



Processing pathways for emotional vocalizations

Tiffany Grisendi¹ · Olivier Reynaud² · Stephanie Clarke¹ · Sandra Da Costa²

Received: 3 April 2019 / Accepted: 12 June 2019 / Published online: 6 July 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Emotional sounds are processed within a large cortico-subcortical network, of which the auditory cortex, the voice area, and the amygdala are the core regions. Using 7T fMRI, we have compared the effect of emotional valence (positive, neutral, and negative) and the effect of the type of environmental sounds (human vocalizations and non-vocalizations) on neural activity within individual early stage auditory areas, the voice area, and the amygdala. A two-way ANOVA was applied to the BOLD time course within each ROI. In several early stage auditory areas, it yielded a significant main effect of vocalizations and of valence, but not a significant interaction. Significant interaction as well as significant main effects of vocalization and of valence were present in the voice area; the former was driven by a significant emotional modulation of vocalizations but not of other sounds. Within the amygdala, only the main effect of valence was significant. Post-hoc correlation analysis highlighted coupling between the voice area and early stage auditory areas during the presentation of any vocalizations, and between the voice area and the right amygdala during positive vocalizations. Thus, the voice area is selectively devoted to the encoding of the emotional valence of vocalizations; it shares with several early stage auditory areas encoding characteristics for vocalizations and with the amygdala for the emotional modulation of vocalizations. These results are indicative of a dual pathway, whereby the emotional modulation of vocalizations within the voice area integrates the input from the lateral early stage auditory areas and from the amygdala.

Keywords Human vocalizations · Emotions · Auditory belt areas · Amygdala · Voice area · 7T fMRI

Abbreviations

AI	Primary auditory area
AMY	Amygdala
HVN	Human vocalizations with negative emotional valence
HVP	Human vocalizations with positive emotional valence
HVO	Human vocalizations with neutral emotional valence
NVN	Non-vocalizations with negative emotional valence
NVP	Non-vocalizations with positive emotional valence

NV0	Non-vocalizations with neutral emotional valence
R	Rostral (primary) auditory area
VA	Voice area

Introduction

The universal nature of human emotional vocalizations and the prominent role that they play in shaping social interactions (Sauter et al. 2010) warrant the highly specialized processing, which have been described in a series of seminal studies.

The emotional valence of human vocalizations is processed by complex cortico-subcortical networks, which include the primary auditory cortex, the surrounding early stage and higher order auditory areas, the inferior frontal gyrus and the amygdala, as well as the medial frontal cortex, the insula, basal ganglia, and the cerebellum (Schirmer and Kotz 2006; Frühholz et al. 2016). As highlighted in a recent review (Frühholz et al. 2016), the amygdala, the auditory areas, including the voice area, and the inferior frontal cortex have strong functional connections and encode not only the

✉ Sandra Da Costa
sandra.dacosta@epfl.ch

Stephanie Clarke
Stephanie.Clarke@chuv.ch

¹ Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, 1011 Lausanne, Switzerland

² Centre d'Imagerie Biomédicale, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

emotional valence of non-verbal vocalizations, but also of other sound categories (e.g., speech prosody: Wildgruber et al. 2009; music: Koelsch 2010).

The *amygdala* (AMY) has been repeatedly shown to be involved in the process of emotional stimuli of different sensory modalities, as demonstrated by fMRI and PET studies (Baas et al. 2004; Costafreda et al. 2008; Ball et al. 2009) and by intracranial recordings (for review Murray et al. 2014). In the auditory modality, emotional modulation was investigated for human vocalizations (Morris et al. 1999; Sander and Scheich 2005; Ethofer et al. 2006a, b; Viinikainen et al. 2012; Pannese et al. 2016), including emotional prosody (reviews: Wildgruber et al. 2006; Liebenenthal et al. 2016); a mixture of human vocalizations and environmental sounds (Viinikainen et al. 2012); or instrumental music (reviews: Koelsch 2010; Frühholz et al. 2014). Cytoarchitecturally AMY is subdivided into three major nuclei groups, the lateral basal, centromedial, and superficial (Amunts et al. 2005). As demonstrated in non-human primate and non-primate species, it receives auditory input via monosynaptic afferents from the medial geniculate nucleus (Ottersen and Ben-Ari 1979; Russchen 1982; LeDoux et al. 1985; Shinonaga et al. 1994) and by reciprocal interconnections with parts of the auditory cortex. The latter involve in non-human primates predominantly the non-primary auditory areas and cortical regions on the postero-superior part of the temporal convexity (Price and Amaral 1981; Yuki 2002); in some species, part of the primary auditory cortex was found to be involved as well (Reser et al. 2009). Among other wide-spread connections, AMY has also reciprocal connections with the prefrontal cortex (Ghashghaei and Barbas 2002; Barbas 2007) and with the mediodorsal nucleus of the thalamus (Russchen et al. 1987). Diffusion tensor imaging in humans revealed fibre tracts between AMY and medial geniculate nucleus-related fibre tracts that are compatible with the connectivity described in non-human species (Keifer et al. 2015; Kamali et al. 2016). As shown in rodents, distinct amygdala circuits process fearful and rewarding stimuli and modulate, via specific outputs, autonomic reactions (Janak and Tye 2015).

The voice area (VA) was defined in the pioneering study of Belin and colleagues (Belin et al. 2000) by its stronger responses to human than animal vocalizations. It is located in the middle part of the superior temporal gyrus and sulcus. A follow-up study using the same stimulus set showed in 218 subjects that the voice-sensitive region is mostly bilateral (94% of subjects). It confirmed the location within the posterior part of the superior temporal sulcus and on the adjacent part of the convexity of the superior temporal gyrus; the rostral extension of VA was shown to reach up to the (lower) lip of the sylvian fissure and to encroach on the lateral brim of the supratemporal plane. The authors reported great inter-individual variability as to its precise anatomical location

(Pernet et al. 2015). The part of the superior temporal gyrus and sulcus, where VA is located, receives auditory input from the primary auditory cortex via a cascade of cortico-cortical connections (Cammoun et al. 2015). In this respect, it is similar to other auditory-processing pathways in human (e.g., Kim and Knösche 2016) and bears strong similarity to the homologous region in non-human primates (e.g., a recent comprehensive study: Scott et al. 2015). VA activity was reported to be modulated by the emotional value of vocalizations (Belin et al. 2002; Grandjean et al. 2005; Ethofer et al. 2006b, 2008, 2009b; Beaucousin et al. 2007a; Obleser et al. 2007, 2008; Bestelmeyer et al. 2017). The emotional voice area has been identified by its stronger response to emotional than neutral pseudo-sentences; it overlaps partially with VA and extends beyond it to the middle part of the superior temporal gyrus, Heschl's gyrus, and the antero-lateral part of the planum temporale (Ethofer et al. 2012). Thus, in addition to VA, the auditory cortex on the supratemporal plane participates significantly in the encoding of emotional vocalizations (Wildgruber et al. 2004a; Meyer et al. 2005; Dietrich et al. 2007, 2008; Leitman et al. 2010; Szameitat et al. 2010; Ethofer et al. 2012). Although multiple subregions of the superior temporal cortex are involved (Frühholz and Grandjean 2013), none of the previous studies analysed individual auditory areas, most likely because of limitations imposed by low spatial resolution. From the above studies, eight used 1.5 T (Belin et al. 2000, 2002; Wildgruber et al. 2004b; Grandjean et al. 2005; Ethofer et al. 2006b, 2008; Beaucousin et al. 2007b; Szameitat et al. 2010) and nine 3 T fMRI (Dietrich et al. 2007; Ethofer et al. 2008, 2009a, 2012; Obleser et al. 2008; Leitman et al. 2010; Arnal et al. 2015; Bestelmeyer et al. 2017; Lavan et al. 2017), often with a voxel size of $3 \times 3 \times 3$ mm or more (11 studies) and smoothing of more than 8 mm (12 studies). In view of the anatomical evidence described below, this type of spatial resolution does not allow to analyse individual early stage auditory areas.

The supratemporal plane comprises several early stage auditory areas, as demonstrated in histological studies (Clarke and Morosan 2012). Investigating them with fMRI represents a challenge on three accounts. First, early stage auditory areas tend to be relatively small (40–310 mm²; (Clarke and Morosan 2012) and their investigation requires high spatial resolution. Second, the realignment of the supratemporal plane tends to be imprecise and landmarks, such as Heschl's gyrus, have been shown to be shifted by as much as 4 mm between individual brains of a group study (Viceic et al. 2009). Thus, a whole brain contrast aligns in a group study regions, which do not correspond across subjects to the same area. To circumvent this problem, several studies used a functional marker for the primary auditory cortex and based its identification in each individual subject on tonotopic mapping (with ultra-high field: Formisano

et al. 2003; Da Costa et al. 2011; Moerel et al. 2014), or by approximating it with Heschl's gyrus (Zilles et al. 1988; Rademacher et al. 2001; Viceic et al. 2006; van der Zwaag et al. 2011). Third, no reliable functional marker is available for individual non-primary auditory areas; they tend to be broadly tuned to complex features such as frequency, pitch, amplitude modulation, or envelop (Hall et al. 2002; Rauschecker and Scott 2009; Chevillet et al. 2011). Their characterization relies on histological criteria and was carried out in post-mortem material (Rivier and Clarke 1997; Clarke and Rivier 1998; Hackett et al. 2001; Wallace et al. 2002; Chiry et al. 2003). Their identification in activation studies can be based on Talairach coordinates published in histological studies (Viceic et al. 2006; van der Zwaag et al. 2011), preferably in combination with tonotopic mapping for the localization of the primary auditory cortex (Da Costa et al. 2015, 2018).

The inferior frontal cortex contributes to the cognitive evaluation of emotional cues of verbal and non-verbal vocalizations and its modulation by attention (review Frühholz and Grandjean 2013). The putatively homologous area in non-human primates, the lateral prefrontal cortex, receives relatively sparse afferents from AMY; this contrasts with the strong, bidirectional connections, which AMY has with the orbitofrontal and medial prefrontal areas (Ghashghaei and Barbas 2002; Barbas 2007; Barbas et al. 2011).

In summary, the early stage auditory areas, VA, and AMY constitute a core network for the process of emotional vocalizations (Frühholz et al. 2016), which is fostered by strong connections, as demonstrated in hodological studies in non-human primates and in man. The primary auditory cortex and AMY receive both direct auditory inputs from the medial geniculate nucleus (Shinonaga et al. 1994). A complex pattern of cortico-cortical connections links the primary and non-primary early stage areas and the adjacent superior temporal convexity (Cammoun et al. 2015). The extended auditory region is interconnected with AMY (Price and Amaral 1981; Yukie 2002; Reser et al. 2009). On the basis of this complex architecture, and notably a dual auditory input via the primary auditory cortex and via AMY, it can be argued that the process of emotional vocalizations may differ between the early stage auditory areas, VA, and AMY, possibly with different selectivities in respect to that of other emotional sounds.

Here, we made use of the high spatial resolution of ultra-high field fMRI at 7 T to investigate the representation of human vocalizations vs. other environmental sounds, and their modulation by emotional valence within early stage auditory areas, VA, and AMY. Based on previous findings, we expected (1) the AMY to process emotional valence both for vocalizations and non-vocalizations; (2) specific auditory belt areas to encode specifically human vocalizations or emotional valence, but not emotional valence of human

vocalizations only; and (3) VA and/or AMY to process emotional valence selectively for human vocalizations. These hypotheses were tested by comparing the BOLD responses within the above regions of interest to human vocalizations and to other environmental sounds with positive, neutral, or negative emotional valence using various repeated measures ANOVA. In addition, we explored functional coupling between individual early stage auditory areas, VA and AMY, expecting to find a signature of the dual auditory input via the primary auditory cortex and the amygdala.

Materials and methods

Subjects

Eleven subjects (7 female, 8 right-handed, mean age 25.3 ± 4.27 years) participated in the study after giving written, informed consent. None of the participants reported hearing deficits or history of neurological or psychiatric illness. Hearing thresholds and mental states were measured prior to testing. All participants were native speakers of French, without musical training. All experimental procedures were approved by the Ethics Committee of the Canton de Vaud. The data set of one subject was discarded due to data acquisition problems (ghosting), and data from the remaining ten subjects were used in the following analysis.

Participants provided informative health status and then completed five questionnaires: the Edinburgh Handedness Inventory (Oldfield 1971), the Hospital Anxiety and Depression (HAD) scale (Zigmond and Snaith 1983), the Big-Five Inventory (Plaisant et al. 2010), and a musical aptitude questionnaire. The participants were representative of the general population. The results of the Big-Five Inventory showed that the N (Neuroticism) score had greater between-subject variability than the other scores. The scores for factors A (Agreeableness), C (Conscientiousness), and E (Extraversion) in our subject sample were higher than in the sample from (Plaisant et al. 2010), whereas the N score was smaller and had greater variability. The distribution of the O (Openness) score of the current study was similar to that of (Plaisant et al. 2010). Concerning the HAD scale, the between-subject variability was greater for the anxiety score compared to the depression score. Despite this score variability, no subjects were excluded based on these results.

Experimental design and statistical analysis

The experimental design included a single fMRI session (~55–60 min in total) during which participants listened passively (i.e., without performing a task) to human vocalizations or other environmental sounds with positive, neutral, or negative emotional valence while fixating on a red cross

on a black background. The baseline condition was resting silently with the same fixation. The fMRI session was followed by a rating of the emotional valence of the 66 stimuli used in the experiment. A debriefing was then performed outside the MRI scanner.

Auditory stimuli were presented in blocks of 11 different sounds from the same category (human vocalizations or other environmental sounds) and with the same emotional valence (positive, neutral, or negative). On the whole, six stimulus conditions were presented: (1) human vocalizations with neutral valence (vowels or consonant vowels without significance); (2) human vocalizations with positive valence (e.g., baby or adult laughing and erotic vocalizations by man or woman); (3) human vocalizations with negative valence (e.g., frightened scream and vomiting, brawl); (4) non-vocalizations with neutral valence (e.g., running car engine, wind blowing, and train); (5) non-vocalizations with positive valence (e.g., applause, opening beer can, and pouring into the glass; river); and (6) non-vocalizations with negative valence (e.g., ticking and exploding bomb; tire skids, breaking glass). Each subject listened to three runs, in which blocks and their sequence order were pseudo-randomized. Each fMRI run began with a 30-s silent rest condition, followed by 14 blocks, each of which lasted 30 s (22 s of sounds + 8 s of silence), followed again by a 30-s silent rest condition (total of 8 min). Sounds were presented using MATLAB (R2015b, The MathWorks, Inc., Natick, Massachusetts, United States) and the Psychophysics Toolbox (www.psychtoolbox.org). Stimuli were delivered binaurally at 80 ± 8 dB SPL via MRI-compatible headphones (SensiMetrics S14, SensiMetrics, USA), following prior filtering with the SensiMetrics filters to obtain a flat frequency transmission. The auditory stimuli used in this experiment were the same as in (Aeschlimann et al. 2008), who showed in their study that human vocalizations is a separate category within the environmental sounds. In this battery, 66 different emotional sound files of 2 s were selected and equally distributed in the following six categories: Human Vocalizations Positive (HVP), Human Vocalizations Neutral (HV0), Human Vocalizations Negative (HVN), Non-vocalizations Positive (NVP), Non-vocalizations Neutral (NV0), and Non-vocalizations Negative (NVN). Categories were controlled for their acoustic characteristics: the percentage of points showing a significant difference between the mean spectrogram of two different sound categories was calculated and maintained below 1% to avoid acoustic differences between the six categories of sound, as in (De Meo et al. 2015). All the sounds were also tested using the PRAAT software (<http://www.fon.hum.uva.nl/praat/>) and homemade MATLAB scripts to determine their mean fundamental frequency, mean intensity, harmonics-to-noise ratio, power, center of gravity, mean Wiener entropy, and spectral structure variation (Reddy et al. 2009). Two-way repeated measure

ANOVA with the factors Vocalization (Human Vocalizations, Non-vocalizations) \times Valence (Positive, Neutral, Negative) were performed to compare the effect of each acoustic feature on the sound categories. We found a main effect of Vocalization [$F(1,64) = 18.68, p = 0.0015$], a main effect of Valence [$F(2,63) = 21.14, p = 1.17E - 5$] and an interaction Vocalizations \times Valence [$F(2,63) = 8.28, p = 0.002$] on the mean Wiener entropy. We found a main effect of Valence [$F(2,63) = 10.51, p = 0.0007$] on the center of gravity. There was a main effect of Vocalizations [$F(1,64) = 134.23, p = 4.06E - 7$], a main effect of Valence [$F(2,63) = 69.61, p = 9.78E - 10$] and an interaction of Vocalizations \times Valence [$F(2,63) = 17.91, p = 3.48E - 5$] on the harmonics-to-noise ratio. Finally, there was an interaction of Vocalizations \times Valence on the mean intensity [$F(2,63) = 12.47, p = 0.0003$] and on the power [$F(2,63) = 14.77, p = 0.0001$].

The post-acquisition rating of the emotional valence of each stimulus was performed, while the subject was still lying in the scanner (without any sequence running) to minimize emotional bias and to match at best the experimental conditions. Stimuli were presented in random order; after each sound presentation, the subject was instructed to judge the valence of the sound with a linear visual 7-point scale (1 being very pleasant and 7 being very unpleasant) and to give their answer orally within a 5 s silent gap.

Tonotopic mapping was achieved by presenting each subject pure tones (88–8000 Hz, in half octave steps) in ordered progressions for 2 s, as described previously (Da Costa et al. 2011, 2013, 2015, 2018). A cycle was composed of 28 s of tone presentation (14 frequencies \times 2 s) followed by a 12-s silent pause. A single fMRI run consisted of 12 identical cycles for a total duration of 8 min. Each subject listened to two runs, either with ascending or descending frequency progressions. One subject (the pilot of the study) had a different mapping paradigm (but comparable tonotopic maps), with seven tones (88–8000 Hz, in octave steps) and 12 cycles (composed of 14 s of tone presentation and 14 s of silent pause).

The identification of the regions of interest (ROIs) was performed as follows. First, the identification of the early stage auditory areas was carried out as described previously (Da Costa et al. 2015, 2018). Briefly, individual tonotopic mappings were used to identify in each subject the primary and non-primary areas, which were designated as the primary auditory areas (A1 and R), the lateral belt areas (L1, L2, L3, and L4), and medial belt areas (M1, M2, M3, and M4). These individually defined areas were used as ROIs for the analysis of neural activity (Fig. 1a, orange box, Table 1), and their respective coordinates were in accordance with previously published values (Viceic et al. 2006; van der Zwaag et al. 2011; Da Costa et al. 2015, 2018). Second, VA was identified by the contrast ‘Human Vocalizations vs. Non-vocalizations’ ($p = 0.001$, uncorrected). All significant

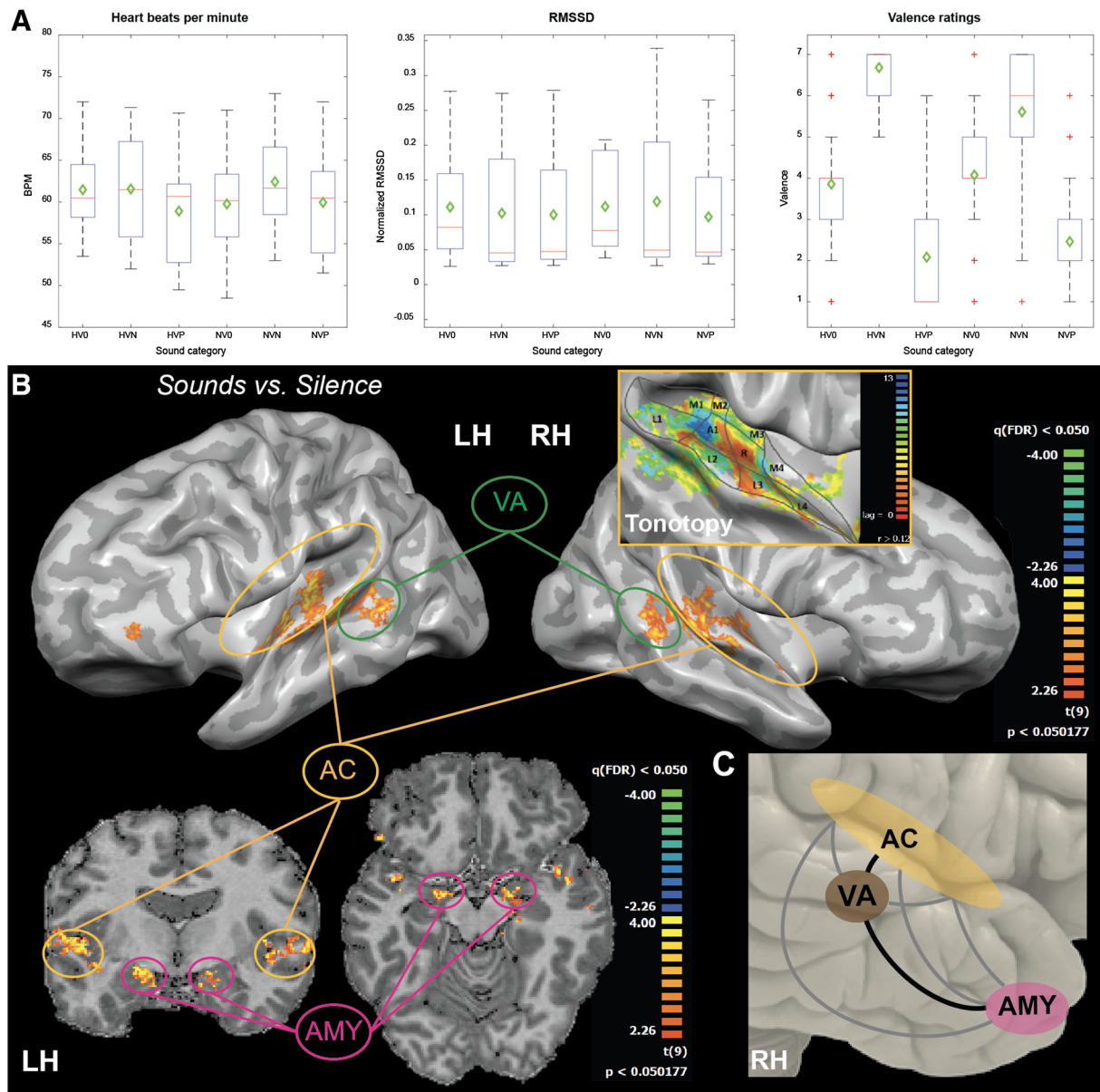


Fig. 1 **a** Behavioral results. Left panel: heart rate (beats per minute). Middle panel: root mean squared successive difference (RMSSD). Right panel: valence ratings. The red line represents the median value, the box indicates the values between the 25th and the 75th percentiles, and the whiskers show the data below the 25th percentile or above the 75th percentile, not considered as outliers. The outliers are depicted with a plus symbol. The green diamond represents the mean value. Refer to the manuscript for the sound categories abbreviations. **b** GLM statistical maps and ROI definition. The maps resulting from the contrast ‘Sounds vs. Silence’ are presented in the lower panel on the volume in the coronal and transverse views and in the upper panel on the surface (FDR corrected, $q < 0.05$). Orange box: tonotopic maps were projected into the individual right hemisphere surface of

a representative subject ($r > 0.12$). The frequency-selective region was divided into 10 ROIs: A1, R, L1, L2, L3, L4, M1, M2, M3, and M4. **c** Hypothesis. We expected (1) the AMY to process emotional valence both for vocalizations and non-vocalizations and to be highly correlated with VA and specific lateral belt auditory areas (within AC, L1, L2, and L3; grey lines); (2) specific auditory belt areas to encode specifically human vocalizations or emotional valence but not emotional valence of human vocalizations only and to be correlated to each other; and (3) VA and/or AMY to process emotional valence selectively for human vocalizations and being modulated by lateral belt auditory areas (within AC, L2; dark lines). *RH* right hemisphere, *LH* left hemisphere. Refer to the manuscript for the ROIs definition and abbreviations

voxels on the posterior part of the superior temporal gyrus and sulcus were considered as part of VA. The Talairach coordinates of VA defined in this way were well within the

general regions of VA as described by Pernet et al. (2015). For some subjects, a few voxels were located within the lateral early stage auditory areas; they were not considered as

Table 1 Mean Talairach coordinates (center of gravity) of all ROIs and mean areas

ROI	X ± STD(X)	Y ± STD(Y)	Z ± STD(Z)	Area ± STD(area) [mm ³]
Left hemisphere				
Amy	− 21.79 ± 4.32	− 4.62 ± 3.06	− 14.41 ± 3.42	1240 ± 218.96
A1	− 41.85 ± 4.77	− 25.63 ± 4.82	10.71 ± 3.87	781.1 ± 113.52
R	− 38.68 ± 4.18	− 20.58 ± 4.77	10.04 ± 3.86	736.2 ± 147.44
L1	− 53.67 ± 5.02	− 35.60 ± 8.63	16.97 ± 8.19	2204 ± 491.04
L2	− 54.40 ± 5.19	− 18.26 ± 5.00	9.22 ± 4.00	794.2 ± 89.6
L3	− 48.86 ± 5.46	− 8.73 ± 5.80	5.23 ± 3.51	1193 ± 286.6
L4	− 42.96 ± 4.16	− 0.63 ± 9.85	− 7.21 ± 6.98	2113 ± 305.24
M1	− 44.50 ± 6.44	− 35.80 ± 5.31	20.71 ± 7.75	1723 ± 352.72
M2	− 32.94 ± 2.71	− 31.42 ± 3.17	16.75 ± 2.87	254.7 ± 65.1
M3	− 30.05 ± 1.89	− 27.67 ± 3.15	16.89 ± 3.46	182.4 ± 55.08
M4	− 34.54 ± 3.19	− 11.19 ± 9.84	− 2.39 ± 10.71	1698 ± 237.5
VA	− 53.30 ± 6.21	− 30.36 ± 5.72	6.13 ± 4.43	435.4 ± 202.36
Right hemisphere				
Amy	20.27 ± 4.36	− 5.05 ± 2.84	− 14.57 ± 3.42	1259 ± 180.96
A1	43.24 ± 4.83	− 26.09 ± 5.00	11.79 ± 3.71	678.3 ± 110.16
R	40.24 ± 4.24	− 20.24 ± 4.96	8.71 ± 4.29	745.9 ± 141.28
L1	55.52 ± 4.80	− 31.55 ± 5.86	19.18 ± 9.55	1903 ± 436.44
L2	56.71 ± 4.23	− 21.04 ± 6.27	9.50 ± 4.49	922.5 ± 195.7
L3	52.15 ± 5.07	− 9.87 ± 5.78	3.67 ± 3.56	1003 ± 174.8
L4	43.62 ± 4.71	0.31 ± 10.10	− 7.29 ± 6.39	2380 ± 343.2
M1	44.73 ± 6.25	− 32.77 ± 5.04	24.99 ± 8.42	1477 ± 144.36
M2	33.25 ± 3.11	− 30.77 ± 4.48	17.78 ± 4.36	236.1 ± 35.3
M3	31.47 ± 2.30	− 26.75 ± 3.97	16.23 ± 4.21	199.3 ± 43.56
M4	34.95 ± 2.92	− 10.74 ± 10.96	− 3.30 ± 10.32	1844 ± 279.36
VA	48.79 ± 7.60	− 31.39 ± 7.37	5.46 ± 4.98	592.3 ± 222.56

STD standard deviation

belonging to VA but to the respective lateral early stage auditory areas in further analysis. Third, AMY was identified in each subject on the anatomical images using BrainVoyager (BrainVoyager QX v2.8, Brain Innovation, Maastricht, The Netherlands) drawing tools. Given that the same sounds are used for the definition of VA and for the following statistical analysis, the results for VA region are only descriptive, except for the post-hoc correlation analysis.

MRI data acquisition was performed on a 7T MRI scanner (Siemens MAGNETOM scanner, Siemens Medical Solutions) with an 8-channel head rf-coil (RAPID Biomedical). To acquire high spatial resolution data sets, a sinusoidal 2D-EPI sequence with 1.5 mm isotropic voxels was used for the functional acquisition (1.5 × 1.5 mm in-plane resolution, slice thickness = 1.5 mm, TR = 2000 ms, TE = 25 ms, flip angle = 70°, slice gap = 0 mm, matrix size = 146 × 146, and field of view = 219 × 219, with 43 oblique slices centered on the superior temporal plane along the lateral sulcus, with a total coverage of ~65 mm and covering the full extent of the superior and medial temporal sulci until the entorhinal cortex). T1-weighted high-resolution 3D anatomical images were acquired with an MP2RAGE sequence

(resolution = 1 × 1 × 1 mm³, TR = 5500 ms, TE = 1.87 ms, T1/TI2 = 750/2350 ms, slice gap = 0 mm, matrix size = 256 × 240, and field of view = 256 × 240 (Marques et al. 2010).

The process of emotional stimuli in AMY has been shown to modulate autonomic reactions, via specific outputs to brain stem nuclei (Janak and Tye 2015). Emotional stimuli of different valence can thus induce distinct changes in heart rate or breathing and introduce a bias to image analysis. We accounted for this in our study and recorded during the experiment, pulse oximetry, and respiration, using a plethysmograph and respiratory belt provided from the MRI scanner vendor. In total, each subject had an imaging session of 54 min with five functional runs: three runs of the auditory emotional experiment and two runs of the tonotopic mapping experiment, which were used for the definition of the ROIs within the superior temporal plane.

The MRI analysis included the following steps. Preprocessing steps included scan time correction (only for the auditory emotional runs), temporal filtering, motion correction, segmentation, and normalization into the Talairach space and were performed with BrainVoyager. These preprocessing steps were common to all fMRI acquisitions, then

depending on the purpose of the fMRI run, the data sets were processed differently. To define early stage auditory areas, a linear cross-correlation analysis was computed for each tonotopic mapping data set, and the resulting correlation maps were averaged together (ascending and descending correlation map) to define the best frequency value for each voxel according to the cycle order (Da Costa et al. 2011, 2013, 2015, 2018). These analyses were performed at the single subject level in the volumetric space, and the resulting maps were projected onto the cortical surface meshes, where the individual early stage auditory areas and VA were defined and projected back to the volumetric space (see the paragraph on “the identification of the regions of interest (ROIs)” for more details). Then, we performed a random effects (RFX) group analysis on the auditory emotional runs, with movement and respiration parameters as regressors, and tested for the contrast ‘Sounds vs. Silence’ with an FDR correction at $q < 0.05$ ($p < 0.05$, Fig. 1b). This GLM analysis was used to verify that specific regions of the brain were activated by our paradigm and particularly that our ROIs were activated by the emotional sounds. The remaining analysis focused only on the BOLD responses extracted from these ROIs, as explained in the next paragraphs.

The BOLD time course extraction and processing were performed as follows. Functional individual time courses for each ROI were extracted in the 3D volume space using BrainVoyager and imported into MATLAB. Each time course was normalized by its mean signal, separated according to the sound category, and averaged (1) spatially within each ROI, (2) temporally over blocks and runs, and (3) across the ten subjects, resulting in a time course of 15 timepoints for each ROI and category. A timepoint-by-timepoint two-way repeated measure ANOVA, 2 vocalization (human vocalizations, non-vocalizations) \times 3 valence (positive, neutral, negative) was performed on the averaged BOLD time courses according to (Da Costa et al. 2015, 2018). Significant results were restricted temporally by only considering the p values lower or equal to 0.05 for at least three consecutive timepoints. It is to be noted that the probability that three consecutive timepoints be false positives is $(0.05 \times 0.05 \times 0.05) = 1.25 \times 10^{-4}$. Therefore, this constraint in time was considered as a valid correction for our analysis (see Supporting Information of Da Costa et al. 2015 for more details). Finally, post-hoc timepoint-by-timepoint paired t tests were performed between each pair of sound categories.

Physiological data were processed with the TAPAS PhysIO toolbox (Kasper et al. 2017). The respiration recordings were used as regressors in the GLM model, whereas the cardiac recordings were processed with the same pipeline as the BOLD signal to obtain a pulse time course for each sound category. Heart rate, interbeat interval (time interval between two successive beats) time courses, and heart rate variability were also extracted from these data. The heart

rate variability was calculated using the root mean squared successive difference between the interbeat interval, normalized by the mean interbeat interval according to (Goedhart et al. 2007). These latter measures were used to evaluate the effect of the emotional content of the auditory stimuli on the cardiac rhythm.

Results

Modulation of behavioral measures by emotional and vocal contents

The average heart rate did not show any significant differences between sound categories (Fig. 1a, left). Heart rate variability, represented by the normalized root mean squared successive difference (RMSSD), did not show any significant difference between sound categories (Fig. 1a, middle). The time courses of the pulses were submitted to a timepoint-by-timepoint two-way repeated measure ANOVA 2 Vocalization (Human Vocalizations, Non-vocalizations) \times 3 Valence (Positive, Neutral, Negative), which highlighted a main effect of Vocalization.

Post-scanning ratings of the valence of the sound stimuli showed a bigger variance in the categories HVP and NVN compared to that in the four other sound categories (Fig. 1a, right). A two-way repeated measure ANOVA 2 Vocalization (Human Vocalizations, Non-vocalizations) \times 3 Valence (Positive, Neutral, Negative) on the valence ratings revealed a main effect of Valence [$F(2, 653) = 532.29$, $p = 7.38E - 138$], no effect of Vocalization [$F(1, 653) = 2.68$, $p = 0.1$], and an interaction of Vocalization \times Valence [$F(2, 653) = 22.31$, $p = 4.23E - 10$]. As indicated by post-hoc t tests, the latter was driven by the difference between vocalizations vs. non-vocalizations, which was significant for negative, but not neutral or positive valence.

Emotional modulation of neural activity elicited by human vocalizations and by other environmental sounds

The RFX GLM analysis with the contrast ‘Sounds vs. Silence’ [$p < 0.005$, $q(\text{FDR}) < 0.05$] resulted in a strong bilateral activation on the supratemporal plane, the posterior part of the superior temporal sulcus corresponding to VA and in the AMY (Fig. 1b and Table 2).

Two-way repeated measure ANOVA on the BOLD responses with factors’ Vocalization (Human Vocalizations and Non-vocalizations) and Valence (Positive, Neutral, and Negative) revealed a significant main effect of Vocalization bilaterally in VA and L2, as well as in the left L1 and right L3 (Fig. 2a). A main effect of Valence was significant bilaterally in VA and L3, as well as in the left L2, right L4 and

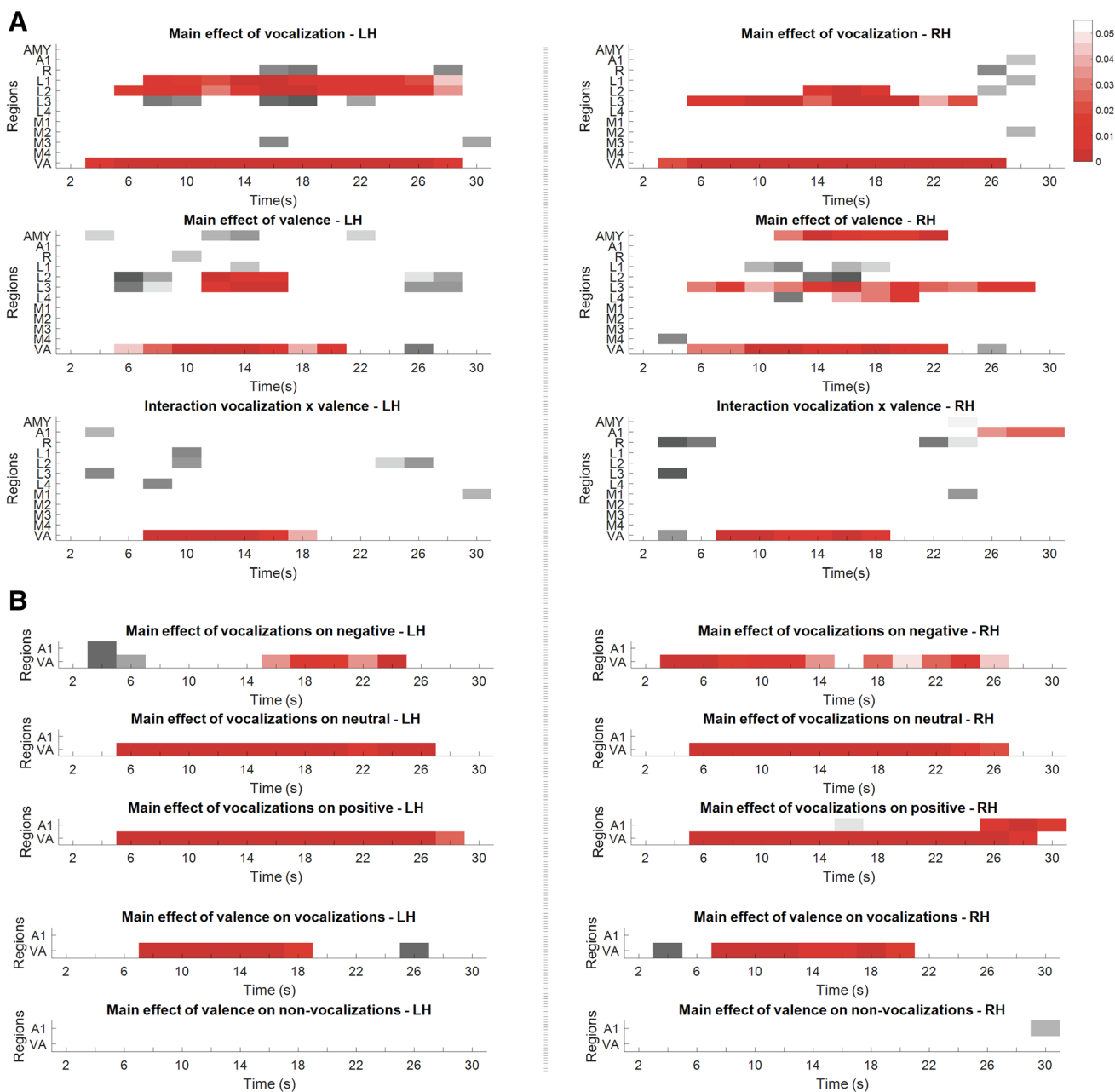


Fig. 2 Statistical results. **a** Two-way ANOVA 2 vocalization (human vocalizations and non-vocalizations)×3 valence (positive, neutral, and negative) on the BOLD signal. The results of the ANOVA vocalization×valence, with the ROIs represented on the y-axis, the time-points on the x-axis, and the color bar indicating different statistical thresholds. The red colors indicate a *p* value lower or equal to 0.05 for at least three consecutive timepoints. Upper panel: main effect of Vocalization. Middle panel: main effect of Valence. Lower panel:

interaction Vocalization×Valence. **b** Separate one-way ANOVA for Human Vocalizations and Non-vocalizations. Three upper panels: main effect of Valence on human vocalization categories (HVP, HV0, HVN). Two lower panels: main effect of Valence on non-vocalization categories (NVP, NV0, NVN). RH: right hemisphere; LH: left hemisphere. Refer to the manuscript for the ROIs’ definition and abbreviations

AMY. The interaction Vocalization × Valence was significant bilaterally in VA, as well as in right A1. To investigate this significant interaction, we performed post-hoc one-way ANOVAs for the factor Valence on the categories of Human Vocalizations and Non-vocalizations separately, as well as

post-hoc one-way ANOVAs for the factor Vocalization on the categories of Positive, Negative, and Neutral sounds separately (Fig. 2b). The analysis of the effect of emotional Valence on Human Vocalizations (HVP, HV0, HVN) using one-way ANOVA yielded a significant main effect

of Valence bilaterally in VA. The analysis of the effect of emotional Valence on Non-vocalizations (NVP, NV0, NVN) using one-way ANOVA did not yield any significant main effect. Post-hoc *t* tests reveal that the main effect of Valence on vocalizations in VA was driven by a significant difference between ‘HVP and HVN’ and ‘HV0 and ‘HVN’. The interaction effect found in right A1 was driven by a main effect of Vocalization only for the positive stimuli and not for the neutral nor negative stimuli. A post-hoc three-way repeated measure ANOVA with factors’ Vocalization (Human Vocalizations and Non-vocalizations), Valence (Positive, Neutral, and Negative), and Time (every 15 time point of the block) revealed a significant main effect for Vocalization in bilateral VA, in left L1 and L2, and right L3; a significant main effect for Valence in right L3 and bilateral VA; a significant main effect for Time in all bilateral ROIs; a significant interaction Vocalization \times Time bilaterally in R, L2, L3, VA, in left L1 and M3, and right M1 and M2; a significant interaction Valence \times Time bilaterally in L2, L3, and VA; and a significant interaction Vocalization \times Valence \times Time in bilateral L2, L3, VA, right AMY, and left A1 and R. There was no significant interaction Vocalization \times Valence when Time was considered as a factor (results not displayed). The main effect for Time in all bilateral ROIs supported our main hypothesis that the BOLD evolved differently for each stimuli along the blocks; therefore, our interpretations are focusing on the results from the two-way ANOVA.

The BOLD time courses for the ROIs with a significant effect in the ANOVA are presented in Fig. 3. VA responded preferentially to human vocalizations of neutral and positive valence, whereas the superior temporal gyrus responded preferentially to neutral sounds. The AMY shows a stronger response for the positive emotional sounds. We observed that the process of the various valences occurs at the beginning of the time course, regardless of the ROI. Moreover, we noticed a habituation effect for all sound categories in all ROIs, with the strongest one (i.e., longest plateau) observed in VA for the vocalizations.

Correlation of BOLD signal between regions of interest

To investigate the coupling between ROIs with significant effects of the two-way repeated measure ANOVA of the BOLD responses, we performed post-hoc correlations between the BOLD signals of the AMY, VA, L1, L2 and L3 for Human Vocalizations and for Non-vocalizations (Fig. 4a). Significant correlations ($p < 0.01$; with an adjusted $R^2 > 0.6$; see Tables 3 and 4) were found among the three belt areas (L1, L2, L3) and VA; their strength varied as a function of stimulus category and partially valence and between the hemispheres. Strikingly, the correlations between VA and the belt areas were stronger for vocalizations than for

non-vocalizations. The right AMY was correlated with the right VA and right L1, L2 and L3 during the presentation of HVP (but not during non-vocalizations nor during HVN or HV0); this effect was not found in the left hemisphere.

Discussion

Our results indicate that emotional valence modulates differentially neural activity that is elicited by human vocalizations vs. non-vocal environmental sounds within individual early stage auditory areas, VA and AMY. Using 7T fMRI and $1.5 \times 1.5 \times 1.5$ mm voxel size, without smoothing, rendered individual auditory areas accessible to investigation. Whereas emotional content modulates both vocalizations and other environmental sounds in early stage auditory areas, it singles out vocalizations in VA. Both types of sounds are also modulated in AMY. The specificity profiles within these regions and the correlations in their activity suggest that VA shares emotional information both with early stage auditory areas and with AMY.

Selectivity for emotional vocalizations

Whereas emotional valence modulates neural activity elicited by different sensory modalities in AMY or by different sound categories in lateral belt areas, its modulation appears to be limited to a single category, human vocalizations, in VA. This area was initially identified by its selectivity for human vocalizations, including speaker’s identity, over other environmental sounds or acoustically similar control stimuli such as scrambled voices or amplitude modulated noise (Belin et al. 2000, 2002; Warren et al. 2006; Latinus et al. 2013; Zäske et al. 2017). VA is located within the superior temporal sulcus and there is inter-individual variability as to its precise location (Pernet et al. 2015). VA neural activity elicited by human vocalizations was shown to be modulated by emotional intensity of happy or angry intonation (Ethofer et al. 2006b), of laughter (Lavan et al. 2017) and of positive or negative valence of non-verbal vocalizations (Bestelmeyer et al. 2017). Our results confirm the selectivity of VA for human vocalizations over other environmental sounds and show that emotional modulation impacts the encoding of vocalizations but not of other environmental sounds. To our knowledge, none of the previous studies investigated this issue specifically, and none reported emotional modulation of environmental sounds other than vocalizations within VA. Further studies need to establish whether the neural populations, which encode emotional vocalizations do so in a categorical way (positive vs. neutral vs. negative) or on a continuous scale.

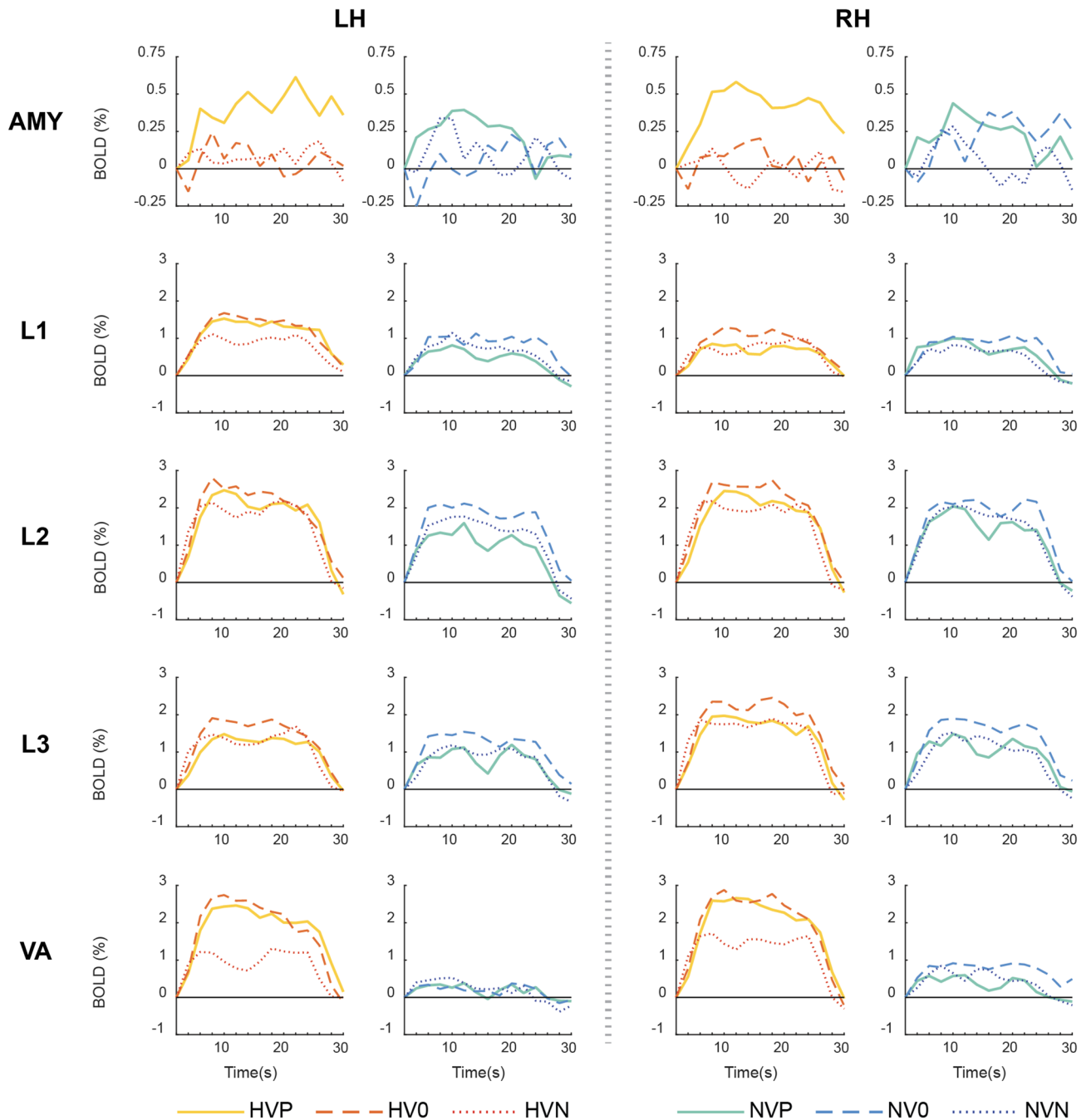


Fig. 3 BOLD time courses for significant ROIs: AMY, L1, L2, L3, and VA. BOLD time courses for the left and right hemispheres are in the left and right parts, respectively. For each hemisphere, the left panel depicts the human vocalization categories [HVP (solid line), HV0 (dashed line), and HVN (dotted line)] in warm colors and the

right panel the non-vocalization categories [NVP (solid line), NV0 (dashed line), NVN (dotted line)] in cold colors. *RH* right hemisphere; *LH* left hemisphere. Refer to the manuscript for the ROIs and sound categories abbreviations

Selectivity for emotions and for vocalizations within early stage auditory areas

Our results indicate that specific lateral belt areas are selective for vocalizations over other environmental sounds and/or modulated by emotional valence but that

emotional modulation is not limited to a specific stimulus category. Within the left hemisphere areas, L1 and L2, which are located postero-laterally on the planum temporale, are selective for vocalizations, whereas L2 and L3, which are located laterally on the planum temporale and Heschl's gyrus, are selective for emotional valence.

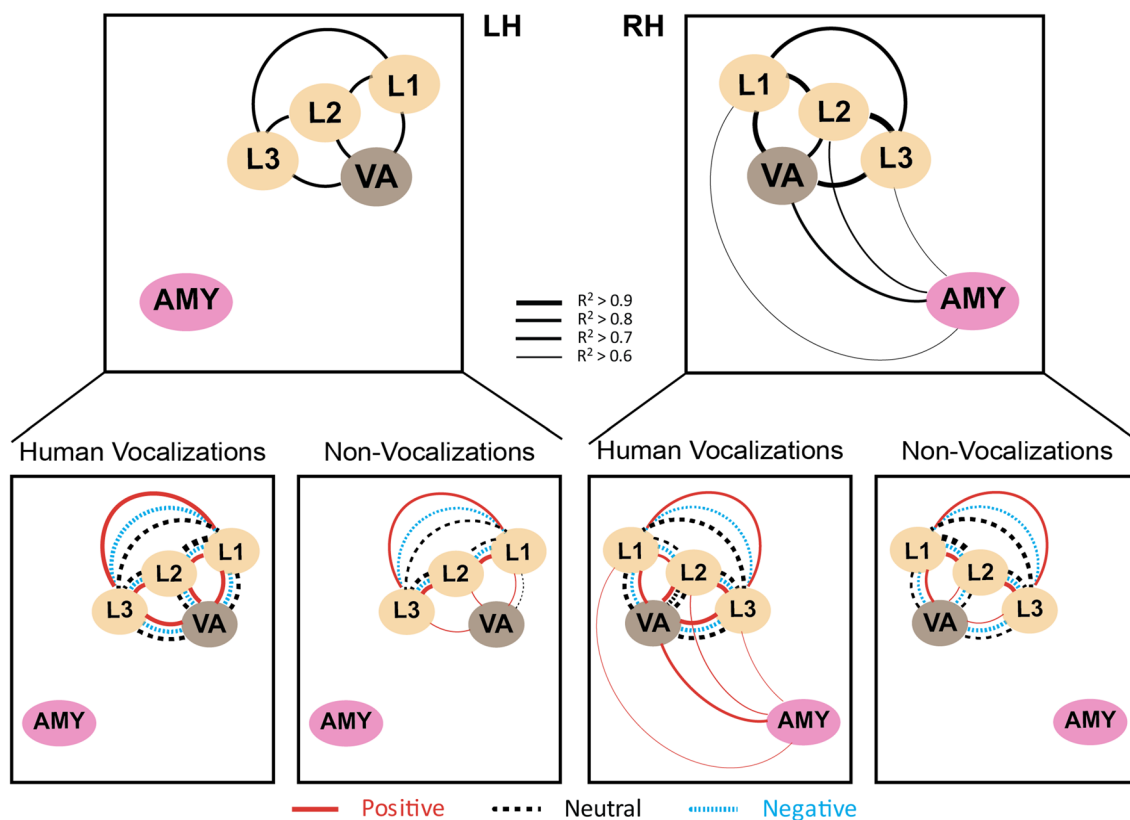


Fig. 4 BOLD correlations. Correlations between the BOLD time courses of the ROIs L1, L2, L3, VA, and AMY, in the left and right hemispheres separately. The thickness of the lines represents the strength of the correlation. Correlations with an adjusted- R^2 smaller than 0.6 are not represented. In the lower part of the figure, the correlations are separated for the human vocalizations categories [HVP

(solid red line), HV0 (dashed black line), HVN (dotted blue line)] and for the non-vocalizations categories [NVP (solid red line), NVO (dashed black line), NVN (dotted blue line)]. *RH* right hemisphere, *LH* left hemisphere. Refer to the manuscript for the ROIs definition and abbreviations

Within the right hemisphere L2 and L3 are selective for vocalizations and L3 and L4 for emotional valence. Voice selectivity within the planum temporale has been documented in previous studies, showing stronger activation to vocal than to non-vocal sounds (Belin et al. 2000) and participating in spectrotemporal analysis of vocalizations, a processing step which is believed to precede speaker identification in the superior temporal sulcus (Warren et al. 2006). Modulation by emotional valence was reported in a region located postero-laterally to the PAC, where emotional vocalizations yielded stronger activations than neutral voices (Wildgruber et al. 2004a; Ethofer et al. 2006b, 2012; Leitman et al. 2010; Bestelmeyer et al. 2017) or distinct spatial response patterns to different emotional categories (Ethofer et al. 2009b). This region on the postero-lateral supratemporal plane is very likely part of the belt or parabelt areas, as suggested by its connectivity pattern. Diffusion Spectrum MRI (DSI) and post-mortem tracing studies have shown that this part of the auditory cortex (1) is interconnected with the primary auditory cortex and with higher order areas on the superior temporal gyrus

(Cammoun et al. 2015); (2) receives monosynaptic callosal afferents from the fusiform gyrus (Di Virgilio and Clarke 1997); and (3) has intrinsic connections that tend to be longer than those within the primary auditory cortex, but shorter than those of Broca’s area (Tardif and Clarke 2001; Tardif et al. 2007). This supratemporal region was included, together with a large part of VA, in the so-called “emotional voice area” (Ethofer et al. 2012) because of its responsiveness to emotionally modulated vocalizations; the specificity of the emotional effect for vocalizations vs. other sound categories has, however, not been investigated prior to our study. The high spatial resolution of the present study allowed us to show that the “emotional voice area” consists of two functionally distinct regions, the early stage auditory areas, where the emotional content modulates neural activity elicited by vocalizations and by other environmental sounds, and VA, where it modulates responses to vocalizations only. The effect appears to be driven by a stronger response to neutral than positive or negative valence in early stage areas and by positive valence in VA.

Table 2 Talairach coordinates of the peaks of the activation clusters of ‘Sounds vs. Silence’

Area	X	Y	Z	<i>t</i> value	#voxels	BA
Left hemisphere						
Transverse gyrus (HG)	– 48	– 19	10	33.75	21,230	41
Triangular part IFG	– 45	16	16	10.05	1282	45
	– 42	19	10	9.33	771	
Dorsolateral PFC	– 47	42	7	9.73	392	46
	– 51	29	13	7.41	247	
Parahippocampal gyrus	– 18	– 11	– 13	9.23	386	54
Orbital part IFG	– 28	27	– 2	9.53	354	47
Insula	– 24	23	12	7.11	307	13
STG	– 52	– 51	12	7.74	164	22
Cerebellum	– 41	– 58	– 23	7.77	154	–
	– 23	– 29	– 22	8.72	145	–
ITG	– 54	– 57	– 15	7.63	150	20
MGB	– 16	– 26	– 5	8.46	140	–
Cuneus (V1)	– 11	– 90	– 8	7.11	137	17
Right hemisphere						
Transverse gyrus (HG)	54	– 16	5	31.22	24,633	41
Dorsolateral PFC	49	38	2	10.42	922	46
AMY	16	– 5	– 20	11.30	715	53
Opercular part IFG	45	16	21	11.87	615	44
	54	16	30	8.51	184	
	36	30	0	6.76	159	
	31	29	0	5.78	121	
Dorsolateral PFC	50	34	19	8.18	213	9
Inferior Colliculus	6	– 32	– 2	11.95	144	–
Secondary visual cortex	7	– 88	– 15	9.12	135	18
Thalamus	11	– 14	6	8.29	111	–

Talairach coordinates, *t* values, number of voxels and corresponding regions for the peaks of all the activation clusters resulting from the contrast sounds vs. silence in the RFX GLM analysis (FDR $q < 0.05$, cluster threshold of 100 voxels). Clusters were sorted according to their size and corresponding region

HG Heschl’s Gyrus, *IFG* inferior frontal gyrus, *PFC* prefrontal gyrus, *STG* superior temporal gyrus, *ITG* inferior temporal gyrus, *MGB* medial geniculate body of the thalamus, *AMY* amygdala

Modulation by emotional valence in the amygdala

Emotional valence in AMY is encoded independently of stimulus category. Although several studies have shown that AMY plays an important role in processing emotions in non-verbal vocalizations (Phillips et al. 1998; Morris et al. 1999; Sander et al. 2003, 2007; Fecteau et al. 2007; Frühholz et al. 2014), modulation by emotional valence concerns other auditory categories, demonstrated here and in a previous study (Frühholz et al. 2014) or other sensory modalities (Baas et al. 2004; Costafreda et al. 2008; Ball et al. 2009). Our finding that the emotional effect was driven by stronger responses to positive than neutral and negative stimuli is consistent with previous studies using human vocalizations (Fecteau et al. 2007; Wiethoff et al. 2009) or other stimuli (O’Doherty et al. 2001; Anderson et al. 2003; Winston et al. 2005; Ball et al. 2007; Hurlmann et al. 2008; Sergerie et al. 2008; Costa et al. 2010). However, our results contrast with

a recent report that screams and alarms, a priori negative stimuli, activate strongly AMY; the authors attribute this selectivity to the acoustic feature of roughness, which is shared by both types of stimuli (Arnal et al. 2015). Several other studies have highlighted the preference of AMY for negative emotional valence (Morris et al. 1998; Phillips et al. 1998, 2001; Wright et al. 2001). In our experimental paradigm modulation by emotional valence was stronger in AMY in the right hemisphere. Previous studies reported right (for laughing and crying sounds: (Sander et al. 2003) or left lateralization (for non-linguistic lateralizations: (Fecteau et al. 2007) or bilateral activation (Aubé et al. 2015). These divergent findings could be explained by the complex structure of AMY, which is composed of several nuclei (Amunts et al. 2005; Roy et al. 2009; Solano-Castiella et al. 2011). Imaging AMY remains difficult because of inhomogeneities in the local magnetic field (Labar et al. 2001), the proximity of large veins (Boubela et al. 2015) and the

Table 3 BOLD correlations for the left hemisphere

Left hemisphere	L1	L2	L3	VA	AMY
	<i>p</i> value (<i>R</i> ²)	<i>p</i> value (<i>R</i> ²)	<i>p</i> value (<i>R</i> ²)	<i>p</i> value (<i>R</i> ²)	<i>p</i> value (<i>R</i> ²)
HV0					
L1		1.56E-09 (0.94)	8.36E-11 (0.96)	2.08E-09 (0.94)	
L2			1.56E-12 (0.98)	5.72E-12 (0.98)	
L3				4.86E-11 (0.97)	
VA					
AMY					
HVN					
L1		2.38E-09 (0.94)	2.20E-09 (0.94)	3.73E-07 (0.87)	
L2			6.73E-12 (0.98)	1.85E-08 (0.92)	
L3				5.38E-09 (0.93)	
VA					
AMY					
HVP					
L1		5.65E-09 (0.93)	5.05E-11 (0.97)	5.42E-13 (0.98)	
L2			1.14E-12 (0.98)	3.85E-10 (0.96)	
L3				1.73E-11 (0.97)	
VA					
AMY					
NV0					
L1		6.99E-08 (0.89)	6.74E-08 (0.89)		
L2			3.07E-13 (0.99)		
L3					
VA					
AMY					
NVN					
L1		3.04E-09 (0.94)	4.24E-08 (0.91)	2.73E-05 (0.75)	
L2			3.20E-10 (0.96)		
L3					
VA					
AMY					
NVP					
L1		1.97E-10 (0.96)	9.74E-08 (0.89)	5.06E-05 (0.73)	
L2			5.24E-08 (0.90)	2.30E-05 (0.76)	
L3				9.46E-06 (0.79)	
VA					
AMY					

p values and *R*² for the BOLD correlations of the left ROIs (L1, L2, L3, VA, and AMY) for the different sound categories (HV0, HVP, HVN, NV0, NVP, and NVN). The results are only presented for correlations with *p* values lower than 0.01 and R-square greater than 0.6. The correlations that do not meet this criterion are hatched. No result in the shaded part of the table, as this is a symmetrical matrix. Refer to the manuscript for the ROIs' definition and abbreviations

lateralization of AMY activation due to the phase-encoding polarity (Mathiak et al. 2012). These limitations prevented us from parcellating AMY accurately and exploring emotional encoding in specific sub-nuclei.

Processing pathway for emotional vocalizations

Our results speak in favour of a module dedicated to the processing of the emotional value of human vocalizations but not of other environmental sounds, which is part of or co-extensive with VA (Fig. 4b). This observation highlights

three features of emotional processing. First, the neural mechanisms underlying this specificity involve most likely the combination of a category-specific input from the lateral belt areas and of emotion-specific input from AMY, as suggested by evidence from activation and connectivity patterns. Our results indicate that emotional information, which is encoded in VA, shares a preference for positive stimuli with AMY and for neutral stimuli with lateral belt areas. As reported in the previous studies, the lateral part of the planum temporale processes temporo-spatial information pertaining to vocalizations and relays this information to

Table 4 BOLD correlations for the right hemisphere

Right hemisphere	L1	L2	L3	VA	AMY
	<i>p</i> value (R^2)	<i>p</i> value (R^2)	<i>p</i> value (R^2)	<i>p</i> value (R^2)	<i>p</i> value (R^2)
HV0					
L1		1.60E-10 (0.96)	6.54E-10 (0.95)	4.82E-11 (0.97)	
L2			1.08E-12 (0.98)	4.31E-15 (0.99)	
L3				3.14E-13 (0.98)	
VA					
AMY					
HVN					
L1		1.43E-06 (0.84)	2.29E-06 (0.83)	1.34E-07 (0.89)	
L2			3.48E-13 (0.98)	1.71E-10 (0.96)	
L3				3.01E-09 (0.94)	
VA					
AMY					
HVP					
L1		2.08E-07 (0.88)	1.01E-07 (0.89)	3.45E-07 (0.87)	1.61E-04 (0.68)
L2			9.18E-12 (0.97)	4.48E-13 (0.98)	2.96E-05 (0.75)
L3				1.76E-10 (0.96)	1.48E-04 (0.68)
VA					3.69E-06 (0.79)
AMY					
NV0					
L1		5.88E-10 (0.95)	5.42E-10 (0.95)	4.65E-07 (0.87)	
L2			1.79E-12 (0.98)	1.99E-06 (0.83)	
L3				2.70E-07 (0.88)	
VA					
AMY					
NVN					
L1		1.12E-08 (0.92)	4.55E-07 (0.87)	2.43E-06 (0.83)	
L2			3.82E-11 (0.97)	4.33E-07 (0.87)	
L3				3.98E-08 (0.91)	
VA					
AMY					
NVP					
L1		2.26E-08 (0.92)	6.72E-08 (0.89)	2.87E-07 (0.88)	
L2			8.02E-10 (0.95)	3.26E-05 (0.75)	
L3				7.76E-06 (0.79)	
VA					
AMY					

Same conventions as in Table 3

VA, where higher order analysis, including voice identification, is conducted (Belin et al. 2000; Warren et al. 2006). The corresponding interconnection between lateral belt areas and the region of the superior temporal sulcus, where VA is located, was demonstrated using DSI tract tracing (Cammoun et al. 2015). Furthermore, functional connections were described between VA and the supratemporal region (Pernet et al. 2015). Functional connectivity between AMY and VA, reported in an early study (Roy et al. 2009), was not confirmed in a later study, which proposed that the AMY-VA link passes via the prefrontal cortex (Pernet et al. 2015). The model of dual input to VA, from the lateral belt areas and from AMY, is consistent with the multi-stage concept for the processing of vocalizations and valence (Schirmer and Gunter 2017); evidence from EEG studies suggests that

vocalizations and valence are processed first independently, before being integrated in higher order auditory or frontal regions. At the neuronal level, we can only hypothesize as to the mechanisms that underlie the emotion-vocalization selectivity by postulating the existence of populations of “human-vocalization neurons”, which are driven by inputs from lateral belt areas and AMY. Second, VA is not the only category-specific area that is selectively modulated by emotional valence. The fusiform face area shares the same feature. Emotional expressions were shown to modulate neural activity within the fusiform face area (intracranial recordings: Pourtois et al. 2009); fMRI: (Jehna et al. 2011; Harry et al. 2013), albeit less than within the face area in the superior temporal sulcus (Zhang et al. 2016). As revealed by a meta-analysis of fMRI studies, the fusiform face area

appears to process emotional content for faces but not for other categories (i.e., visual scenes, Sabatinelli et al. 2011). Thus, predominantly auditory and predominantly visual cortices each comprise an area, where category-specific processing—voices and faces, respectively—is modulated by emotional valence. In both cases, this emotion-linked encoding concerns stimuli of high social relevance. Third, the emotion–vocalization specific module within VA is very likely a stepping stone towards a more global, hetero-modal representation of emotionally relevant information about people (Watson et al. 2014). The combined encoding of voices and faces, including emotional aspects, was shown to involve a small part of the superior temporal sulcus at the intersection of VA and the more posterior lying face area (Kreifelts et al. 2009; Ethofer et al. 2013).

Several aspects of the processing of emotional vocalizations, which remain to be explored, could be addressed in future studies with an event-related paradigm at 7T. This would allow to correlate the perceived valence by a given subject with the activation within a ROI and compare thus more precisely emotional modulation of vocalizations and non-vocalizations. This same design would be particularly adapted to investigate neural coupling between ROIs.

Conclusions

Our results highlighted different stages in the processing of emotional vocalizations. Within the supratemporal plane, several lateral early stage auditory areas responded strongly to non-verbal vocalizations and/or were modulated by emotional valence. However, none of these areas appeared to be dedicated to emotional processing of vocalizations only. This role was assumed by VA, where emotional valence modulated selectively responses to human vocalizations but not to other environmental sounds. In contrast, emotional valence modulated neural responses to both types of stimuli in right AMY. Correlation analysis revealed coupling between VA and early stage auditory areas during the presentation of any vocalization, and between VA and right AMY during positive vocalizations. Thus, emotional vocalizations are processed in a dual pathway, whereby the emotion–vocalization module within VA integrates the input from the lateral early stage auditory areas and from AMY.

Funding This work was supported by the Swiss National Science Foundation Grant to S. Clarke (FNS 320030-159708) and by the Centre d’Imagerie BioMédicale (CIBM) of the UNIL, UNIGE, HUG, CHUV, EPFL and the Leenaards and Jeantet Foundations.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest.

Ethical approval All procedures performed were in accordance and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards and ethical approval was obtained from the Ethical Committee of the Canton de Vaud (reference number 282/08).

References

- Aeschlimann M, Knebel J-F, Murray MM, Clarke S (2008) Emotional pre-eminence of human vocalizations. *Brain Topogr* 20:239–248
- Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, Habel U, Schneider F, Zilles K (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol (Berl)* 210:343–352
- Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G, Gabrieli JDE, Sobel N (2003) Dissociated neural representations of intensity and valence in human olfaction. *Nat Neurosci* 6:196–202
- Arnal LH, Flinker A, Kleinschmidt A, Giraud A-L, Poeppel D (2015) Human screams occupy a privileged niche in the communication soundscape. *Curr Biol* 25:2051–2056
- Aubé W, Angulo-Perkins A, Peretz I, Concha L, Armony JL (2015) Fear across the senses: brain responses to music, vocalizations and facial expressions. *Soc Cogn Affect Neurosci* 10:399–407
- Baas D, Aleman A, Kahn RS (2004) Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. *Brain Res Rev* 45:96–103
- Ball T, Rahm B, Eickhoff SB, Schulze-Bonhage A, Speck O, Mutschler I (2007) Response properties of human amygdala subregions: evidence based on functional mri combined with probabilistic anatomical maps. *PLoS One* 2:e307
- Ball T, Derix J, Wentlandt J, Wieckhorst B, Speck O, Schulze-Bonhage A, Mutschler I (2009) Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. *J Neurosci Methods* 180:57–70
- Barbas H (2007) Flow of information for emotions through temporal and orbitofrontal pathways. *J Anat* 211:237–249
- Barbas H, Zikopoulos B, Timbie C (2011) Sensory pathways and emotional context for action in primate prefrontal cortex. *Biol Psychiatry* 69:1133–1139
- Beaucousin V, Lacheret A, Turbelin M-R, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007a) FMRI study of emotional speech comprehension. *Cereb Cortex* 17:339–352
- Beaucousin V, Lacheret A, Turbelin M-R, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007b) FMRI study of emotional speech comprehension. *Cereb Cortex* 17:339–352
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312
- Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. *Cogn Brain Res* 13:17–26
- Bestelmeyer PEG, Kotz SA, Belin P (2017) Effects of emotional valence and arousal on the voice perception network. *Soc Cogn Affect Neurosci*. <https://academic.oup.com/scan/article/doi/10.1093/scan/nsx059/3753554/Effects-of-emotional-valence-and-arousal-on-the-voice-perception-network>. Accessed 30 May 2017
- Boubela RN, Kalcher K, Huf W, Seidel E-M, Derntl B, Pezawas L, Našel C, Moser E (2015) fMRI measurements of amygdala activation are confounded by stimulus correlated signal fluctuation in nearby veins draining distant brain regions. *Sci Rep* 5:10499
- Cammoun L, Thiran JP, Griffa A, Meuli R, Hagmann P, Clarke S (2015) Intrahemispheric cortico-cortical connections of the human auditory cortex. *Brain Struct Funct* 220:3537–3553

- Chevillet M, Riesenhuber M, Rauschecker JP (2011) Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *J Neurosci* 31:9345–9352
- Chiry O, Tardif E, Magistretti PJ, Clarke S (2003) Patterns of calcium-binding proteins support parallel and hierarchical organization of human auditory areas. *Eur J Neurosci* 17:397–410
- Clarke S, Morosan P (2012) Architecture, connectivity and transmitter receptors of human auditory cortex. In: Poeppel D, Overath T, Popper AN, Fay RR (eds) *Human auditory cortex*. Springer, New York
- Clarke S, Rivier F (1998) Compartments within human primary auditory cortex: evidence from cytochrome oxidase and acetylcholinesterase staining. *Eur J Neurosci* 10:741–745
- Costa VD, Lang PJ, Sabatinelli D, Versace F, Bradley MM (2010) Emotional imagery: assessing pleasure and arousal in the brain's reward circuitry. *Hum Brain Mapp* 31:1446–1457
- Costafreda SG, Brammer MJ, David AS, Fu CHY (2008) Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Res Rev* 58:57–70
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31:14067–14075
- Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M (2013) Tuning into sound: frequency-selective attentional filter in human primary auditory cortex. *J Neurosci* 33:1858–1863
- Da Costa S, Bourquin NM-P, Knebel J-F, Saenz M, van der Zwaag W, Clarke S (2015) Representation of sound objects within early-stage auditory areas: a repetition effect study using 7T fMRI. *PLoS One* 10:e0124072
- Da Costa S, Clarke S, Crottaz-Herbette S (2018) Keeping track of sound objects in space: the contribution of early-stage auditory areas. *Hear Res*. <http://www.sciencedirect.com/science/article/pii/S0378595517306032>
- De Meo R, Bourquin NM-P, Knebel J-F, Murray MM, Clarke S (2015) From bird to sparrow: learning-induced modulations in fine-grained semantic discrimination. *NeuroImage* 118:163–173
- Di Virgilio G, Clarke S (1997) Direct interhemispheric visual input to human speech areas. *Hum Brain Mapp* 5:347–354
- Dietrich S, Hertrich I, Alter K, Ischebeck A, Ackermann H (2007) Semiotic aspects of human nonverbal vocalizations: a functional imaging study. *NeuroReport* 18:1891–1894
- Dietrich S, Hertrich I, Alter K, Ischebeck A, Ackermann H (2008) Understanding the emotional expression of verbal interjections: a functional MRI study. *NeuroReport* 19:1751–1755
- Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D (2006a) Impact of voice on emotional judgment of faces: an event-related fMRI study. *Hum Brain Mapp* 27:707–714
- Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D (2006b) Effects of prosodic emotional intensity on activation of associative auditory cortex. *NeuroReport* 17:249–253
- Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P, Wildgruber D (2008) Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J Cogn Neurosci* 21:1255–1268
- Ethofer T, Kreifelts B, Wiethoff S, Wolf J, Grodd W, Vuilleumier P, Wildgruber D (2009a) Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J Cogn Neurosci* 21:1255–1268
- Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009b) Decoding of emotional information in voice-sensitive cortices. *Curr Biol* 19:1028–1033
- Ethofer T, Brettecher J, Gschwind M, Kreifelts B, Wildgruber D, Vuilleumier P (2012) Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cereb Cortex* 22:191–200
- Ethofer T, Brettecher J, Wiethoff S, Bisch J, Schlipf S, Wildgruber D, Kreifelts B (2013) Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. *NeuroImage* 76:45–56
- Fecteau S, Belin P, Joanette Y, Armony JL (2007) Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage* 36:480–487
- Formisano E, Kim D-S, Salle FD, van de Moortele P-F, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859–869
- Frühholz S, Grandjean D (2013) Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci Biobehav Rev* 37:2847–2855
- Frühholz S, Trost W, Grandjean D (2014) The role of the medial temporal limbic system in processing emotions in voice and music. *Prog Neurobiol* 123:1–17
- Frühholz S, Trost W, Kotz SA (2016) The sound of emotions—towards a unifying neural network perspective of affective sound processing. *Neurosci Biobehav Rev* 68:96–110
- Ghashghaei HT, Barbas H (2002) Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* 115:1261–1279
- Goedhart AD, Van Der Sluis S, Houtveen JH, Willemsen G, De Geus EJC (2007) Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology* 44:203–215
- Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci* 8:145–146
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197–222
- Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140–149
- Harry BB, Williams M, Davis C, Kim J (2013) Emotional expressions evoke a differential response in the fusiform face area. *Front Hum Neurosci*. <https://www.frontiersin.org/articles/10.3389/fnhum.2013.00692/full>. Accessed 17 May 2018
- Hurlmann R, Rehme AK, Diessel M, Kukulja J, Maier W, Walter H, Cohen MX (2008) Segregating intra-amygdalar responses to dynamic facial emotion with cytoarchitectonic maximum probability maps. *J Neurosci Methods* 172:13–20
- Janak PH, Tye KM (2015) From circuits to behaviour in the amygdala. *Nature* 517:284–292
- Jehna M, Langkammer C, Wallner-Blazek M, Neuper C, Loitfelder M, Ropele S, Fuchs S, Khalil M, Pluta-Fuerst A, Fazekas F, Enzinger C (2011) Cognitively preserved MS patients demonstrate functional differences in processing neutral and emotional faces. *Brain Imaging Behav* 5:241–251
- Kamali A, Sair HI, Blitz AM, Riascos RF, Mirbagheri S, Keser Z, Hasan KM (2016) Revealing the ventral amygdalofugal pathway of the human limbic system using high spatial resolution diffusion tensor tractography. *Brain Struct Funct* 221:3561–3569
- Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly Z-M, Pruessmann KP, Stephan KE (2017) The PhysIO toolbox for modeling physiological noise in fMRI data. *J Neurosci Methods* 276:56–72
- Keifer OP, Gutman DA, Hecht EE, Keilholz SD, Ressler KJ (2015) A comparative analysis of mouse and human medial geniculate nucleus connectivity: a DTI and anterograde tracing study. *NeuroImage* 105:53–66
- Kim S-G, Knösche TR (2016) Intracortical myelination in musicians with absolute pitch: quantitative morphometry using 7-T MRI. *Hum Brain Mapp* 37:3486–3501

- Koelsch S (2010) Towards a neural basis of music-evoked emotions. *Trends Cogn Sci* 14:131–137
- Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D (2009) Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia* 47:3059–3066
- Labar KS, Gitelman DR, Mesulam MM, Parrish TB (2001) Impact of signal-to-noise on functional MRI of the human amygdala. *NeuroReport* 12:3461–3464
- Latinus M, McAleer P, Bestelmeyer PEG, Belin P (2013) Norm-based coding of voice identity in human auditory cortex. *Curr Biol* 23:1075–1080
- Lavan N, Rankin G, Lorking N, Scott S, McGettigan C (2017) Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia* 95:30–39
- LeDoux JE, Ruggiero DA, Reis DJ (1985) Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat. *J Comp Neurol* 242:182–213
- Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC (2010) “It’s Not What You Say, But How You Say it”: a reciprocal temporo-frontal network for affective prosody. *Front Hum Neurosci*. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831710/>. Accessed 21 Mar 2016
- Liebenthal E, Silbersweig DA, Stern E (2016) The language, tone and prosody of emotions: neural substrates and dynamics of spoken-word emotion perception. *Front Neurosci*. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5099784/>. Accessed 8 Mar 2017
- Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R (2010) MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage* 49:1271–1281
- Mathiak KA, Zvyagintsev M, Ackermann H, Mathiak K (2012) Lateralization of amygdala activation in fMRI may depend on phase-encoding polarity. *Magn Reson Mater Phys Biol Med* 25:177–182
- Meyer M, Zysset S, von Cramon DY, Alter K (2005) Distinct fMRI responses to laughter, speech, and sounds along the human perisylvian cortex. *Brain Res Cogn Brain Res* 24:291–306
- Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. *Front Neurosci*. <http://journal.frontiersin.org/article/10.3389/fnins.2014.00225/abstract>. Accessed 20 Feb 2017
- Morris JS, Friston KJ, Büchel C, Frith CD, Young AW, Calder AJ, Dolan RJ (1998) A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121:47–57
- Morris JS, Scott SK, Dolan RJ (1999) Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia* 37:1155–1163
- Murray RJ, Brosch T, Sander D (2014) The functional profile of the human amygdala in affective processing: insights from intracranial recordings. *Cortex* 60:10–33
- O’Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F (2001) Representation of pleasant and aversive taste in the human brain. *J Neurophysiol* 85:1315–1321
- Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex* 17:2251–2257
- Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* 28:8116–8123
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113
- Ottersen OP, Ben-Ari Y (1979) Afferent connections to the amygdaloid complex of the rat and cat. I. Projections from the thalamus. *J Comp Neurol* 187:401–424
- Pannese A, Grandjean D, Frühholz S (2016) Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. *Cortex* 85:116–125
- Pernet CR, McAleer P, Latinus M, Gorgolewski KJ, Charest I, Bestelmeyer PEG, Watson RH, Fleming D, Crabbe F, Valdes-Sosa M, Belin P (2015) The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage* 119:164–174
- Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams SCR, Bullmore ET, Brammer M, Gray JA (1998) Neural responses to facial and vocal expressions of fear and disgust. *Proc R Soc Lond B Biol Sci* 265:1809–1817
- Phillips ML, Medford N, Young AW, Williams L, Williams SC, Bullmore ET, Gray JA, Brammer MJ (2001) Time courses of left and right amygdalar responses to fearful facial expressions. *Hum Brain Mapp* 12:193–202
- Plaisant O, Courtois R, Réveillère C, Mendelsohn GA, John OP (2010) Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Ann Méd-Psychol Rev Psychiatr* 168:97–106
- Pourtois G, Spinelli L, Seeck M, Vuilleumier P (2009) Modulation of face processing by emotional expression and gaze direction during intracranial recordings in right fusiform cortex. *J Cogn Neurosci* 22:2086–2107
- Price JL, Amaral DG (1981) An autoradiographic study of the projections of the central nucleus of the monkey amygdala. *J Neurosci Off J Soc Neurosci* 1:1242–1259
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund H-J, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13:669–683
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724
- Reddy RK, Ramachandra V, Kumar N, Singh NC (2009) Categorization of environmental sounds. *Biol Cybern* 100:299
- Reser DH, Burman KJ, Richardson KE, Spitzer MW, Rosa MGP (2009) Connections of the marmoset rostrotemporal auditory area: express pathways for analysis of affective content in hearing. *Eur J Neurosci* 30:578–592
- Rivier F, Clarke S (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *NeuroImage* 6:288–304
- Roy AK, Shehzad Z, Margulies DS, Kelly AMC, Uddin LQ, Gotimer K, Biswal BB, Castellanos FX, Milham MP (2009) Functional connectivity of the human amygdala using resting state fMRI. *NeuroImage* 45:614–626
- Russchen FT (1982) Amygdalopetal projections in the cat. II. Subcortical afferent connections. A study with retrograde tracing techniques. *J Comp Neurol* 207:157–176
- Russchen FT, Amaral DG, Price JL (1987) The afferent input to the magnocellular division of the mediodorsal thalamic nucleus in the monkey, *Macaca fascicularis*. *J Comp Neurol* 256:175–210
- Sabatinelli D, Fortune EE, Li Q, Siddiqui A, Krafft C, Oliver WT, Beck S, Jeffries J (2011) Emotional perception: meta-analyses of face and natural scene processing. *NeuroImage* 54:2524–2533
- Sander K, Scheich H (2005) Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *J Cogn Neurosci* 17:1519–1531
- Sander K, Brechmann A, Scheich H (2003) Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Res Protoc* 11:81–91
- Sander K, Frome Y, Scheich H (2007) FMRI activations of amygdala, cingulate cortex, and auditory cortex by infant laughing and crying. *Hum Brain Mapp* 28:1007–1022

- Sauter DA, Eisner F, Ekman P, Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc Natl Acad Sci* 107:2408–2412
- Schirmer A, Gunter TC (2017) Temporal signatures of processing voiceness and emotion in sound. *Soc Cogn Affect Neurosci* 12:902–909
- Schirmer A, Kotz SA (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci* 10:24–30
- Scott BH, Leccese PA, Saleem KS, Kikuchi Y, Mullarkey MP, Fukushima M, Mishkin M, Saunders RC (2015) Intrinsic connections of the core auditory cortical regions and rostral supratemporal plane in the Macaque Monkey. *Cereb Cortex*:bhv277
- Sergerie K, Chochol C, Armony JL (2008) The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev* 32:811–830
- Shinonaga Y, Takada M, Mizuno N (1994) Direct projections from the non-laminated divisions of the medial geniculate nucleus to the temporal polar cortex and amygdala in the cat. *J Comp Neurol* 340:405–426
- Solano-Castiella E, Schäfer A, Reimer E, Türke E, Pröger T, Lohmann G, Trampel R, Turner R (2011) Parcellation of human amygdala in vivo using ultra high field structural MRI. *NeuroImage* 58:741–748
- Szameitat DP, Kreifelts B, Alter K, Szameitat AJ, Sterr A, Grodd W, Wildgruber D (2010) It is not always tickling: distinct cerebral responses during perception of different laughter types. *NeuroImage* 53:1264–1271
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91:1282–1296
- Tardif E, Clarke S (2001) Intrinsic connectivity of human auditory areas: a tracing study with DiI. *Eur J Neurosci* 13:1045–1050
- Tardif E, Probst A, Clarke S (2007) Laminar specificity of intrinsic connections in Broca's area. *Cereb Cortex* 17:2949–2960
- van der Zwaag W, Gentile G, Gruetter R, Spierer L, Clarke S (2011) Where sound position influences sound object representations: a 7-T fMRI study. *NeuroImage* 54:1803–1811
- Viceic D, Fornari E, Thiran J-P, Maeder PP, Meuli R, Adriani M, Clarke S (2006) Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. *NeuroReport* 17:1659–1662
- Viceic D, Campos R, Fornari E, Spierer L, Meuli R, Clarke S, Thiran J-P (2009) Local landmark-based registration for fMRI group studies of nonprimary auditory cortex. *NeuroImage* 44:145–153
- Viinikainen M, Kätsyri J, Sams M (2012) Representation of perceived sound valence in the human brain. *Hum Brain Mapp* 33:2295–2305
- Wallace MN, Johnston PW, Palmer AR (2002) Histochemical identification of cortical areas in the auditory region of the human brain. *Exp Brain Res* 143:499–508
- Warren JD, Scott SK, Price CJ, Griffiths TD (2006) Human brain mechanisms for the early analysis of voices. *NeuroImage* 31:1389–1397
- Watson R, Latinus M, Charest I, Crabbe F, Belin P (2014) People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex* 50:125–136
- Wiethoff SAB, Wildgruber DA, Grodd WB, Ethofer TAB (2009) Response and habituation of the amygdala during processing of emotional prosody. [Miscellaneous Article]. *Neuroreport* 20:1356–1360
- Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H (2004a) Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb Cortex* 14:1384–1389
- Wildgruber D, Hertrich I, Riecker A, Erb M, Anders S, Grodd W, Ackermann H (2004b) Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb Cortex* 14:1384–1389
- Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. In: Anders S, Ende G, Junghofer M, Kissler J, Wildgruber D (eds) *Progress in brain research. Understanding emotions*. Elsevier, Amsterdam, pp 249–268. <http://www.sciencedirect.com/science/article/pii/S0079612306560133>. Accessed 20 Mar 2019
- Wildgruber D, Ethofer T, Grandjean D, Kreifelts B (2009) A cerebral network model of speech prosody comprehension: international Journal of Speech-Language Pathology: vol 11, no 4. *Int J Speech Lang Pathol* 11:277–281
- Winston JS, Gottfried JA, Kilner JM, Dolan RJ (2005) Integrated neural representations of odor intensity and affective valence in human amygdala. *J Neurosci* 25:8903–8907
- Wright CI, Fischer H, Whalen PJ, McInerney SC, Shin LM, Rauch SL (2001) *Wright_2001.pdf*. *NeuroReport* 12:379–383
- Yukie M (2002) Connections between the amygdala and auditory cortical areas in the macaque monkey. *Neurosci Res* 42:219–229
- Zäske R, Awwad Shiekh Hasan B, Belin P (2017) It doesn't matter what you say: fMRI correlates of voice learning and recognition independent of speech content. *Cortex* 94:100–112
- Zhang H, Japee S, Nolan R, Chu C, Liu N, Ungerleider LG (2016) Face-selective regions differ in their ability to classify facial expressions. *NeuroImage* 130:77–90
- Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. *Acta Psychiatr Scand* 67:361–370
- Zilles K, Armstrong E, Schleicher A, Kretschmann H-J (1988) The human pattern of gyrification in the cerebral cortex. *Anat Embryol (Berl)* 179:173–179

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.