CORTEX 58 (2014) 170–185

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Automatic domain-general processing of sound source identity in the left posterior middle frontal gyrus

Bruno L. Giordano [a,*], Cyril Pernet [b], Ian Charest [c], Guylaine Belizaire [d,e], Robert J. Zatorre [f,d] and Pascal Belin [a,g,d]

[a] Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, UK
[b] Brain Research Imaging Center, Neuroimaging Sciences, University of Edinburgh, Western General Hospital, Edinburgh, Scotland, UK
[c] Medical Research Council – Cognition and Brain Sciences Unit, Cambridge, UK
[d] International Laboratory for Brain, Music and Sound (BRAMS), Université de Montréal, Montréal, QC, Canada
[e] Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Université de Montréal, Montréal, Québec, Canada
[f] Montréal Neurological Institute, McGill University, Montreal, QC, Canada
[g] Institut des Neurosciences de la Timone, UMR7289, CNRS-Université Aix Marseille, Marseille, France

ARTICLE INFO

Article history:
Received 29 August 2013
Reviewed 28 October 2013
Revised 24 March 2014
Accepted 9 June 2014
Action editor Norihiro Sadato
Published online 18 June 2014

Keywords:
Auditory cortex
Prefrontal cortex
Auditory object
fMRI
Adaptation

ABSTRACT

Identifying sound sources is fundamental to developing a stable representation of the environment in the face of variable auditory information. The cortical processes underlying this ability have received little attention. In two fMRI experiments, we investigated passive adaptation to (Exp. 1) and explicit discrimination of (Exp. 2) source identities for different categories of auditory objects (voices, musical instruments, environmental sounds). All cortical effects of source identity were independent of high-level category information, and were accounted for by sound-to-sound differences in low-level structure (e.g., loudness). A conjunction analysis revealed that the left posterior middle frontal gyrus (pMFG) adapted to identity repetitions during both passive listening and active discrimination tasks. These results indicate that the comparison of sound source identities in a stream of auditory stimulation recruits the pMFG in a domain-general way, i.e., independent of the sound category, based on information contained in the low-level acoustical structure. pMFG recruitment during both passive listening and explicit identity comparison tasks also suggests its automatic engagement in sound source identity processing.

* *Corresponding author*. Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK.
E-mail addresses: brunog@psy.gla.ac.uk, brungio@gmail.com (B.L. Giordano), cyril.pernet@ed.ac.uk (C. Pernet), ian.charest@mrc-cbu.cam.ac.uk (I. Charest), guylaine.belizaire@umontreal.ca (G. Belizaire), robert.zatorre@mcgill.ca (R.J. Zatorre), pascal.belin@glasgow.ac.uk (P. Belin).
http://dx.doi.org/10.1016/j.cortex.2014.06.005

1. Introduction

Natural objects in the environment produce variable sounds: a speaker utters different phonemes; a guitar plays different tones; a drinking glass produces different impact sounds depending on how it is struck. Recognizing the identity of a sound source in the face of this acoustical variability is thus fundamental to developing a stable and meaningful representation of the auditory environment. The cortical architecture underlying this ability still represent a largely unexplored frontier of auditory neuroscience (King & Nelken, 2009), and only a handful of studies addressed the cortical processing of sound source identity (Andics et al., 2010; Belin & Zatorre, 2003; Formisano, De Martino, Bonte, & Goebel, 2008; Imaizumi et al., 1997; Latinus, Crabbe, & Belin, 2011; Von Kriegstein & Giraud, 2006; Zatorre, Bouffard, & Belin, 2004). Consistently with the dual-stream model of auditory processing (Rauschecker & Scott, 2009; Romanski et al., 1999), these studies reveal that the cortical processing of source identity relies on regions part of the ventral "what" stream such as the anterior temporal sulcus (aSTS, Andics et al., 2010; Belin & Zatorre, 2003; Zatorre et al., 2004), and the posterior aspects of the inferior prefrontal cortex/premotor cortex (Latinus et al., 2011; Von Kriegstein & Giraud, 2006; Zatorre et al., 2004). Because most brain imaging studies on the cortical processing of source identities used speech stimuli, it is unclear whether the cortical analysis of source identity relies directly on regions that respond selectively to their category membership, such as voice-selective regions of middle and anterior STG/STS (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), or instead relies on cortical mechanisms that are independent of category information. Indeed, the close proximity of the above-mentioned posterior inferior prefrontal regions to regions that implement content-independent control and working-memory processes in the premotor/middle prefrontal cortex (Power & Petersen, 2013; Rottschy et al., 2012) may suggest an involvement in domain-general identity processing, i.e., an involvement in identity processes not modulated by the particular category of the sound stimulus.

Within the dual-stream model of auditory processing, cortical representations of sound identity are assumed to progressively disengage from low-level representations of the sound signal along an anterior gradient originating from the primary auditory cortex (A1; Leaver & Rauschecker, 2010; Rauschecker & Scott, 2009) or, alternatively, with the distance from A1 in both the anterior and posterior directions (Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2013; Peelle, Johnsrude, & Davis, 2010). This account suggests that frontal regions to which cortical regions part of the ventral stream are projecting, such as inferior prefrontal cortex, do not encode low-level representations of the sound signal, but rather high-level attributes resulting from complex transformations of the low-level information. The support for this hypothesis is mixed. For example, Cohen, Theunissen, Russ, and Gill (2007) showed that spectrotemporal receptive field (STRF) models that account for A1 and subcortical processing (e.g., Miller, Escabí, Read, & Schreiner, 2002) do not explain activity in the ventrolateral prefrontal cortex (vlPFC) of Rhesus macaques. In contrast, Romanski, Averbeck, and Diltz (2005) hypothesizes low-level encoding in the vlPFC because neural responses for the same animal model do not appear to process effectively high-level attributes such as call function or meaning. More importantly, both Andics et al. (2010) and Latinus et al. (2011) report evidence for acoustical sensitivity in inferior prefrontal cortex, an area also sensitive to speaker identity (Latinus et al., 2011; Von Kriegstein & Giraud, 2006). Although at odds with the dual-stream model, the presence of a cortical process for the identification of sound sources that relies on low-level structure is instead consistent with psychophysics evidence showing that low-level structure accounts well for the perception of the properties of a sound source (Giordano & McAdams, 2006; Giordano, Rocchesso, & McAdams, 2010; McAdams, Roussarie, Chaigne, & Giordano, 2010).

Different studies of the cortical encoding of source identity were carried out using different tasks. For example, whereas Belin and Zatorre (2003) investigated adaptation to speaker identity in passive-listening conditions, participants in Von Kriegstein and Giraud (2006) explicitly identified speakers. The effect of task on the cortical processing of source identity is unclear, and might indeed explain part of the variation in the functional-imaging literature. More importantly, no previous study assessed whether the processing of source identity in passive listening and explicit judgment conditions relies on overlapping cortical networks. As such, no evidence is available concerning the robustness of the cortical processing of source identities to variation in task demands and whether explicit source identification is necessary to recruit prefrontal cortex regions.

We investigated these issues in two fMRI experiments on the passive-listening adaptation and explicit discrimination of source identities. We sought to ascertain: (1) which cortical regions are involved in the processing of source identity and whether common cortical regions process identity in passive and explicit-task conditions; (2) whether separate cortical modules process the identity of sources belonging to different categories [Vocalizations (speech and non-speech); Music; Environmental sounds]; (3) the extent to which low-level structure accounts for cortical identity processing. Overall, results suggest that (1) a region in the left posterior middle frontal gyrus (pMFG) is involved in the processing of the identity of sound sources in both active and passive-listening conditions, (2) identity-related activity in this region is not modulated by sound category and (3) it is accounted for by low-level acoustical structure.

2. Materials and methods

The two fMRI experiments were carried out during two subsequent scanning sessions on the same participants. During Experiment 1, they listened passively to sequences of sounds that differed in the number of identity repetitions. During Experiment 2, they were presented with pairs of sounds and were asked to evaluate whether the two sounds had been produced with the same sound-generating object or not.

2.1. Stimuli

On each trial of Experiment 1, participants heard one of 48 different sequences of six natural sounds [median/interquartile range (IQR) of sequence duration = 7250/10.5 msec; median/IQR duration of sound within sequence = 1000/4 msec; median/IQR within-sequence interstimulus interval (ISI) = 250/5 msec]. Each of the sequences comprised stimuli belonging to one of three sound categories: environmental sounds (16 sequences); music-instrument tones (16 sequences); human vocalizations (8 speech sequences; 8 non-speech vocalization sequences). The sound-category factor was crossed factorially with a four-level sequence-type factor measuring the number of same-stimulus repetitions within the sequence (N repetitions from now on): [1 (all stimuli different); 2 (3 different stimuli, each repeated twice); 3 (2 different stimuli each repeated three times); 6 (all stimuli equal); each stimulus was presented in only one of the sequences]. In all sequences, non-repeated sounds were generated with different sound sources. Each sound category thus comprised 4 sequences for each level of the sequence type/N-repetitions factor (2 sequences for each N-repetitions level for speech and non-speech vocalization sequences). The order of stimuli within each sequence was the same for all participants. For the 2-repetition sequences, each of the three different stimuli was presented in random order once during the first half of the sequence, and again during the second half. Neither the 2- nor the 3-repetition sequences included subsequent presentations of the same stimulus (e.g., [sounds S1-S2-S1-S2-S1-S2] and [S1-S2-S3-S2-S3-S1] and for a 2- and 3-repetition sequence, respectively).

On each trial of the Experiment 2, participants were presented with one of 108 pairs of sounds [median/IQR sound-pair duration = 2701/3334 msec; median/IQR duration of within-pair sound = 746/183 msec; within-pair ISI = 1200 msec]. Each of the pairs comprised sounds belonging to one of three different categories (36 pairs for each category): environmental sounds; three-tone synthetic-instrument melodies; human vocalizations (phonemes). The sound-category factor was crossed factorially with a three-level pair-type factor (12 pairs for each pair-type level within each sound category). In different sound/different source pairs (DD), the paired sounds were different, and were generated with two different sound sources although highly similar sound sources (e.g., two vacuum cleaners or two helicopters; two same-gender adults – 6 pairs – or two same-gender children – 4 pairs – or two different-gender children – 2 pairs; two different models of the same musical instrument such as two trumpets or two organs). For the musical-instrument category, the two instruments played the same melody. For the environmental-sound category the paired sources produced sounds based on the same interaction type (e.g., breaking for the two glass-breaking sounds; multiple impacts for two different sets of coins; two squeaking bottle corks). For the vocal category, half of the pairs included the same phoneme (both/a/or both/i/) and half of the pairs included different phonemes (/a/and/i/). In different sound/same source pairs (DS), the paired sounds were different, but were generated with the same sound source (e.g., two different card-shuffles with the same deck; two different rings with the same bicycle ring; two different melodies with the same musical instrument; two different phonemes uttered by the same speaker). In same sound/same source pairs (SS), the same sound was repeated twice. The same sound was used as initial sound in each of three pairs, one for each of the pair types, for a total of 36 unique initial sounds across all of the set of pairs (e.g., the same phonemic sound was paired with itself [SS], or with a different phoneme uttered by the same speaker [DS], or with a different phoneme uttered by a different speaker [DD]). The second sound was different across all of the sound pairs, with the exception of two vocal sounds which were used to create both one DS and one DD pair. The design for the stimuli in both experiments is summarized in Fig. 1.

2.1.1. Low-level dissimilarity

We modeled the dissimilarity of sounds within the presented sequences and pairs relative to a set of 12 low-level features, and analyzed the effect of the experimental factors on the low-level dissimilarity. The results of this analysis indicate that low-level within-sequence/pair dissimilarity decreased with an increase in the number of same-sound repetitions for Experiment 1, and from SS to DS to DD pairs in Experiment 2. They are summarized in Fig. 2.

Low-level features were extracted based on the approach described in Giordano et al. (2013; see for a more extensive description of the modeling approach). For each of the sound stimuli within each of the sequences, we initially quantified the time-varying profile of four different low-level features (temporal resolution = 1 msec): (1) loudness in sones, defined for each frame of analysis as the sum of the specific loudness for the different cochlear filters; (2) spectral centroid in ERB-rate units (ERB = Equivalent Rectangular Bandwidth, Moore & Glasberg, 1983), defined as the specific-loudness weighted average of the spectral frequency; (3) harmonic-to-noise ratio (HNR) – periodicity in short – a measure of the ratio of periodic-to-non-periodic energy in the sound signal in dB; (4) pitch in ERB-rate units. Time-varying loudness and spectral centroid were derived from the time-varying specific loudness of the sound signals, as computed according to the model of Glasberg and Moore (2002). Time-varying periodicity and pitch were computed using the Praat software (Boersma & Weenink, 2009). For each of the sequences, we computed 12 different measures of low-level dissimilarity based on the application of each of 3 different operators on each of the 4 time-varying features. For each of the low-level time-varying features, median dissimilarities were computed as follows: (i) extract the median of the time-varying feature for each of the sounds in the sequence or pair; (ii) compute a matrix of within-sequence pairwise dissimilarities defined as the absolute difference of the median feature between each pair of sounds within the sequence; (iii) average the pairwise dissimilarities within the sequence to hold an overall score of within-sequence dissimilarity of medians. Interquartile-range dissimilarities were computed by adopting the same pairwise absolute-difference approach as for the median dissimilarities. Finally, the cross-correlation dissimilarity between sounds belonging to the same sequence measured the lag- and scale-independent dissimilarity of the temporal patterns of variation of a target low-level feature. The cross-correlation

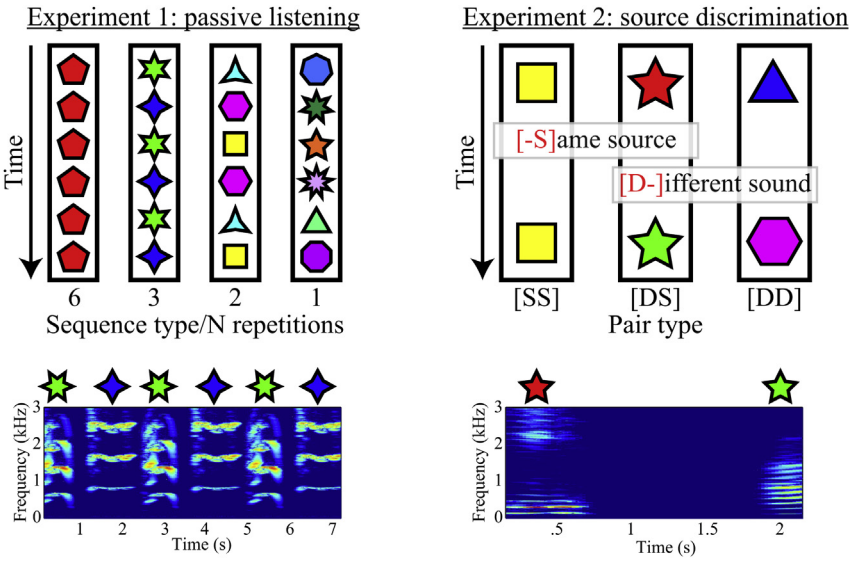

Fig. 1 – Experimental design. Schematics of the stimuli (sound sequences or pairs) presented on each trial of the two experiments. Different symbols and colors denote different sound sources and different sounds, respectively (e.g., the two paired sounds presented during the DS condition of Experiment 2 were different but generated with the same sound source). Sound category (environmental sound, music or human vocalizations) was constant within each of the presented sound sequences and pairs. Regions involved in the processing of the identity of sound sources are expected to show a decrease in fMRI activation for an increase in the number of same-stimulus repetitions in Experiment 1, and lower fMRI activations for SS than DS than DD pairs in Experiment 2. The bottom part of the figure shows a detail of the spectrogram for a vocal-sound sequence (Experiment 1; 3-repetitions condition) and for a vocal sound pair (Experiment 2; DS condition; level increases from blue to red).

dissimilarity between a time-varying feature for two sounds was defined as 1 minus the maximum cross-correlation between their time-varying features. The cross-correlation was normalized so as to yield a value of −1 for the cross-correlation between one signal and its negative at lag 0, and a value of 1 for the cross-correlation between one signal and its replica (i.e., autocorrelation) at lag 0. In order to yield a scale-independent measure of the temporal-pattern dissimilarity, time-varying features were range-normalized between 0 and 1 before being analyzed with the cross-correlation algorithm. The matrix of cross-correlation dissimilarities computed between each pair of sounds within the sequence was averaged in order to yield a sequence-specific measure of cross-correlation dissimilarity. No averaging was necessary for the pairs presented in Experiment 2 because they comprised only two sounds.

We then assessed the effects of the experimental factors on the within-sequence/pair low-level dissimilarity in each of the Experiments. We did not adopt a multivariate approach (e.g., MANOVA) because of the low number of data points, i.e., sound sequences, within each cell of the experimental designs (Tabachnick & Fidell, 2007). We instead carried out

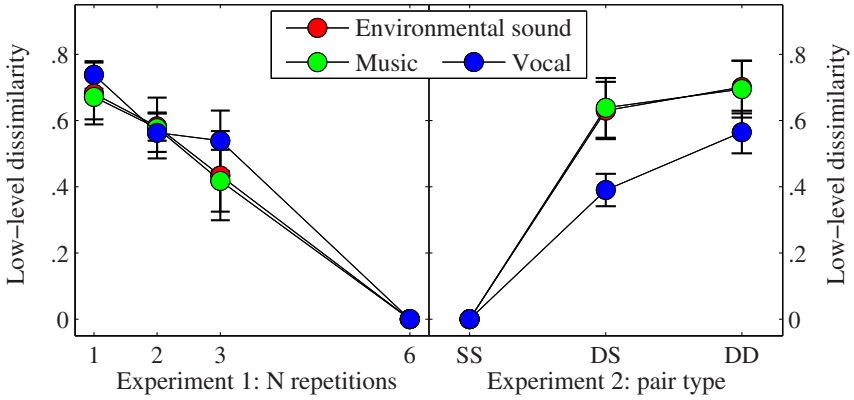

Fig. 2 – Low-level within-sequence and within-pair stimulus. Low-level scores of the dissimilarity between the sounds presented within the same sequence (Experiment 1) or pair (Experiment 2). Error bar = ±1 standard deviation. SS = same sound/same source; DS = different sound/same source; DD = different sound/different source.

analyses on a score of overall low-level dissimilarity computed by collapsing information across the 12 acoustical dissimilarities. To this purpose, we: (i) rank-transformed each of the 12 dissimilarities; (ii) normalized each of them between 0 (lowest dissimilarity) and 1 (highest dissimilarity); (iii) averaged the 12 ranked and normalized dissimilarities to yield one final score of low-level dissimilarity. Fig. 2 shows the average low-level dissimilarity score measured within each cell of the experimental design for each of the experiments. A two-factor generalized linear model (GLM) was adopted to ascertain significant effects of the experimental factors on the low-level dissimilarity scores. For Experiment 1, the effect of the interaction between the category and sequence type/N-repetition factors, and the main effect of category were not significant [$F(6,36) = .89$ and $F(2,36) = 1.62$, respectively, $p > .214$]. The number of same-stimulus repetitions had instead a significant effect on low-level dissimilarity [$F(3,36) = 209.24$, $p < .001$]. In particular, a linear contrast showed that low-level dissimilarity decreased significantly with an increase in the number of same-stimulus repetitions [$F(1,36) = 547.11$, $p < .001$]. For Experiment 2, we observed a significant main effect of the category and pair-type factors, and a significant interaction between these factors [$F(2,99) = 47.57$, $F(2,99) = 1120.41$, $F(4,99) = 15.06$, $p < .001$]. Post-hoc contrasts showed that, overall, low-level dissimilarity decreased significantly from the DD to the DS to the SS condition [$F(1,99) > 1382.58$, $p < .001$], and that whereas pairs of environmental sounds and of musical tones did not differ in low-level dissimilarity [$F(1,99) = 0$, $p = .959$], paired sounds from both of these categories were significantly more dissimilar than pairs of vocal stimuli [$F(1,99) > 70.92$, $p < .001$]. Finally, the significant interaction between the category and pair-type factors appeared to be caused by the presence of significant differences between sound categories for the DS and DD conditions [$F(2,33) > 11.87$, $p < .001$], but not for the SS condition, where the dissimilarity between identical sounds was, by definition, equal to 0 for all of the sound categories.

2.2. Participants

Fifteen normal-hearing individuals took part in this study (nine females, six males; median/IQR age = 24/2 yrs; fourteen right handed). Informed consent was obtained from all individuals, and the protocol was approved by the research ethics board of the Regroupement Neuroimagerie du Québec (RNQ) at Université de Montréal.

2.3. Design and procedure

On each trial of Experiment 1, participants were presented with either one of the 48 sound sequences, or with one of 12 8-sec periods of silence, for a total of 60 trials. The order of the sequences and silent periods was established at random for each of the participants, who were instructed to listen passively to the sound stimuli. On each trial of Experiment 2, participants were presented with either one of the 108 sound pairs, or with one of 12 8-sec periods of silence, for a total of 120 trials. Participants carried out a same/different source categorization task: on each trial, they were asked to press one response key with the index if the two sounds were generated with the same source, and a different response key with the middle finger if they were not. Reaction times for this task were measured from the onset of the second sound in the pair. Anticipations and exceedingly long responses (RT > 3 sec from onset of second sound) were considered as missing responses.

2.4. fMRI data acquisition

Participants were scanned with a Siemens 3T Trio scanner, using a Siemens CP head coil. Sound sequences were presented through MRI-compatible electrodynamic headphones (MR-Confon Starter F system with Peltor Optime H510A earmuffs) at a comfortable level. For both experiments, the echo time was 30 msec, and the TR was 10 sec, composed of a 2-sec acquisition time and an 8-sec silent period during which sound sequences were presented against a silent background. During Experiment 1, sequence playback began at the end of the volume acquisition, i.e., ISI = TR. During Experiment 2, the onset of the sound sequence was jittered between 1.5 and 2.5 sec after the end of the volume acquisition. Each brain volume contained 28 slices of 3.2 mm thickness (inter-slice gap = .64 mm) in an axial orientation along the direction of the temporal lobe, providing near full-brain coverage. The in-plane voxel size was 3.2×3.2 mm^2 (64×64 matrix). A whole-brain, high-resolution, structural T_1-weighted MP-RAGE image (176 sagittal slices, 256×256 matrix size, $1 \times 1 \times 1$ mm^3 voxel size) was also acquired to characterize the subjects' anatomy.

2.5. fMRI data analysis

Analyses were carried out using SPM8 and custom Matlab code. Functional images were slice-time corrected to the onset of the first slice and spatially realigned using a 6-parameter affine transformation. The registration procedure considered images from both Experiments 1 and 2. For each of the participants, high-resolution T_1 images were co-registered to the average functional image and segmented into gray matter, white matter and cerebrospinal fluid (Ashburner & Friston, 2005). Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL, Ashburner, 2007) was used to register gray and white matter probability maps for the different individuals and derive a common gray-matter template. An affine transformation was finally estimated to normalize the DARTEL template to MNI space.

For both experiments, the first step of the pipeline for the analysis of the functional images involved fitting a GLM to the data for each participant and estimating the average BOLD response in each of the cells of the experimental design. To this purpose, the first-level GLMs included one condition-specific regressor for each of the cells of the experimental design (12 and 9 condition-specific regressors for Experiment 1 and 2, respectively). Regressors for each of the experimental conditions were obtained by convolving a boxcar function modeling the presentation of the condition-specific sequences with the canonical hemodynamic response function (HRF). For both experiments, the GLM also included: (1) head-motion parameters estimated during the spatial realignment step; (2) an intercept term modeling activation during the implicit silent baseline condition; (3) a standard high-pass

filter (cutoff = 1/128 Hz). For Experiment 2, the first level GLMs also included condition-specific parametric modulators modeling the effects of response correctness on the difference between trials from the same condition, and one additional condition-specific regressor for missing-answer trials. For this experiment, condition-specific activation levels were estimated by considering both correct- and incorrect-response trials (see e.g., VanRullen, 2011; for a discussion of potential analysis biases associated with trial-selection procedures based on behavioral responses). Temporal dependencies between functional images from the same participant were accounted with a standard AR(1) autoregressive model. First-level models were fit to unsmoothed native-space EPI data within a gray matter analysis mask [p(gray matter) $\geq$ .5, as based on segmentation of the T_1 scan].

Second-level random-effect (RFX) models aimed to ascertain the effect of the experimental factors and of their interaction on fMRI activation levels in the population of participants. For both experiments, models were fit to the condition minus silent baseline contrast images (12 and 9 contrast for each of the participants in Experiment 1 and 2, respectively) deformed to the group DARTEL template, smoothed using a Gaussian kernel (8 mm full-width at half-maximum, FWHM), and normalized to MNI space (voxel size after normalization = 2 mm^3). Analyses were carried out within a gray-matter mask [p(gray matter) for group DARTEL template $\geq$ .5]. Contrast images for each experiment were analyzed within a flexible factorial SPM GLM with regressors for: (1) the main effect of each of the two experimental factors [sound category for both Experiment 1 and 2 and sequence type or pair type for Experiment 1 and 2, respectively; for each experimental factor, N regressors = N factor levels]; (2) the interaction between the experimental factors [N regressors = N levels factor 1 $\times$ N levels factor 2]; (3) subject-specific effects [N regressors = N participants]. F tests based on linear contrasts of the regressor-specific GLM estimates were used to ascertain significant main effects and interactions. Significant omnibus F tests (FWE < .05; extent threshold = 10 voxels) were further qualified based on pairwise post-hoc F contrasts carried out within significant omnibus test analysis masks (FWE < .05, Bonferroni-corrected for number of post-hoc contrasts; extent threshold = 10 voxels). For Experiment 1, further post-hoc analyses were carried out to assess significant repetition suppression or enhancement effects within those clusters characterized by a significant effect of the sequence type/N-repetition factor. In particular, repetition-suppression effects were assessed based on a *T* test of the linear contrast with weights [.75, .25, −.25, −.75] for [1, 2, 3, 6 repetitions], whereas repetition-enhancement effects were assessed with a *T* test of the linear contrast with weights [−.75, −.25, .25, .75] for [1, 2, 3, 6 repetitions] (FWE < .05; extent = 10 voxels).

A final set of RFX analyses aimed to assess the extent to which the cortical effects of the experimental manipulations were accounted for by changes in within-sequence low-level dissimilarity across experimental conditions. The analysis strategy followed the pipeline described above for the unaltered fMRI data, with some notable exceptions concerning the first-level models. In general, the first-level models for both experiments aimed to produce residual condition-specific contrast images that did not contain variance explained by the low-level dissimilarities. For Experiment 1, an initial native-space first-level model was fit to the unsmoothed data with one single non-baseline experimental condition, and with the head motion parameters, the intercept term, and the high-pass filter. Importantly, the first-level model also included 12 additional regressors (parametric modulators for non-baseline condition), one for each of the low-level sequence-specific dissimilarities. These low-level regressors had a value of zero for baseline trials and an average value of zero for non-baseline trials, and were convolved with the HRF. These first-level GLMs thus estimated the effect of the low-level dissimilarities on the differences in fMRI activation between non-baseline trials, and did not alter the estimates of the baseline activity. The residuals of the GLM prediction based on the low-level dissimilarities alone, i.e., without considering the baseline intercept term of the GLM, were finally analyzed as specified above for the unaltered fMRI data (first-level activity estimate for each cell of the experimental design and subsequent RFX analysis of the cell-specific contrast images). The same strategy was adopted for Experiment 2 data, exception done for the fact that the initial first-level model fit to estimate the effects of the low-level dissimilarities also contained a separate condition for no-answer trials for which low-level dissimilarities were assumed to have a value of zero. The same residual-based strategy was finally adopted to assess the extent to which the effects of the experimental manipulations in Experiment 2 were accounted for by between-condition differences in RT (RT = 0 for no-answer trials).

3. Results

3.1. *Experiment 1 – passive listening*

Sound category had significant effects on activation levels in two large bilateral STG clusters, extending from posterior to anterior aspects, and comprising the lateral portion of the Heschl's gyrus (HG), and the planum temporale (PT). For both clusters, the effect of sound category appeared to be driven by a stronger activation for vocal sounds than for either music or environmental sounds (see Fig. 3 for details). Importantly, a large portion of the clusters characterized by significant category-sensitivity effects in the analysis of unaltered fMRI data continued to exhibit category sensitivity when the fMRI data were cleaned of the variance explained by low-level dissimilarity [38 and 27% of the unaltered-data category-sensitive voxels in the left and right STG cluster, respectively].

Significant effects of the sequence type/N repetitions factor emerged in three left temporal clusters (middle superior temporal gyrus – mSTG; posterior STG – pSTG; posterior middle temporal gyrus – pMTG), in the left Rolandic operculum (RolOp), in a comparatively larger right cluster in the middle temporal plane (mTP), extending to the STG, and in the posterior aspects of the left middle frontal gyrus (pMFG; see Table 1, and Fig. 4). Post-hoc contrasts between levels of the N-repetitions factor revealed that these effects were, overall, driven by a significantly stronger activation for the 2-repetition sequence relative to the 3- and 6-repetition

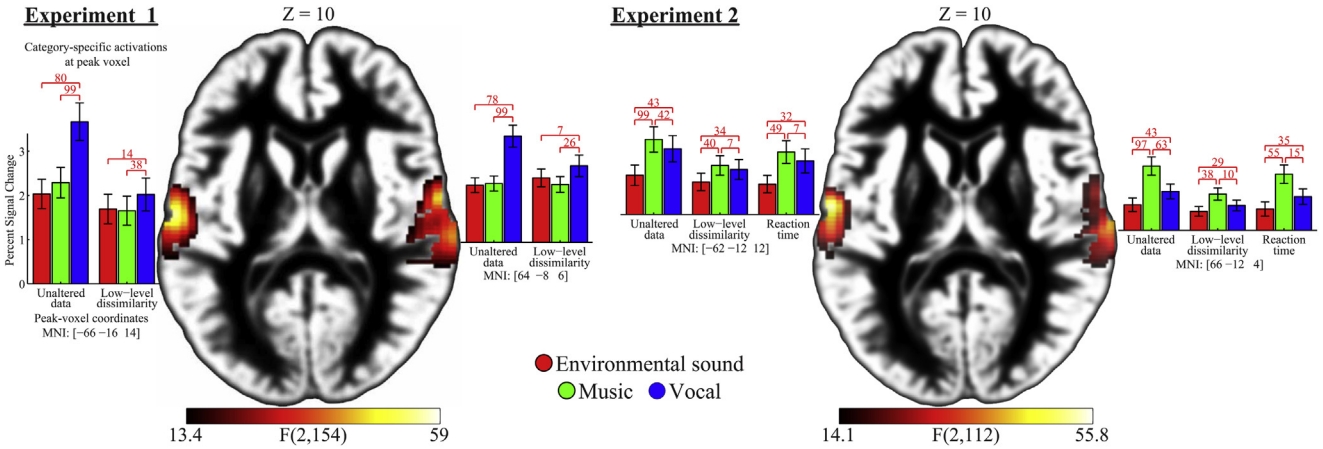

Fig. 3 – Experiments 1 and 2: cortical sensitivity to sound categories. Axial gray-matter probability slices are derived from the group DARTEL template normalized to MNI space, and are displayed along with the statistical parametric maps for assessing a significant omnibus effect of sound category on the unaltered fMRI data (FWE < .05; extent threshold = 10 voxels). For each of the clusters, we show the category-specific percent signal change (PSC) averaged across participants (error bar = ±1 SEM) in the peak-effect voxel for the analysis of the unaltered fMRI data. PSC was computed using condition-specific regressors in the first-level GLMas discussed in Pernet (2014). In particular, rather than using a standard trial height as scaling factor, the PSC was computed in reference to condition [environmental sound/one repetition] in Experiment 1 (scaling factor of 1), and in reference to condition [environmental sound/same sound same source] in Experiment 2 (scaling factor of .5). For the same voxel, we also show the category-specific PSC for the various additional analyses of the fMRI data from which the variance explained by the low-level dissimilarities (Experiments 1–2), or by reaction time in the same/different source categorization task (Experiment 2) was partialled out. The numbers on top of the red lines that connect different bars show the percentage of voxels in the unaltered-data omnibus-effect clusters associated with a significant difference between the connected bars (FWE < .05, Bonferroni-corrected for multiple post-hoc comparisons; extent threshold = 10 voxels; percentage sign omitted). Bars not connected by a red line refer to categories for which the fMRI activations did not differ significantly.

sequences (see Fig. 4 for details), with the exception of an additional significantly stronger activation for 1- than for 6-repetition sequences in the left pMFG cluster, and of the absence of a significant difference between the activation for 2- and 3-repetition sequences in the left pMTG cluster. More importantly, T-based post-hoc contrasts revealed a significant repetition-suppression effect for the vast majority of the voxels in the left pMFG, right mTP and left RolOp clusters (see

Table 1 – Summary of random-effect analyses. All effects significant at FWE < .05, extent threshold = 10 voxels.

Location	Left hemisphere					Right hemisphere				
	Z	C. size	x	y	z	Z	C. size	x	y	z
Experiment 1: Sound category										
STG	>8	975	−66	−16	14	>8	1081	64	−8	6
Experiment 1: N repetitions										
mSTG/mTP	4.77	25	−58	−18	10	5.43[a]	250	54	−20	12
pSTG	5.10	45	−66	−42	16	–	–	–	–	–
pMFG	5.06[a]	128	−52	20	34	–	–	–	–	–
pMTG	4.90	31	−62	−64	2	–	–	–	–	–
RolOp	4.89[a]	44	−52	−26	20	–	–	–	–	–
Experiment 2: Sound category										
STG	>8	698	−62	−12	12	>8	674	66	−12	4
Experiment 2: Pair type										
FrOp/aIns	5.28	93	−38	18	6	5.63	131	42	18	4
pMFG	5.26	132	−48	10	38	5.62	129	54	20	38
SFGpm	5.37	64	0	16	54	4.96	19	4	18	52

Note. C. size = cluster size; mSTG = middle superior temporal gyrus; mTP = middle temporal plane; pMFG = posterior middle frontal gyrus; pMTG = posterior middle temporal gyrus; RolOp = Rolandic operculum; FrOp = frontal operculum; aIns = anterior insula; SFGpm = superior frontal gyrus, pars medialis.

[a] marks clusters that exhibited a significant effect of sequence type/N repetitions, and a significant repetition-suppression effect, see Fig. 4 and text, for more details.

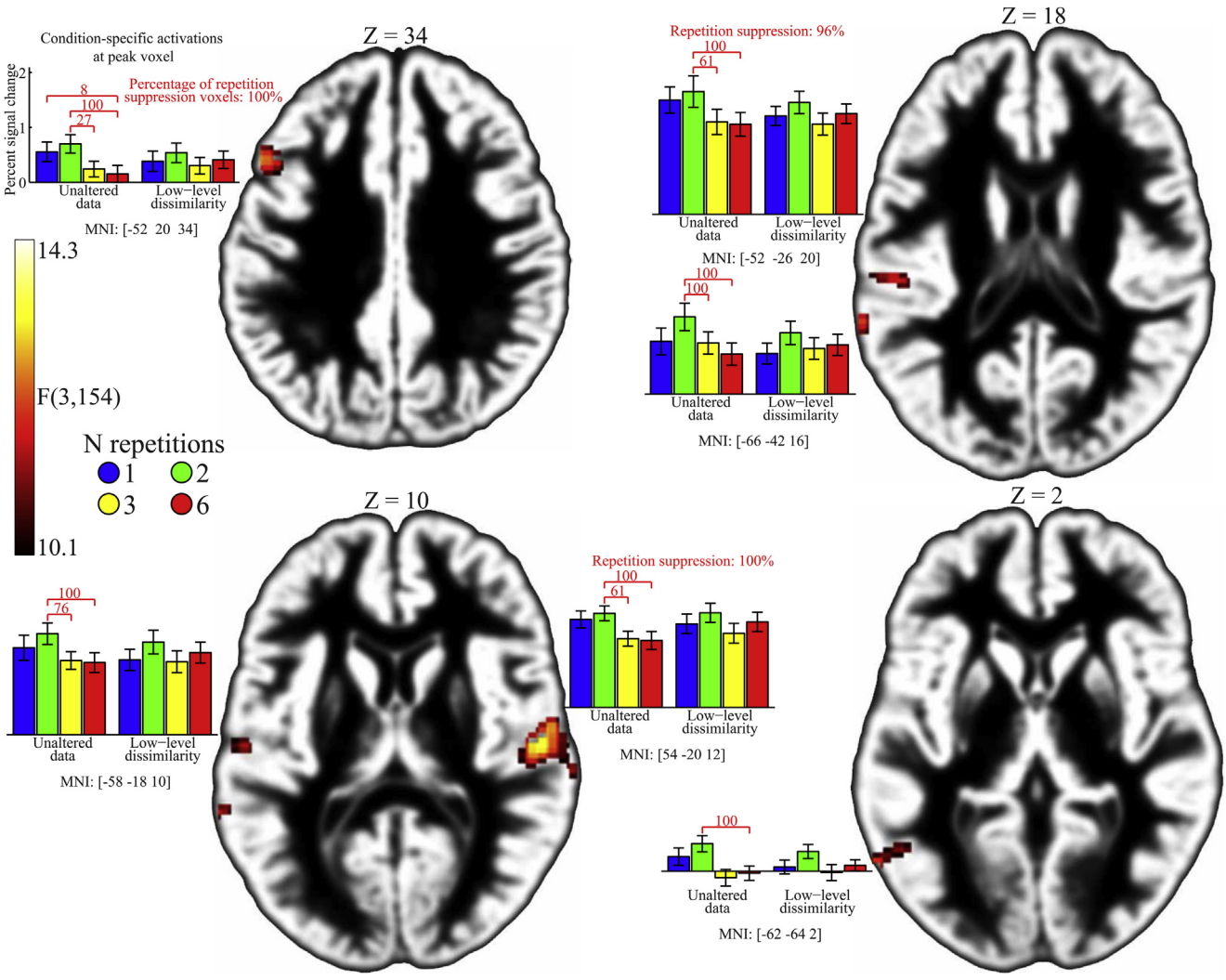

Fig. 4 – Experiment 1: Passive adaptation to source-identity repetitions. Significant omnibus effect of the sequence type/N identity repetitions factor (FWE < .05; extent threshold = 10 voxels). Red bars connect sequence types that induced significantly different levels of fMRI activation (post-hoc contrasts; Bonferroni-corrected FWE < .05; extent threshold = 10 voxels; red numbers = percentage of cluster voxel associated with a significant difference). For each cluster, we also report the percentage of voxels that showed a significant repetition-suppression effect in the post-hoc analyses (FWE < .05; extent threshold = 10 voxels). No cluster was characterized by a significant repetition-enhancement effect. Error bar = ±1 SEM. See legend of Fig. 3 for further details.

Fig. 4), and no repetition-enhancement effect. No significant effect of sequence type/N repetitions emerged after the fMRI-data variance explained by within-sequence low-level dissimilarity was removed. Finally, no significant effect was observed for the interaction between the sequence type/N-repetition and sound-category factors either for the analysis of the unaltered fMRI data or of the same data from which the low-level dissimilarity variance was removed.

3.2. *Experiment 2 – active source-identity discrimination*

3.2.1. *Behavioral results*

A very low number of missing responses (no response; anticipations; RT > 3 sec) was observed [across-participant median of percentage of missing responses = 1.9%; IQR = 3.5%]. They were not considered in the following analyses of the behavioral data. For each participant, we initially computed the proportion of correct answers and the average RT (both correct and incorrect answers) for each of the 9 experimental conditions (3 sound categories × 3 sound–pair types; see Fig. 5). Analyses of the across-participants average proportion correct in each experimental condition revealed performance to be better than a chance-level random-response performance of 50% correct [$t(14) \geq 2.38$, $p \leq .032$]. For each of the participants, we then computed the Spearman correlation between the average RT and proportion correct for the different experimental conditions. A t-test carried out on the average of the Fisher Z-transformed correlations revealed a significant group-average negative correlation between RT and proportion correct [$t(14) = -9.72$, $p < .001$, 95% confidence interval for the average correlation = −.829 to −.639]. The

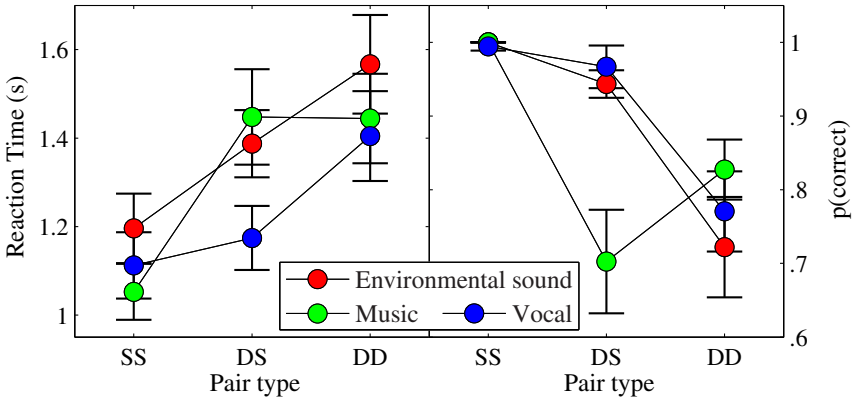

Fig. 5 – Experiment 2: Behavioral discrimination of sound source identities. Across-participants average reaction time (left panel) and proportion correct (right panel) for the same/different source categorization task in each of the experimental conditions. SS = same sound/same source; DS = different sound/same source; DD = different sound/different source. Error bar = ±1 SEM.

same procedure was adopted to assess the significance of the group-average correlation between the average RT and average low-level dissimilarity for the different experimental conditions. The correlation between condition-specific RTs and low-level dissimilarity was significant and positive [$t(14) = 7.15$, $p < .001$, 95% confidence interval for the average correlation = .631–.881], i.e., participants responded more slowly for sounds whose low-level structure was very different.

A 3 × 3 repeated measure ANOVA was carried out to analyze the effects of sound category and pair type on RT. Both the main effect of category and of pair type were significant [$F(2,28) \geq 11.52$, $p \leq .001$]. For the category factor, post-hoc contrasts revealed only significantly faster responses for vocalizations compared to environmental sounds [$F(1,14) = 22.91$, Bonferroni-corrected $p < .001$; $F(1,14) \geq 5.15$, Bonferroni-corrected $p > .05$ for the other post-hoc contrasts]. All of the post-hoc contrasts for the main effect of pair type were instead significant, showing an increase in RT from SS to DS to DD sequences [$F(1,14) \geq 11.05$, Bonferroni-corrected $p \leq .015$]. The ANOVA model also revealed a significant interaction between the category and pair-type factors [$F(4,56) = 5.20$, Greenhouse-Geisser corrected $p = .006$]. The simple effects of both of the experimental factors were all significant [$F(2,28) \geq 4.57$, $p \leq .019$]. A significant category X pair-type interaction could thus be potentially interpreted as revealing a modulation of the strength of the effect of either experimental factor across the levels of the other experimental factor.

3.2.2. *fMRI results*

The modulation effect of pair type on brain activation was significant in six clusters (see Fig. 6): the bilateral pars medialis of the superior frontal gyrus (SFGpm), in a region often labeled as pre supplementary motor cortex (preSMA); the bilateral frontal operculum/anterior insula (FrOp/aIns); the bilateral pMFG)occupying part of the pre-central gyrus. Post-hoc contrasts revealed that all of the voxels in these clusters were most strongly activated by DD than SS pairs. In addition, 55, 64 and 74% of the voxels in the left pMFG, left SFGpm, and right SFGpm were more strongly activated by DD than by DS pairs, respectively. Finally, a minimum of 58% of the voxels in all of the clusters that showed a significant omnibus effect of pair type were most strongly activated by DS than by SS pairs (see Fig. 6).

No significant effect of pair type emerged after the fMRI data were cleaned of the variance explained by low-level dissimilarities or by RT. Two large STG clusters, one per hemisphere, were characterized by a significant effect of sound category (Fig. 3). Voxels in both clusters were activated more strongly by music than by vocal than by environmental-sound pairs. Importantly, neither low-level dissimilarity nor RT explained the totality of the category-sensitivity effects observed for the analysis of unaltered fMRI data. Indeed, when low-level dissimilarities were partialled out, 44 and 31% of the left- and right-hemispheric category-sensitive voxels still exhibited a significant effect of this factor. When RT was partialled out, 52 and 59% of the left- and right-hemispheric category-sensitive voxels retained their category sensitivity. In no analysis did we observe a significant effect of the interaction between sound category and pair type.

3.3. *Conjunction of identity-related effects in Experiments 1 and 2*

A cluster of 58 voxels in the left pMFG [Brodmann area 9/44] exhibited both a significant effect of sequence type/N repetitions in Experiment 1 and of pair type in Experiment 2 [average xyz MNI coordinates of within-cluster voxels = −49 18 35]. Thirty-seven of the voxels in the conjunction cluster were characterized by a significant repetition-suppression effect in Experiment 1. For 33, and 58 of these voxels, DD pairs induced greater activation levels than DS and SS pairs, respectively. Finally, for all of the voxels in this cluster, the effects of N repetitions, that of pair type, and the repetition-suppression effect did not reach significance after the low-level dissimilarity variance was partialled out of the fMRI data. Fig. 7 shows the conjunction of the significant effects of sound-sequence and sound-pair type in Experiments 1 and 2, respectively, and displays the association between the

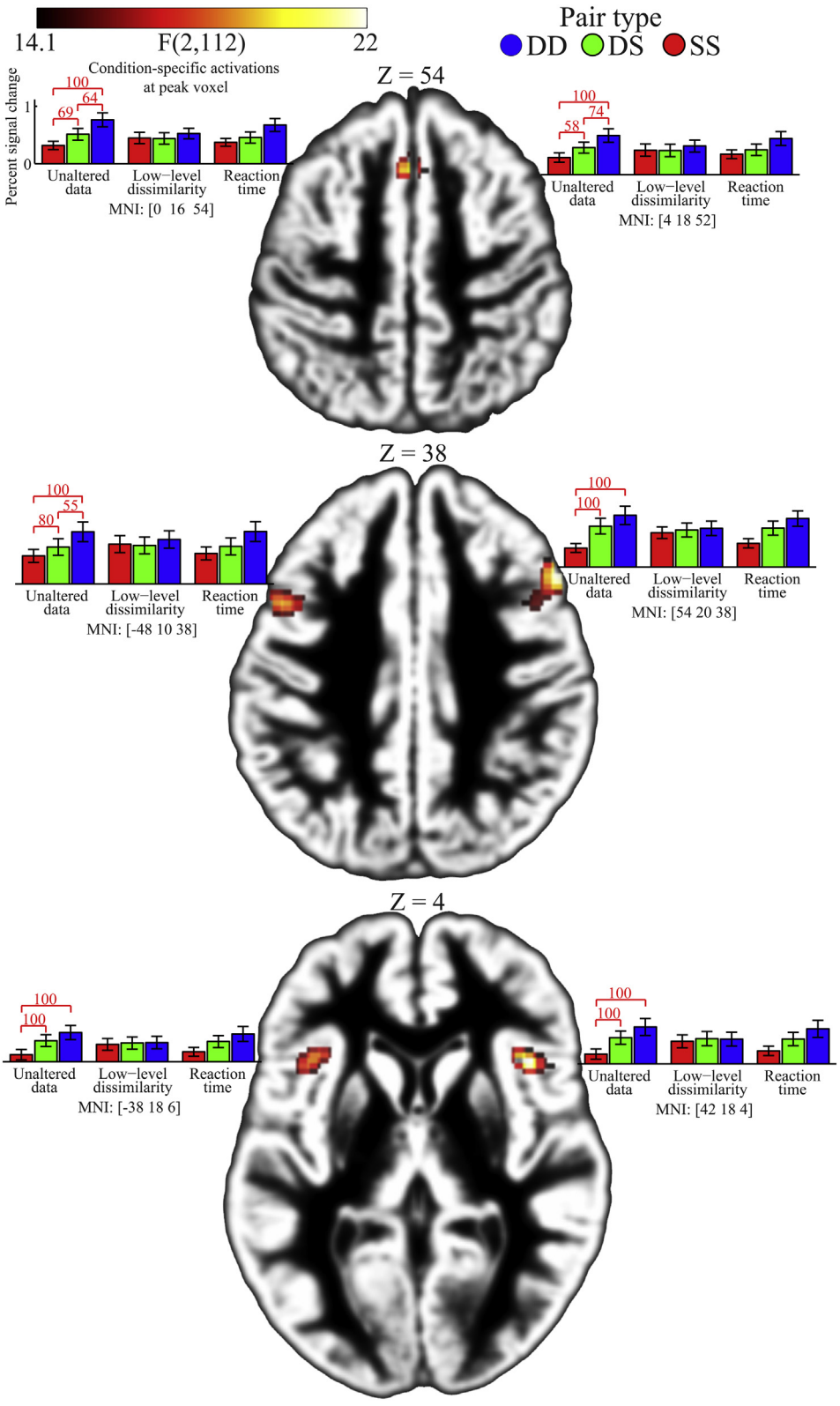

Fig. 6 – Experiment 2: Modulation of cortical activation by pair type during sound-identity discrimination. Significant fMRI effect of pair type (FWE < .05; extent threshold = 10 voxels; DD = different sound/different source; DS = different sound/same source; SS = same sound/same source). Red lines connect pair types that induced significantly different activation levels, as assessed within a post-hoc contrast analysis (Bonferroni-corrected FWE < .05; extent threshold = 10 voxels; red numbers = percentage of voxels that exhibited a significant difference). Error bar = ±1 SEM. See legend of Fig. 3 for further details.

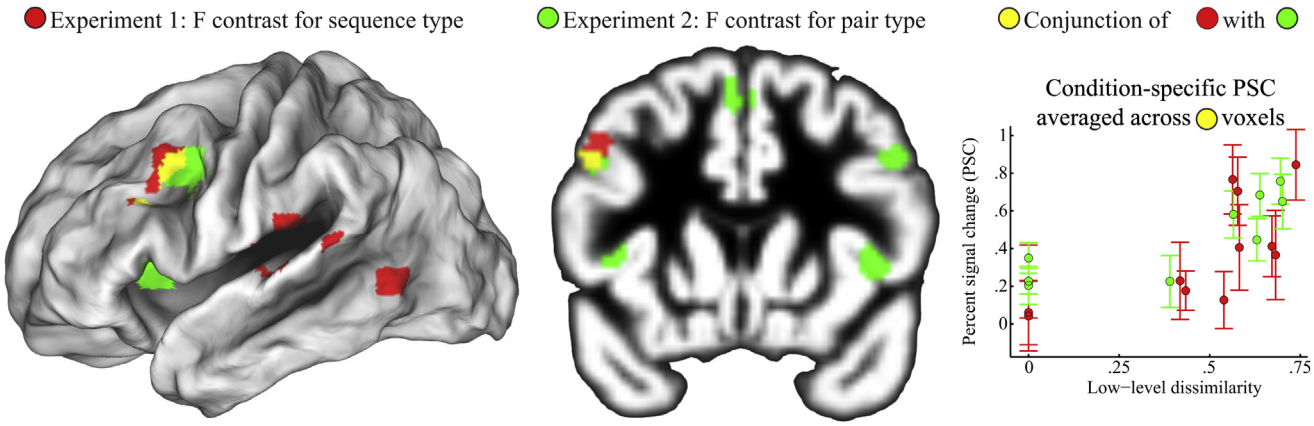

Fig. 7 – Experiments 1 and 2: conjunction of effects of sequence type/N identity repetitions during passive listening and of the effect of pair type during active source-identity discrimination (FWE < .05; extent threshold = 10 voxels). The conjunction reveals a left pMFG cluster of 18 voxels. This cluster was characterized by fMRI adaptation for the repetition of source identities, and by stronger levels of activation for paired sounds generated by different sources. The left panel displays the projection of the significant-effect masks onto the PALS atlas. The right panel displays the association between condition-specific scores of low-level dissimilarity and cerebral activation in both experiments (average percent signal change – PSC – across participants and voxels within conjunction cluster; error bar = ±1 SEM). No effect of the number of same-sound repetitions or of the type of sound pair emerged when the variance explained by low-level dissimilarities was partialled out of the fMRI data. Note that when a more lenient statistical threshold is adopted ($p = .0001$ uncorrected, extent = 20 voxels), a significant conjunction emerges also for the right pMFG (Supplementary Figure).

average condition-specific activation and low-level dissimilarity. Consistently with the analysis of the effects of the experimental conditions on low-level dissimilarity, and with the analyses of the effects of experimental conditions on fMRI activations, voxels within this left pMFG cluster appeared to be most strongly activated by sound-sequences/pairs characterized by a larger low-level dissimilarity.

4. Discussion

Two main results emerged from this study. First, adaptation effects in temporal cortex induced by sound repetition appear independent of sound category and are largely explained by within-sequence similarities in low-level acoustical structure. Second, left pMFG cortex is involved in processing sound source identity in both passive and active listening tasks similarly across sound categories.

4.1. *Adaptation to source identity in the temporal, parietal and frontal lobes*

The level of activation of various cortical areas was modulated by the number of passively-heard sound source identities independently of sound category. All of these effects were accounted for by low-level within-sequence dissimilarity. None of the areas was characterized by a repetition-enhancement effect (Segaert, Weber, de Lange, Petersson, & Hagoort, 2012), i.e., fMRI activation did not increase with the number of identity repetitions. A significant repetition-suppression effect (Grill-Spector, Henson, & Martin, 2006) was instead observed in the right mTP extending to the STG, in the left Rolandic operculum (RolOp), and in the ventral aspect of the pMFG, extending rostrally to an MFG area dorsal to the pars opercularis of the inferior frontal gyrus (IFG). The presence of repetition-suppression effects in the right mTP/STG replicates the results of a large number of imaging studies of natural sounds (Altmann, Bledowski, Wibral, & Kaiser, 2007; Altmann, Júnior, Heinemann, & Kaiser, 2010; Andics et al., 2010; Bergerbest, Ghahremani, & Gabrieli, 2004; Charest, Pernet, Latinus, Crabbe, & Belin, 2013; De Lucia et al., 2010; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Latinus et al., 2011; Zatorre et al., 2004). Consistent with our observation of category-independent adaptation in this region, Doehrmann et al. (2008) observed that a large extent of the right superior temporal plane exhibited repetition-suppression effects for both animal and tool sounds. Importantly, the low-level account for the repetition-suppression effects in this region maps recent observations of cortical sensitivity to the same set low-level features considered in the current study, and in particular, of the selective encoding of the median of the time-varying pitch and HNR, and of the IQR of the time-varying spectral centroid in various location along the right TP/STG (Giordano et al., 2013). Additional studies of speech sounds also confirm the adaptation to low-level structure in this region (Andics et al., 2010; Charest et al., 2013; Latinus et al., 2011).

The Rolandic operculum is often considered to be devoted to processes of auditory-motor integration for speech (Vigneau et al., 2006) and music (Engel et al., 2012). For example, brain imaging and lesion studies link the RolOp to the differentiation, recognition, and gestural reproduction of limb- and mouth-action sounds (Galati et al., 2008; Pazzaglia, Pizzamiglio, Pes, & Aglioti, 2008), or to the differentiation between laughter and speech (Meyer, Zysset, von Cramon, & Alter, 2005). Importantly, activity in this region is also

modulated by action-independent attributes of sound stimuli, such as music pleasantness (Koelsch, Fritz, Müller, & Friederici, 2005) and, consistently with our observation of low-level sensitivity, by timbre-related features of music materials such as brightness, fullness and complexity (Alluri et al., 2011).

Adaptation in the right PFC cortex is reported by three speaker-identification studies (Andics et al., 2010; Latinus et al., 2011; Von Kriegstein & Giraud, 2006). Consistently with our observations, both Andics et al. (2010) and Latinus et al. (2011) report adaptation to low-level structure in this region, although only the latter links this region also to identity-processing mechanisms. The right PFC region reported by previous studies of speaker identity is contralateral to the left pMFG identity-adapting region observed in this study. It should thus be noted that the number of identity-repetitions appears to modulate activity on both the left and right pMFG when a more lenient threshold is considered [$p = .0001$; extent = 20 voxels, see Supplementary Figure]. As such, the divergence in laterality between our study and previous ones could be potentially explained by the presence of semantic speech content in the speech stimuli used in this study but not in previous studies, or by an eventual presence of larger amounts of spectrotemporal variability in the stimulus set investigated in the current study (Boemio, Fromm, Braun, & Poeppel, 2005; Zatorre & Belin, 2001; Zatorre & Gandour, 2008).

Overall, activity in the right mTP, and in the left RolOp and pMFG appears consistent with the hypothesis of domain-general automatic processing of source identity during passive listening, and also reveals that such processing relies on the analysis of sound-to-sound variations in low-level structure. However, based on the results of Experiment 1 alone, it is not possible to state whether these regions implement a mechanism dedicated specifically to the analysis of source-identity information, or whether they are more simply implicated in an analysis of short-term low-level variability active also for sounds that do not contain source-identity information.

4.2. *Discriminating source identities explicitly involves the cingulo-opercular and fronto-parietal control networks*

During an explicit source-identity discrimination task, activity in the bilateral SFGpm, FrOp/aIns and in the pMFG was modulated by the sound-pair type. In general, activity in all of these regions increased from same sound/same source to different sound/same source to different sound/different source pairs, although some important variations across regions concerning the significance of between-pair activation differences emerged. Paralleling the results observed when listening passively to sequences of source identities, the cortical effects of pair type were independent of the category of the paired sounds, and appeared to be accounted for by low-level dissimilarity. Additionally, the effect of pair type in these regions was also accounted for by between-condition differences in reaction times.

In the behavioral task, reaction times increased in the same direction as activation levels in these regions, i.e., from SS to DS to DD pairs, and longer reaction times led to impaired behavioral performance. Confirming published evidence (e.g., Wig, Grafton, Demos, & Kelley, 2005), repetition of either sounds (SS pairs) or of identities (SS and DS pairs) thus produced a behavioral advantage in the source-discrimination task. Consistently, reaction times were significantly longer for sound pairs characterized by a large low-level difference, and, as could also be expected based on the lawful relationship between sound source mechanics and acoustical structure (e.g., Fletcher & Rossing, 1991), sound pairs generated by different sound sources were characterized by a larger low-level dissimilarity than sound pairs generated by the same source.

The SFGpm, an area often labeled as pre-SMA, has been linked to various processes such as the generation and sequencing of motor speech plans (Hartwigsen et al., 2012), or to working memory processes related to the encoding and retrieval of speech materials (Marvel & Desmond, 2010). The anterior portion of the bilateral insula is reported as among the regions most frequently activated in brain imaging studies (Nelson et al., 2010), and recent meta-analyses confirmed the involvement of this region in a large variety of tasks targeting higher-level cognition (Cauda et al., 2012; Chang, Yarkoni, Khaw, & Sanfey, 2013). More specifically, the SGFpm and FrOp/aIns are part of a cingulo-opercular control network responsible of maintaining task-dependent cognitive sets across multiple trials (Dosenbach et al., 2006; Power & Petersen, 2013). The cingulo-opercular network is also considered a "salience" network (Seeley et al., 2007) that produces reliable error-related signals (Dosenbach et al., 2006) and is implicated in the conscious perception of errors (Ullsperger, Harsay, Wessel, & Ridderinkhof, 2010).

The bilateral pMFG region whose activity differentiated between pair types is in close proximity to the right-hemispheric clusters implicated in the analysis of identity and low-level structure by previous studies on the cortical processing of speaker identity (Andics et al., 2010; Latinus et al., 2011; Von Kriegstein & Giraud, 2006). This region is part of a fronto-parietal control network (Dosenbach et al., 2006; Power & Petersen, 2013). The fronto-parietal network is thought to control task sets at shorter time-scales than the cingulo-opercular network (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008), i.e., it is involved in the adjustment of task sets on a trial-to-trial basis, and is less likely involved in error-related processes because it shows less reliable error-related signals (Dosenbach et al., 2006). The fronto-parietal network is also hypothesized to be involved in top-down attentional control (Seeley et al., 2007). Not in contrast with the network model of executive control, for this frontal region the competing cascade model assumes a control process that relies on short-term contextual information (Koechlin & Summerfield, 2007).

The influence of pair type on the activation levels in all of these regions was explained by reaction time. Consistently, for all of these areas Yarkoni, Barch, Gray, Conturo, and Braver (2009) observed a significant increase of activation levels with RT. Our current understanding of executive control in the frontal lobe appears to support the hypothesis that the differential activity for pair type was a performance effect in SFGpm and FrOp/aIns, but not in pMFG. However, the data for Experiment 2 alone are not conclusive to this purpose. Interestingly, a recent meta-analysis of brain-imaging studies of

working-memory reveals that regions in close proximity to the left SFGpm and pMFG clusters observed in this study are part of a "core" working memory network engaged by a wide variety of tasks and contents (Rottschy et al., 2012). Both of these regions are the only ones that, in our study, were characterized by significantly higher levels of activation for different sound/different source pairs than for different sound/same source pairs. In other words, whereas all of the regions influenced by pair type demonstrated significant activity-suppression effect associated with the repetition of the same sound, only these two regions where characterized by a significant adaptation effect for the repetition of source identity.

4.3. *Automatic domain-general processing of source identity in the left posterior middle frontal gyrus*

A region in the left pMFG showed suppressed activity for the repetition of source identities during passive listening, and adapted to the repetition of source identities during the explicit identity-discrimination task. All of these repetition effects were not modulated by sound category, and were accounted for by the higher low-level dissimilarity of different-identity sounds than for same-identity sounds.

The pMFG is considered to implement control processes that operate at short time scales (Dosenbach et al., 2006; Koechlin & Summerfield, 2007; Power & Petersen, 2013). The presence of significant identity-adaptation effects in this area during both passive listening and explicit identity discrimination suggests that the identity-adaptation effects observed during the explicit task are not a mere performance effect related to the processing error signals, but are instead indicative of neural processes involved in the processing of identity information. This view is consistent with the observation that the fronto-parietal control network is characterized by less reliable error signals than regions in the cingulo-opercular network (Dosenbach et al., 2006), from which error signals are thought to originate to mark errors as salient cognitive events (Ullsperger et al., 2010). Conversely, the same anatomical overlap of fMRI effects during passive listening and explicit identity discrimination suggests that during passive listening the pMFG was not merely adapting to the repetition of low-level patterns, but to the repetition of source identities. The intriguing implication of this interpretation is that listeners process identity information automatically when hearing sequences of natural sounds, a setting rather frequent outside the laboratory. Automatic comparisons of source identities is also evocative of the functional interpretation of stream segregation processes, thought to aid the grouping of subsequent bits of auditory information generated by the same sound source (Bregman, 1990).

All of the identity-repetition effects in this region were explained by dissimilarity in low-level structure: sound sequences/pairs whose low-level structure was highly dissimilar activated more strongly the pMFG. Interestingly, empirical evidence suggests that the pMFG or regions in close proximity is particularly sensitive to working memory load (Rottschy et al., 2012), and, in particular, that it is dedicated to updating working memory contents during exposure to widely different stimuli (Nee et al., 2012; Roth, Serences, & Courtney, 2006). This empirical evidence thus suggests that for this prefrontal region, hearing sequences of different-identity sounds, characterized by a widely diverse low-level structure, leads to the update of larger amounts of working memory information, whereas the repetition of identities or the presentation of sounds with a similar low-level structure does not require a complete refresh of working memory storage.

The observation of adaptation to low-level structure in the pMFG is an interesting addition to the debate on the representational architecture of the final projections of the ventral auditory stream to the frontal cortex (Rauschecker & Scott, 2009). The failure of STRF decompositions of the auditory stimulus at accounting for neural activity in the vLPFC of Rhesus macaques has, for example, been considered as evidence of abstract processing of auditory objects in the frontal lobe, independent of low-level structure (Cohen et al., 2007). It should be initially noted that the STRF model is substantially different than the model adopted in this study to characterize the low-level structure of the auditory stimuli. Whereas the former is in essence a decomposition of the sound signal into independent spectrotemporal molecules, our model targets a higher representational level because it considers the statistics of time-varying measures of sensory dimensions such as loudness or pitch. To a very rough approximation, dimensions of auditory sensation are themselves global statistics of molecular sound decompositions. For example, loudness is quite simply the sum of the energy output from different spectral filters. Such statistics of the sound structure guide the perception of natural sounds (Gygi, Kidd, & Watson, 2007; McDermott & Simoncelli, 2011), the recognition of the properties of sound sources (Giordano & McAdams, 2006; Giordano et al., 2010), and the cortical representation of natural sounds (Giordano et al., 2013). An important distinction is however necessary when considering a low-level account of sound processing in the frontal cortex, that between processors, such as the cortical mechanisms for the computation of low-level features, and controllers, mechanisms that influence processor operations based on top-down projections (Power & Petersen, 2013). It is thus unlikely that the pMFG is involved in the computation of low-level features: the very same low-level features considered in this study are indeed likely computed in non-frontal regions such as the temporal cortex (Giordano et al., 2013). More plausible is instead that the pMFG brings together the output of low-level processes in order to aid more complex computations such as the discrimination of the identity of sound sources, whose properties are indeed lawfully reflected in the acoustics of the signals they generate (Fletcher & Rossing, 1991).

4.4. *Category sensitivity in the bilateral temporal cortex*

Consistently with our knowledge of the encoding of natural sound categories in the cortex, activation levels in large extents of the bilateral STG extending both posterior and anterior to A1 were modulated by the category of the sound sequences. Overall, these cortical-sensitivity effects appeared to be independent of the low-level dissimilarity of sounds within the same sequence. This result maps the weak to non-significant effects of category on within-sequence low-level

dissimilarity in both experiments. Based on previous studies of the relationship between cortical encoding of sound categories and low-level structure (Giordano et al., 2013; Leaver & Rauschecker, 2010), part of the category-sensitivity effects observed in our study was likely a product of low-level structure across sound sequences, where category membership varied only across and not within sequences. Overall, in this study we replicate the known sensitivity of the temporal-cortex components of the ventral auditory stream to sound-category information.

All of the identity-adaptation effects revealed in this study were independent of the category-membership of the sound stimuli, i.e., were not modulated by whether sounds were human vocalizations, or musical-instrument tones, or non-speech non-music environmental sounds. Brain-imaging studies of humans reveal that high-level category information is represented selectively, and independent of low-level structure, in the temporal cortex but not in the frontal lobes (Giordano et al., 2013; Leaver & Rauschecker, 2010). Our result thus suggests that the process of identifying sound sources is robust to variations in high-level abstract information. The hypothesis of robust cortical processing of sound source identities is also consistent with its task independence in this study, and is supported by the observation of task- and content-independent control and working-memory processes in the pMFG (Power & Petersen, 2013; Rottschy et al., 2012). It is thus reasonable to hypothesize that the extraction of auditory-object "what" information as complex as source identity does not stop in the ventral frontal cortex, but also takes advantage of the domain- and task-general executive-control machinery implemented in more dorsal aspects of the frontal lobes.

5. Conclusions

We investigated the cortical processing of the identity of speech- and non-speech sound sources. During both a passive-listening and explicit identity-discrimination task, identity repetitions led to reductions of activity in the left pMFG. This region is known for implementing domain-general executive-control functions that operate at short timescales, and working memory processes. Consistently, the cortical adaptation to source identities appeared to be independent of high-level sound-category information, i.e., it showed traits of domain-generality. All of the identity-adaptation effects were explained by the dissimilarity of the low-level sound structure: different sound sources tended to produce dissimilar sounds, and cortical activity was enhanced during the presentation of sounds with a dissimilar low-level structure. The pMFG thus appears to exploit domain-general mechanisms that rely on low-level sound structure when solving complex ecological problems such as the extraction of source-identity information. The presence of cortical effects of sound identity during both passive listening and explicit identity comparison also suggests that when exposed to sequences of natural sounds, a frequent event during everyday life, the cortex automatically engages in processes devoted to the extraction of the identity of sound sources.

Acknowledgments

This work was supported by the Marie Curie Intra-European Fellowships Program (FP7 PEOPLE-2011-IEF-30153, project BrainInNaturalSound to BLG and PB).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2014.06.005.

REFERENCES

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2011). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage, 59*, 3677–3689.

Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage, 35*, 1192–1200.

Altmann, C. F., Júnior, C. G. O., Heinemann, L., & Kaiser, J. (2010). Processing of spectral and amplitude envelope of animal vocalizations in the human auditory cortex. *Neuropsychologia*, 2824–2832.

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage, 52*(4), 1528–1540.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage, 38*, 95–113.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage, 26*, 839–851.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport, 14*, 2105.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309–312.

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience, 16*, 966–977.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8*, 389–395.

Boersma, P., & Weenink, D. (2009). *Praat: Doing Phonetics by Computer* (Version 5.1.05) [Computer program]. Retrieved May, 1, 2009.

Bregman, A. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Cauda, F., Costa, T., Torta, D. M., Sacco, K., D'Agata, F., Duca, S., et al. (2012). Meta-analytic clustering of the insular cortex: characterizing the meta-analytic connectivity of the insula when involved in active tasks. *NeuroImage, 62*, 343–355.

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cerebral Cortex, 23*, 739–749.

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex, 23*, 958–966.

Cohen, Y., Theunissen, F., Russ, B., & Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *Journal of Neurophysiology, 97*, 1470.

De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R., Clarke, S., & Murray, M. (2010). Perceptual and semantic contributions to repetition priming of environmental sounds. *Cerebral Cortex, 20*, 1676.
Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia, 46*, 2776–2786.
Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*, 99–105.
Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron, 50*, 799–812.
Engel, A., Bangert, M., Horbank, D., Hijmans, B. S., Wilkens, K., Keller, P. E., et al. (2012). Learning piano melodies in visuo-motor or audio-motor training conditions and the neural correlates of their cross-modal transfer. *NeuroImage, 63*, 966–978.
Fletcher, N. H., & Rossing, T. D. (1991). *The physics of musical instruments*. New York, NY: Springer-Verlag.
Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "What"? Brain-based decoding of human voice and speech. *Science, 322*, 970–973.
Galati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., et al. (2008). A selective representation of the meaning of actions in the auditory mirror system. *NeuroImage, 40*, 1274–1286.
Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America, 119*, 1171–1181.
Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex, 23*, 2025–2037.
Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 462–479.
Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50*, 331–342.
Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*, 14–23.
Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics, 69*, 839–855.
Hartwigsen, G., Saur, D., Price, C. J., Baumgaertner, A., Ulmer, S., & Siebner, S. R. H. (2012). Increased facilitatory connectivity from the pre-SMA to the left dorsal premotor cortex during pseudoword repetition. *Journal of Cognitive Neuroscience (Early Access)*, 1–14.
Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *NeuroReport, 8*(12), 2809–2812.
King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience, 12*, 698–701.
Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*, 229–235.
Koelsch, S., Fritz, T., Müller, K., & Friederici, A. D. (2005). Investigating emotion with music: an fMRI study. *Human Brain Mapping, 27*, 239–250.
Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-induced changes in the cerebral processing of voice identity. *Cerebral Cortex, 21*(12), 2820–2828.
Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience, 30*, 7604–7612.
Marvel, C. L., & Desmond, J. E. (2010). The contributions of cerebro-cerebellar circuitry to executive verbal working memory. *Cortex, 46*(7), 880–895.
McAdams, S., Roussarie, V., Chaigne, A., & Giordano, B. L. (2010). The psychomechanics of simulated sound sources: material properties of impacted thin plates. *Journal of the Acoustical Society of America, 128*, 1401–1413.
McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron, 71*, 926–940.
Meyer, M., Zysset, S., von Cramon, D., & Alter, K. (2005). Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Cognitive Brain Research, 24*, 291–306.
Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology, 87*, 516–527.
Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory–filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*, 750–753.
Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., et al. (2012). A meta-analysis of executive components of working memory. *Cerebral Cortex, 23*, 264–282.
Nelson, S. M., Dosenbach, N. U. F., Cohen, A. L., Wheeler, M. E., Schlaggar, B. L., & Petersen, S. E. (2010). Role of the anterior insula in task-level control and focal attention. *Brain Structure and Function, 214*, 669–680.
Pazzaglia, M., Pizzamiglio, L., Pes, E., & Aglioti, S. M. (2008). The sound of actions in apraxia. *Current Biology, 18*, 1766–1772.
Peelle, J., Johnsrude, I., & Davis, M. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience, 4*.
Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology, 23*, 223–228.
Pernet, C. R. (2014). Misconceptions in the use of the general linear model applied to functional MRI: a tutorial for junior neuro-imagers. *Frontiers in Neuroscience, 8*.
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*, 718–724.
Romanski, L. M., Averbeck, B. B., & Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *Journal of Neurophysiology, 93*, 734–747.
Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman–Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience, 1999*, 1131–1136.
Roth, J. K., Serences, J. T., & Courtney, S. M. (2006). Neural system for controlling the contents of object working memory in humans. *Cerebral Cortex, 16*, 1595–1603.
Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *NeuroImage, 60*, 830–846.
Seeley, W. M., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. G., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*, 2349–2356.
Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2012). The suppression of repetition enhancement: a review of fMRI studies. *Neuropsychologia, 51*, 59–66.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education Inc.

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function, 214*, 629–643.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology, 2*.

Vigneau, M., Beaucousin, V., Hervé, P., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage, 30*, 1414–1432.

Von Kriegstein, K., & Giraud, A.-L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, e326.

Wig, G. S., Grafton, S. T., Demos, K. E., & Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nature Neuroscience, 8*, 1228–1233.

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., & Braver, T. S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. *PLoS One, 4*, e4257.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience, 24*, 3637–3642.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 363*, 1087–1104.