average).

During stimulus presentation, participants were requested to fixate a white cross on a black screen and listen to the audio stimuli. In the 16% of cases, the participants were asked to answer to a question (catch trial) related to presented category of the stimuli (Have you listened a robot voice, human voice, or a sound? ). The catch trials was randomly presented and lasted 2 s.

### 2.5. fMRI data acquisition

Anatomical T1-weighted and functional T2*-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 T2*-weighted gradient-echo, echo-planar (EPI) pulse sequence acceleration factor asset 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 mm$^2$, in-plane resolution 2.5 × 2.5 mm$^2$). The scanning sequence comprised 177 ascending sequential volumes. Additionally, a T1 weighted structural image
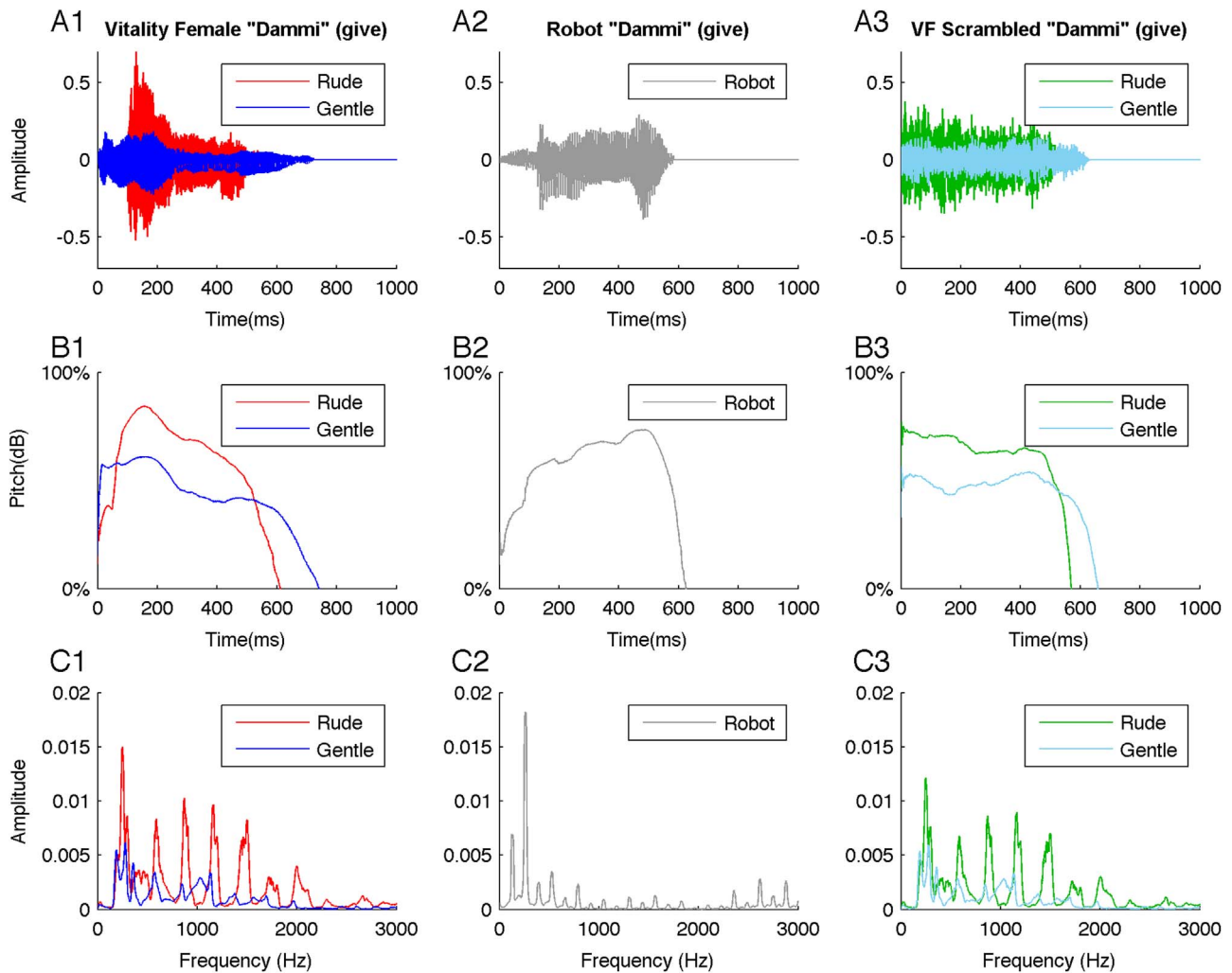
**Fig. 1.** Physical characteristics for the verb "dammi" (give). Graph A shows the audio wave amplitude for all categories [A1. Vitality: Rude, Gentle; A2. Robot; A3. Scrambled VF: Rude, Gentle]. Graph B shows the related pitch (mobile average 100 ms of the power of signal). Finally, graph C shows the frequency spectrum (magnitude of the Discret FFT).

was acquired for each participant (acceleration factor arc 2, 156 sagittal slices, matrix $256 \times 256$, isotropic resolution $1 \times 1 \times 1$ mm$^3$, 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 the 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 et al., 1999), implemented in a two-level procedure. In the first level, single-subject fMRI BOLD signal was modelled 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 GLM model was composed of seven regressors as follows: *Vitality Rude, Vitality Gentle, Scrambled Rude (SB Rude), Scrambled Gentle (SB Gentle), Robot, Silence, and Response*.

Audio stimuli were presented in blocks of three consecutive stimuli of the same condition (Vitality forms: *Vitality Rude, Vitality Gentle;* Scrambled Vitality forms: *Scrambled Rude, Scrambled Gentle; Robot; Silence*). Within each block, the audio stimuli were modelled as a single event lasting 2 s. The silence and response were also modelled as a single event lasting 2 s.

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 five regressors by contrasting the activation patterns obtained for each condition with silence (*Vitality Rude vs. Silence, Vitality Gentle vs. Silence, SB Rude vs. Silence, SB Gentle vs. Silence, Robot vs. Silence*). Within this model, we considered the activation patterns obtained for different conditions (*Vitality forms, Robot, Scrambled VF*) and the activations resulting from the direct contrast between conditions (*Vitality Rude vs. Robot, Vitality Gentle vs. Robot, Vitality Rude vs. SB Rude, Vitality Gentle vs. SB Gentle*). The location of the activation foci was determined in the stereotaxic space of MNI coordinates system.

On the basis of the obtained functional maps, we overlaid the

brain activations resulting from the contrasts *VF Rude vs. Robot* and *VF Gentle vs. Robot* to highlight the region of insula selective for both rude and gentle vitality forms as well as to assess possible differences between them. In the area resulting from this overlap, using MarsBaR ROI Toolbox for SPM (release 0.42), two regions of interest (ROIs) were created. The ROIs were defined centering the sphere (radius 2.5 mm) in the dorsal part (ROI 1: x= −38, y= −2, z=4) and in the central part (ROI 2: x= −38, y=5, z= −5), of the central insula, respectively. Then, the signal change for each subject 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 associated with rude and gentle vitality forms calculated for each subject on the basis of contrast images (second-level analysis) based on the previous flexible ANOVA model.

In addition, to highlight voxels activated either in the contrast *Vitality forms vs. Robot* [(Rude+Gentle) vs. Robot)] either in the contrast *Vitality forms vs. Scrambled VF* [(Rude+Gentle) vs. (SB Rude+SB Gentle)], a conjunction analysis between these two contrasts was performed. In the area resulting from the conjunction analysis, a new ROI was defined centering the sphere (radius 2.5 mm) at the activation maxima (x= −38, y=6, z= −6). Then, as mentioned above, the signal change values in the ROI, associated with vitality forms conditions (*Rude* and *Gentle*) and control conditions (*SB Rude, SB Gentle, Robot*) calculated for each subject were extracted.

## 3. Results

### 3.1. Overall effect of vitality, robot and scrambled VF

Hearing the vitality forms stimuli revealed a signal increase in the superior temporal gyrus (BA 22), left inferior parietal lobule (BA 40), left premotor (BA 6), left prefrontal cortex (BA 45) and posterior part of the inferior frontal gyrus (BA 47). All activations were larger in the left hemisphere (Fig. 2). In addition, there was a bilateral activation of the insula. A very similar activation pattern was observed for the *Robot* condition except for a larger activation of the left inferior parietal lobule in this case (BA 40). Finally listening to *Scrambled VF* produced a signal increase only in the auditory temporal areas (BA 21, 22; for statistical and coordinates see Table 1A, B, C).

### 3.2. Contrasts between vitality forms vs. robot and vitality forms vs. scrambled VF

The direct contrast *Rude vs. Robot* revealed a significant activation pattern in the left middle temporal gyrus (BA 21), left postcentral (BA 4) and precentral (BA 6) gyri and in the left central part of insula (Fig. 3A). The direct contrast *Gentle vs. Robot* revealed a significant activation pattern in the left central part of insula (for statistical and coordinates see Table 2A, B). In addition, the direct

**Table 1.**
Cerebral activity during (A) *Vitality form* listening *vs.* silence; (B) *Robot* listening *vs.* silence; (C) *Scrambled VF* listening *vs.* silence. Local maxima, as shown in Fig. 2, are given in MNI standard brain coordinates, significant threshold set at $P_{FWE} < 0.05$ (voxel-level).

| Anatomical region | Left hemisphere | | | | Right h | | | |
|---|---|---|---|---|---|---|---|---|
| | x | y | z | Z-score | x | y | z | Z-score |
| **(A) Vitality vs. silence** | | | | | | | | |
| Superior temporal gyrus | −60 | −28 | 6 | Inf | 56 | −30 | 2 | Inf |
| Middle temporal gyrus | −56 | −52 | 6 | Inf | 52 | −28 | −4 | Inf |
| Temporal pole | 52 | 6 | −18 | 7.80 | 58 | 4 | −16 | Inf |
| Posterior-Medial frontal gyrus | −4 | 8 | 56 | Inf | | | | |
| Precentral gyrus | −38 | 2 | 44 | 7.30 | 52 | 4 | 48 | 5.17 |
| IFG (*pars orbitalis*) | −50 | 30 | −6 | 5.42 | 54 | 32 | −6 | 5.34 |
| IFG (*pars opercularis*) | −48 | 6 | 12 | 6.69 | | | | |
| IFG (*pars triangularis*) | −36 | 22 | 24 | 6.39 | | | | |
| Inferior parietal lobule | −40 | −46 | 46 | 5.63 | | | | |
| Cerebellum | −30 | −64 | −28 | 5.31 | 32 | −70 | −28 | 5.65 |
| Middle frontal gyrus | −44 | 32 | 28 | 5.84 | | | | |
| Insula | | | | | 36 | 26 | 0 | 5.00 |
| **(B) Robot vs. silence** | | | | | | | | |
| Superior temporal gyrus | −60 | −28 | 6 | Inf | 64 | −28 | 4 | Inf |
| Middle temporal gyrus | −56 | −52 | 6 | 7.49 | | | | |
| Inferior parietal lobule | −34 | −62 | 48 | 6.70 | | | | |
| Posterior medial frontal gyrus | −6 | 8 | 58 | 7.58 | | | | |
| Middle frontal gyrus | −44 | 34 | 30 | 7.46 | 46 | 36 | 36 | 4.82 |
| IFG (*pars triangularis*) | −36 | 22 | 26 | 6.31 | | | | |
| IFG (*pars opercularis*) | −44 | 10 | 24 | 6.01 | | | | |
| IFG (*pars orbitalis*) | −46 | 28 | −8 | 5.28 | | | | |
| Temporal pole | −48 | 4 | −18 | 5.54 | 52 | 6 | −16 | 6.73 |
| Precentral gyrus | −36 | 2 | 44 | 6.92 | | | | |
| Cerebellum | −30 | −64 | −28 | 5.20 | 40 | −70 | −30 | 5.50 |
| **(C) Scrambled VF vs. silence** | | | | | | | | |
| Superior temporal gyrus | −56 | −28 | 6 | Inf | 64 | −28 | 6 | Inf |
| Middle temporal gyrus | −50 | −50 | 12 | 5.19 | | | | |

contrast *Rude vs. SB Rude* and *Gentle vs. SB Gentle* revealed activations in the left central insula, left parietal lobe (BA 40) and left premotor cortex (BA 6), in the auditory cortex (BA 21, 22) and in the inferior frontal gyrus bilaterally (BA 44, 45; see Fig. 3B, for statistical and coordinates see Table 2C, D).

### 3.3. Testing for the vitality effect: ROIs analysis

To highlight the region of insula selective for both rude and
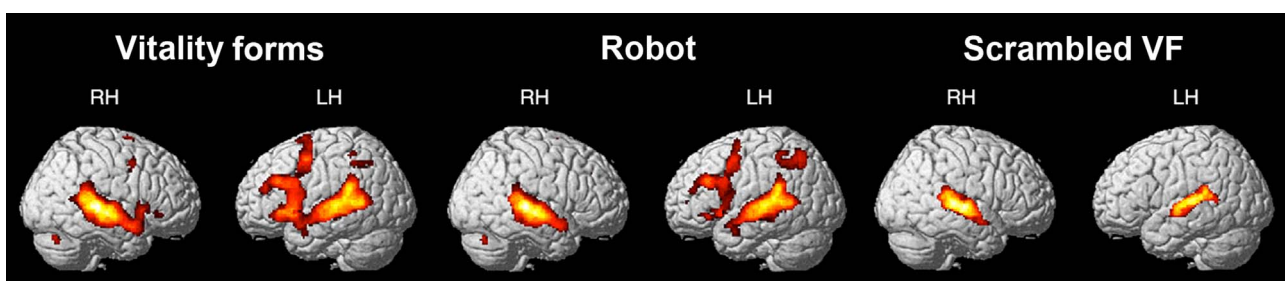


**Fig. 2.** Brain activations resulting from the Vitality, Robot, and Scrambled VF conditions *vs.* Silence. These activations are rendered into a standard Montreal Neurological Institute brain template ($P_{FWE} < 0.05$). LH, left hemisphere; RH, right hemisphere.
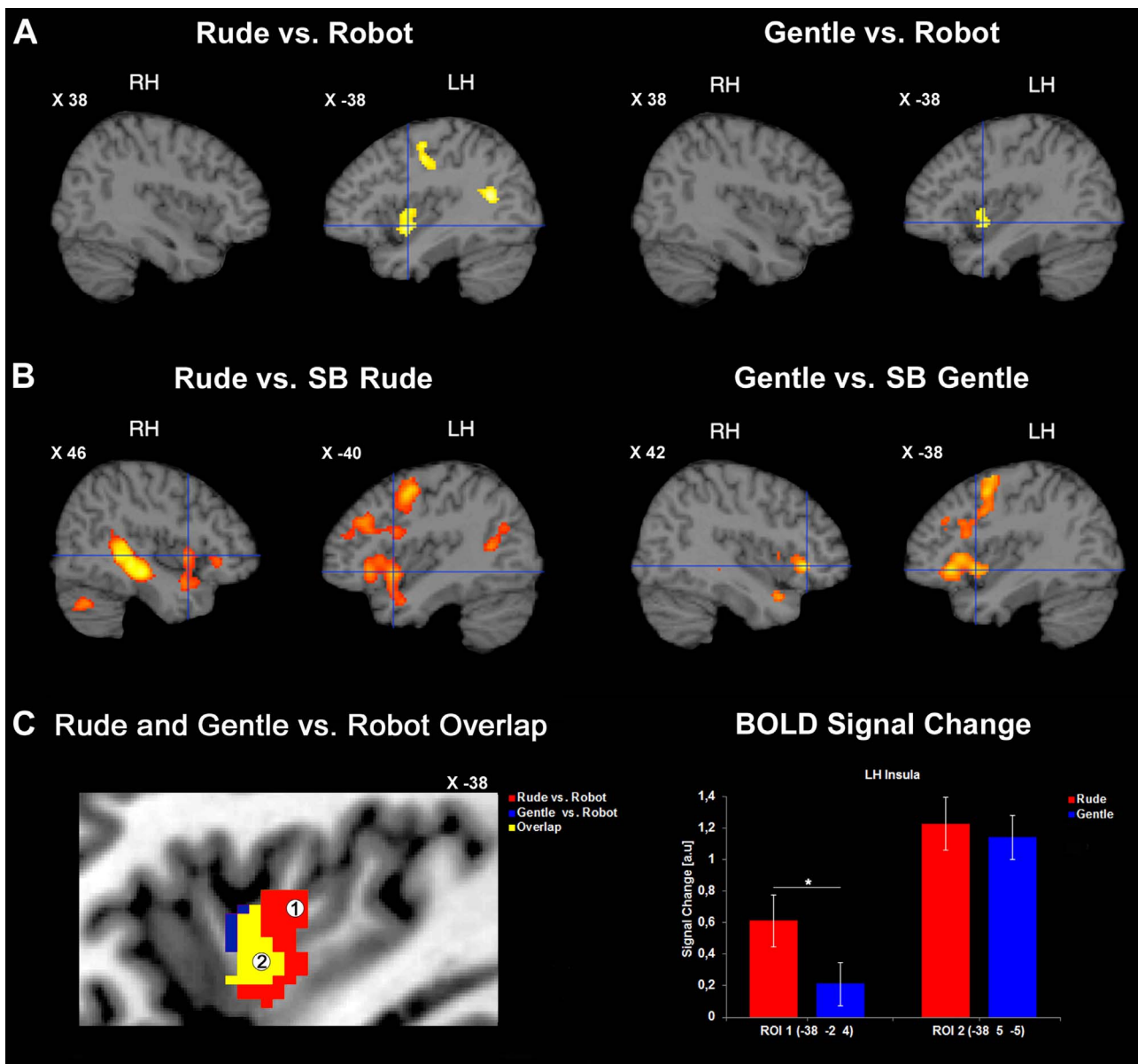
**Fig. 3.** Brain activations obtained during vitality form processing. (A) Parasagittal sections showing the activations resulting from the contrast *Rude vs. Robot* (left side) and *Gentle vs. Robot* (right side). (B) Parasagittal sections showing the activations resulting from the contrast *Rude vs. SB Rude* and *Gentle vs. SB Gentle*. (C) Overlap between insula areas resulting from the contrasts *Rude vs. Robot* and *Gentle vs. Robot* (left side). The dots indicate where the two ROIs were defined centering the sphere respectively in the dorsal part (ROI 1: x= −38, y= −2, z=4) and in the central part (ROI 2: x= −38, y=5, z= −5) of the central insula. BOLD Signal changes in two ROIs created in the central insula (right side). The bars indicate standard error of the mean (SEM). The horizontal line above the columns indicate the comparisons between rude and gentle vitality forms. Asterisk indicates significant difference at p < 0.05.

gentle vitality forms and to assess a possible difference between them, we overlaid the brain activations resulting from the contrasts *Vitality Rude vs. Robot* and *Vitality Gentle vs. Robot* (Fig. 3C, left side). *T*-test revealed a significant difference in the BOLD signal change between rude and gentle vitality forms in ROI 1 (p < 0.05, right side). In contrast, no significant difference was found between vitalities in the ROI 2.

*3.4. Conjunction analysis between vitality forms vs. robot and vitality forms vs. scrambled VF*

The conjunction analysis between the contrasts *Vitality forms vs. Robot (VF vs. Robot)* and *Vitality forms vs. Scrambled VF (VF vs. SB VF)* revealed a significant activation pattern in the left central part of the insula (Fig. 4A, left side; $P_{FWE} < 0.05$). This analysis highlights voxels activated in both contrasts (*Vitality forms vs. Robot*

*and Vitality forms vs. Scrambled VF).* After conjunction analysis, a region of interest (ROI) was created in the central insula (x= −38, y=6, z= −6) and the signal change values associated with vitality form conditions (*Rude* and *Gentle*) and control conditions (*SB Rude, SB Gentle, Robot*) were calculated for each subject on the basis of contrast images (Fig. 4A, right side). The BOLD signal change indicates an enhanced activation in the insula specific for listening to vitality forms. Note that this ROI was based on an overlap of the areas resulting from the main contrast analyses. In this ROI, to avoid the problem of circularity in the analysis, no statistical analysis was carried out (Kriegeskorte et al., 2009).

## 4. Discussion

In the present fMRI study we investigated the neural correlates of auditory vitality forms by presenting action verbs pronounced

**Table 2.**
Cerebral activity during the contrasts (A) *Rude vs. Robot*; (B) *Gentle vs. Robot*; (C) *Rude vs. SB Rude*; (D) *Gentle vs. SB Gentle*. Local maxima, as shown in Fig. 3, are given in MNI standard brain coordinates, significant threshold set at $P_{FWE} < 0.05$ (cluster-level).

| Anatomical region | Left hemisphere | | | | Right hemisphere | | | |
|---|---|---|---|---|---|---|---|---|
| | x | y | z | Z-score | x | y | z | Z-score |
| **(A) Rude vs. Robot** | | | | | | | | |
| Middle temporal gyrus | −54 | −56 | 16 | 3.88 | 62 | −58 | 8 | 4.51 |
| Posterior-Medial frontal gyrus | −12 | 0 | 68 | 4.60 | 4 | −4 | 64 | 4.07 |
| Precentral gyrus | −20 | −20 | 72 | 4.11 | | | | |
| Postcentral gyrus | −20 | −32 | 68 | 4.43 | | | | |
| Middle cingulate cortex | −14 | −8 | 42 | 4.02 | | | | |
| Putamen | −24 | 12 | 8 | 4.28 | | | | |
| Superior medial gyrus | −10 | 50 | 4 | 4.11 | | | | |
| Insula | −38 | 2 | −6 | 4.05 | | | | |
| **(B) Gentle vs. Robot** | | | | | | | | |
| Superior medial gyrus | | | | | 12 | 52 | 14 | 4.46 |
| Putamen | −26 | 12 | 6 | 4.09 | | | | |
| Insula | −38 | 6 | −6 | 3.78 | | | | |
| **(C) Rude vs. SB Rude** | | | | | | | | |
| Superior temporal gyrus | −56 | −44 | 22 | 5.37 | 64 | −42 | 22 | 4.41 |
| Middle temporal gyrus | −60 | −28 | 2 | 6.10 | 52 | −22 | −12 | 7.05 |
| Temporal pole | −50 | 12 | −2 | 5.48 | 52 | 8 | −18 | 4.95 |
| IFG (*pars orbitalis*) | −48 | 30 | −6 | 5.70 | 54 | 28 | −4 | 4.81 |
| IFG (*pars opercularis*) | −52 | 16 | 20 | 5.21 | | | | |
| Precentral gyrus | −38 | 0 | 62 | 4.25 | | | | |
| Posterior medial frontal gyrus | −4 | 12 | 52 | 4.99 | 0 | 6 | 60 | 5.16 |
| Middle frontal gyrus | −38 | 2 | 50 | 5.15 | | | | |
| Insula | −40 | 12 | −6 | 4.72 | 46 | 10 | 0 | 4.21 |
| Cerebellum | | | | | 40 | −66 | −32 | 4.55 |
| **(D) Gentle vs. SB Gentle** | | | | | | | | |
| Superior temporal gyrus | −56 | −44 | 22 | 3.76 | 62 | 0 | −10 | 5.14 |
| Middle temporal gyrus | −62 | −28 | 2 | 5.61 | 52 | −22 | −12 | 6.40 |
| IFG (*pars orbitalis*) | −44 | 28 | −10 | 5.45 | 42 | 32 | −4 | 5.05 |
| IFG (*pars opercularis*) | −50 | 12 | 10 | 5.34 | | | | |
| IFG (*pars triangularis*) | −54 | 20 | 20 | 5.29 | 56 | 20 | 0 | 4.48 |
| Precentral gyrus | −42 | 2 | 58 | 4.74 | | | | |
| Supramarginal gyrus | −58 | −40 | 34 | 3.56 | | | | |
| Posterior medial frontal gyrus | −10 | 8 | 70 | 3.57 | | | | |
| Insula | −38 | 12 | −6 | 4.49 | | | | |
| Temporal Pole | −52 | 12 | −16 | 4.56 | 52 | 10 | −18 | 4.92 |

in a rude or gentle way. The most important result of our study was the demonstration that, listening to action verbs, pronounced by a human voice, produces a selective activation of the middle and posterior short gyri of the left insula. This activation was larger when the words were pronounced in a rude way than gently. In contrast, it was absent during listening to the same verbs via the robot voice.

The activation of the central insula during listening to vitality forms cannot be merely accounted for by the physical properties of the stimuli used. In fact, the mean sound intensity and sound frequency were the same in the scrambled control condition as in the auditory vitality forms. Yet, this activation was completely absent in the insula when participants listened to scrambled stimuli. It is important to note that the robot voice intensity was also matched to the mean loudness value of the two vitality forms stimuli (rude and gentle). This control further excludes that the physical properties of the acoustic stimuli could explain the vitality form effect.

One can argue that the best control stimulus for vitality form could involve the same words pronounced by a human voice using a neutral tone. However, according to the definition of vitality form, a neutral human voice also contains a vitality form, which conveys a specific peaceful and quiet state. For this reason we used a robotic voice which, whilst conveying a verbal message, was devoid of any human vitality form content. In favor of this

conclusion are also our previous findings on action vitality forms (Di Cesare et al., 2015). In all tasks of that experiment (action observation, imagination and execution) the control condition, which consisted of simply placing a ball into a box or observing this action, produced activation of the central insula, although much weaker than in tasks endowed with a clear vitality form.

The results confirmed the validity of the robot voice control. In fact, listening to this stimulus activated the language areas and the motor circuit typically involved in listening to action verbs, but failed to activate the central part of insula involved in vitality form processing.

The insula activation in response to auditory stimuli endowed with vitality forms is in agreement with a previous study on vitality forms related to arm actions (Di Cesare et al., 2015). In this study participants observed, imagined, and performed actions with different vitality forms. In all these conditions there was an overlapping activation of the middle and posterior short gyri. There is, however, a difference between the previous and the present study. In the present study, the activation of the insula overlaid, in part, with that found during the observation and production of vitality forms related to arm actions, but also extended more ventrally (Fig. 4B). In addition, the listening to verbal material induced a left-dominant insula activation, in line with the dominance of the left hemisphere in language processing.

An interesting question is how the central insula, which is
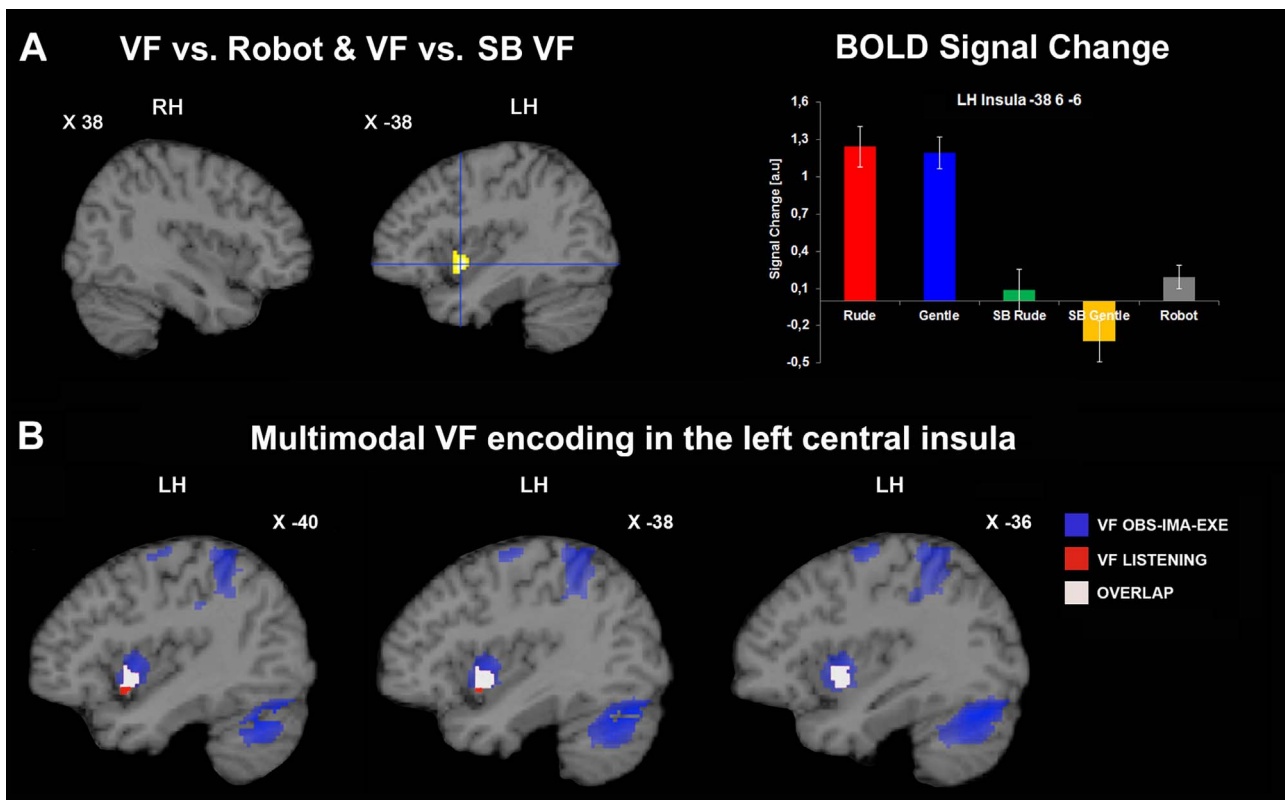
**Fig. 4.** Vitality form processing in the central part of the insula. (A) Brain activation resulting from the conjunction analysis of the contrasts *VF vs. Robot* [(Rude+Gentle) *vs.* Robot] and *VF vs. SB VF* [(Rude+Gentle) *vs.* (SB Rude+SB Gentle), left side]. Signal changes in the ROI created centering the sphere around the maxima (x=−38, y=6, z=−6) of the functional maps resulting from the conjunction analysis (right side). The bars indicate the standard error of the mean (SEM). For each direct contrast, the significance threshold was set at ($P_{FWE} < 0.05$) cluster level. (B) The figure indicates the areas found active in a previous experiment during the Observation, Imagination and Execution of the Rude and Gentle vitality forms (Di Cesare et al., 2015) and the activation of the insula obtained in the present study resulting from the conjunction analysis shown in the panel A, left side. The overlapping between the two experiments is shown in white color and corresponds to the central part of the insula. LH, left hemisphere; RH, right hemisphere.

clearly selective for vitality form recognition, may modulate the cortical areas that control the mouth, the pharyngeal, and the laryngeal movements. Is there an insulo-cortical circuit that may modulate speech production according to different vitality forms? Anatomical data in the monkey support the view that the pre-central opercular cortex (area PrCO; Roberts and Akert, 1963) controls mouth, pharyngeal, and laryngeal movements (*e.g.* Oga-wa, 1994; Martin and Sessle, 1993; Jürgens and Ehrenreich, 2007) and indeed a part of it, was defined by Simonyan and Jürgens (2005) as "laryngeal motor cortex".

The connectivity pattern of PrCO has been recently re-investigated by Gerbella et al. (2014) in monkeys. In particular they found, following injection in the rostral part of PrCO, a rich retrograde labeling in the central part of the insula. These data has been confirmed by Jezzini et al. (2015) by injecting tracers in different sectors of insula. Taken together, these data clearly indicate a connection between the central part of the insula and the oro-laryngeal representation in the opercular part of the frontal cortex. In line with these findings are the data from Remedios et al. (2009) who reported in a monkey study that the vocal communication sounds activate the central posterior part of the insula. This sector roughly corresponds to the part of the insula found to be active during vitality forms listening in our experiment.

In conclusion, our study shows that the key region for vitality form processing is the central sector of the insular cortex. During social interactions, this area is triggered not only by the vitality forms of observed actions but also by listening to action verbs endowed with rude and gentle vitality forms.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2016.06.017.

## References

Di Cesare, G., Di Dio, C., Marchi, M., Rizzolatti, G., 2015. Expressing and under-standing our internal states and those of others. Proc. Natl. Acad. Sci. USA 112 (33), 10331–10335.

Di Cesare, G., Di Dio, C., Rochat, M., Sinigaglia, C., Bruschweiler, N., Stern, D.N., Rizzolatti, G., 2013. The neural correlates of "vitality form" recognition: an fMRI study. Soc. Cognit. Affect. Neurosci. 9, 951–960.

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.

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

Gerbella, M., Borra, E., Rozzi, S., Luppino, G., 2014. Connections of the macaque

Granular Frontal Opercular (GrFO) area: a possible neural substrate for the contribution of limbic inputs for controlling hand and face/mouth actions. Brain Struct. Funct. . http://dx.doi.org/10.1007/s00429-014-0892-8

Jezzini, A., Rozzi, S., Gallese, V., Caruana, F., Gerbella, M., 2015. A shared neural network for emotional expression and perception: an anatomical study in the macaque monkey. Front. Behav. Neurosci. 9, 243.

Jürgens, U., Ehrenreich, L., 2007. The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. Brain Res. 1148, 90–95.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540.

Martin, R.E., Sessle, B.J., 1993. The role of the cerebral cortex in swallowing. Dysphagia 8, 195–202.

Marwick, H., Murray, L., 2009. The effects of maternal depression on the "musicality" of infant-directed speech and conversational engagement. In: Malloch, S., Trevarthen, C. (Eds.), Communicative Musicality. Oxford University Press, New York, pp. 281–300.

Ogawa, H., 1994. Gustatory cortex of primates: anatomy and physiology. Neurosci. Res. 20, 1–13.

Remedios, R., Logothetis, N.K., Kayser, C., 2009. Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. Proc. Natl. Acad. Sci. USA 106 (42), 18010–18015.

Roberts, T., Akert, K., 1963. Insular and opercular cortex and its thalamic projections in Macaca mulatta. Sweiz Arch. Neurol. Neurochir. Psychiatr. 92, 1–43.

Simonyan, K., Jürgens, U., 2005. Afferent subcortical connections into the motor cortical larynx area in the rhesus monkey. Neuroscience 130, 119–131.

Stern, D.N., 1985. The Interpersonal World of the Infant. Basic Books, New York.

Stern, D.N., 2010. Forms of Vitality Exploring Dynamic Experience in Psychology, Arts, Psychotherapy, and Development. Oxford University Press, United Kingdom.

Trevarthen, C., 1998. The concept and foundations of infant intersubjectivity. In: Braten, S. (Ed.), Intersubjective Communication and Emotion in Early Ontogeny. Cambridge University Press, New York.

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