2018

Who's That Knocking at My Door? Neural Bases of Sound Source Identification

Guillaume Lemaitre

John A. Pyles

Andrea R. Halpern
Bucknell University, ahalpern@bucknell.edu

Nicole Navolio

Matthew Lehet

See next page for additional authors

Recommended Citation

Authors

Guillaume Lemaitre, John A. Pyles, Andrea R. Halpern, Nicole Navolio, Matthew Lehet, and Laurie M. Heller

ORIGINAL ARTICLE

Who's that Knocking at My Door? Neural Bases of Sound Source Identification

Guillaume Lemaitre[1,3], John A. Pyles[1], Andrea R. Halpern[2], Nicole Navolio[1], Matthew Lehet[1] and Laurie M. Heller[1]

[1]Carnegie Mellon University, Department of Psychology and Center for Neural Basis of Cognition, Pittsburgh, PA 15213, USA, [2]Bucknell University, Department of Psychology, Lewisburg 17837, PA, USA and [3]Current address: Ircam, one place Stravinsky, 75004 Paris, France

Address correspondence to Guillaume Lemaitre, Email: GuillaumeJLemaitre@gmail.com

Abstract

When hearing knocking on a door, a listener typically identifies both the action (forceful and repeated impacts) and the object (a thick wooden board) causing the sound. The current work studied the neural bases of sound source identification by switching listeners' attention toward these different aspects of a set of simple sounds during functional magnetic resonance imaging scanning: participants either discriminated the action or the material that caused the sounds, or they simply discriminated meaningless scrambled versions of them. Overall, discriminating action and material elicited neural activity in a left-lateralized frontoparietal network found in other studies of sound identification, wherein the inferior frontal sulcus and the ventral premotor cortex were under the control of selective attention and sensitive to task demand. More strikingly, discriminating materials elicited increased activity in cortical regions connecting auditory inputs to semantic, motor, and even visual representations, whereas discriminating actions did not increase activity in any regions. These results indicate that discriminating and identifying material requires deeper processing of the stimuli than discriminating actions. These results are consistent with previous studies suggesting that auditory perception is better suited to comprehend the actions than the objects producing sounds in the listeners' environment.

Key words: action sounds, auditory cognition, dorsal and ventral pathways, sound identification

Introduction

Behavioral and neuroimaging studies investigating the perception of environmental sounds typically address different issues (i.e., sounds produced in a human environment, excluding music, speech, and animal vocalizations). On the one hand, behavioral studies have almost exclusively focused on the *objects* producing the sound: their shape, size, or material (Lakatos et al. 1997; Klatzky et al. 2000; Kunkler-Peck and Turvey 2000; Carello et al. 2005; Giordano et al. 2010; Lemaitre et al. 2010; McAdams et al. 2010; Houix et al. 2012). On the other hand, many auditory neuroimaging studies have focused on sound of human-generated *actions*. No study has contended with the fact that *both* types of information are usually simultaneously available in sounds:

when someone is knocking on a door, listeners hear both the knocking *and* the door. We recently compared identification of objects and actions for the *same* sounds, and found a notable advantage for action identification (Lemaitre and Heller 2012). Following up on these results, the present work studied neural activations elicited by focusing participants' attention either toward the objects or the actions that created the same sounds in a discrimination task.

Auditory processing is hierarchically organized, with processing becoming more abstract as the distance to primary auditory cortex increases (Binder et al. 2000; Goll et al. 2011; see also Giordano et al. 2010 for a review). One model further characterizes the auditory system as having both a ventral and a

dorsal stream, mirroring the dual stream model for vision (Goodale and Milner 1992; see McIntosh and Schenk 2009 for review). In this model, the ventral stream projects anteriorly along the superior temporal gyrus (STG) toward the inferior frontal gyrus (IFG), and the dorsal stream projects posteriorly through STG and middle temporal gyrus (MTG) toward inferior parietal lobule (IPL) and intraparietal sulcus (IPS) and then to motor and premotor regions in the frontal lobe. The ventral stream (the "what pathway," "hearing for perception") is interpreted as processing invariant object properties, and the dorsal stream (the "how" and "where" pathway, or "hearing for action") processes spatial properties necessary to guide actions (Alain et al. 2001; Arnott et al. 2004; Zatorre et al. 2004, 2007; Borst et al. 2011; Arnott and Alain 2011).

The dual stream hypothesis is illustrated by a series of studies of human action sound identification (Lewis et al. 2004, 2005, 2011, 2012; Lewis 2006, 2010; Engel et al. 2009). Overall, these studies show a striking distinction between the cortical networks involved in processing different types of sounds. Sounds made by human actions activate a large network of cortical areas (the "action-sound network"): a left-lateralized frontoparietal network corresponding to the dorsal stream, and a bilateral complex located in the posterior superior temporal sulcus and MTG (the pSTS/MTG complex). In contrast, sounds produced by nonliving (nonbiological) sources preferentially activate cortical regions affiliated with visual form, feature and object recognition (occipital and parahippocampal regions), corresponding to the ventral stream.

The frontoparietal network includes IFG, inferior frontal sulcus (IFS), and particularly the junction of the IFS and the precentral sulcus (corresponding to the ventral premotor cortex (VPMC)). It also includes a region centered on the IPS and spreading to the inferior and superior parietal lobules (IPL and SPL). IFG is thought to be binding sounds' spectrotemporal features and conceptual representations (Adams and Janata 2002; Amedi et al. 2007; Kaplan and Iacoboni 2007) and motor programs (in VPMC in particular, see Keysers et al. 2003; Aglioti and Pazzaglia 2010). The functional roles of IPS and IPL are usually interpreted as the integration of auditory information about actions with other sensory modalities (mainly vision), and a form of "action replay" (Arnott et al. 2008). Furthermore, this frontoparietal network overlaps with regions activated when participants manipulate tools. Therefore, Lewis and colleagues interpreted the frontoparietal network as "praxis" or "audio-motor association" system, presumably linking one's own motor repertoire with sound production within the dorsal stream. The pSTS/MTG complex is sensitive to audio, visual, and audiovisual presentations of nonbiological and biological motions (Beauchamp et al. 2002, 2003, 2004; Bidet-Caulet et al. 2005; Grossman et al. 2005; Doehrmann et al. 2008), and even words denoting action (Kiefer et al. 2012). Activity in this region is usually interpreted as reflecting a multimodal (and possibly amodal or supramodal) representation of biological motion.

Fewer neuroimaging studies have focused on auditory recovery of object properties, with somewhat inconsistent results (see for instance Arnott et al. 2008). In particular, James et al. (2011) had participants match 2 consecutive sound stimuli produced by dropping objects either based on the material or the shape of the objects. Although the stimuli were the same, the experiment shifted the participants' attention to either the shape or the material of the objects producing the sounds. Their results showed that matching of shape resulted in more activation in an area located at the junction of the left posterior MTG and the occipital lobe. Because this region is also sensitive to

visual and haptic shape identification, the authors interpreted it as an amodal shape operator. There was no evidence of a material specific area, although a more liberal criterion resulted in activity in the lingual gyri, a visual area found for material processing (Cant and Goodale 2007). Note also that activation of early visual areas by auditory stimuli has been reported in other studies (Vetter et al. 2014).

In a recent study, we compared the cortical areas coding different actions (hitting, shaking, and squeezing), different objects (liquids and solids), and different sound events resulting from the combination of the actions and objects (Lehet et al. submitted). The results showed that there is a large overlap between the cortical networks coding the different actions and the different sound events (including the dorsal auditory pathway, the hand motor area, and some visual area), whereas only a small area in the primary auditory cortex is responsive to the different materials. This suggests that listening to environmental sound events is coupled to identifying underlying actions.

Many of the aforementioned studies have contrasted different categories of sounds (e.g., sounds of actions vs. control sounds) defined by both their semantic content (e.g., perceived action or not) and consisting of different sounds. Thus, Giordano et al. (2012) have noted that some of the differences observed between action sounds and their controls could have been at least partially driven by acoustic differences. Furthermore, because all previous studies focused participants' attention toward the actions or the objects creating the sounds (but not both), it is also not clear how much of the observed activations reflected the result of the sound categories or the participants' attention to a particular type of information.

The goal of the current research was to isolate the influence of selective auditory attention toward object or action on the cortical activity within the networks subserving sound identification. We used sets of sounds for which listeners could successfully identify both the action and the object, and we manipulated the focus of attention toward either action or object properties by using different discrimination tasks. These sounds consisted of recordings of easily discriminable sounds created by tapping and scraping metal and plastic pipes. Each stimulus consisted of a pair of sounds, and we focused participants' attention by requiring them to indicate whether the actions or the material of the 2 sounds were the same or different. We also added a control task in which participants compared scrambled versions of the action sounds (meaningless noises sharing the same long-term acoustical properties of other stimuli). This allowed for the comparison of 2 "meaningful" tasks (discrimination based on material or action) and an "acoustic" task (discrimination based solely on meaningless acoustic differences). We used a slow event-related design with different blocks (one task per block), with a fixed inter-trial interval (one stimulus per trial; see also Figure A in Supplementary material).

Two aspects of the design could modulate neural activity: the selection and processing of relevant auditory features, and the change of attention from block to block (see Petersen and Posner (2012) for a review of the different roles of attention). Several studies have shown that attention to different auditory features (spatial location, pitch, intensity, and timbre) can modulate activity at all levels of auditory processing, including auditory cortices, brainstem and the cochlea, though there are still debates as to whether attention does modulate activity in brainstem and cochlea (Pugh et al. 1996; Lipschutz et al. 2002; Falkenberg et al. 2011; Paltoglou et al. 2011; see Fritz et al. 2007; Caporello Bluvas and Gentner 2013 for a review). Attention can also modulate cortical activity in auditory areas thought to

perform early categorization of sounds (Ahveninen et al. 2006; Bidelman et al. 2013). For instance, Mesgarani and Chang (2012) showed that activity in posterior STG could predict which speaker listeners were attending to while they were hearing a mixture of 2 concurrent speech signals. Therefore, a first hypothesis was that attention-based differences between tasks would modulate activity in primary and secondary auditory regions.

It is in theory possible that discriminating action or material would rely only on processing auditory features (e.g., attack time for discriminating actions, spectral decay for discriminating material), without involving deeper processing of the identity of the sound sources: listeners could simply listen to acoustic parameters without making sense of the sounds. However, Kaplan and Iacoboni (2007) have reported activity in the action-sound network even when participants were passively listening to the sounds, suggesting that this network would process the sounds' identity even if the listeners were only attending the sounds' superficial properties. Thus, another hypothesis was that selective attention toward objects or actions would be reflected in the cortical networks subserving sound identification (action and object). In such an account, focusing on the actions that produce the sounds may result in an overall increased activity in the action-sound network (Lewis 2010). It may also result in different patterns of activity reflecting the coding of the different pieces of information.

Another possibility was that the changes of attention themselves could modulate cortical activity. For instance, Corbetta and Shulman (2002) have shown that there is a frontoparietal network that responds to changes of visual attention toward different locations, shape or color (i.e., top down control of attention): the dorsal attention network. It comprises the intraparietal cortex, the superior frontal cortex (frontal eye field). Frontoparietal signals driven by attentional shifts are observed in most of the studies manipulating auditory attention, suggesting a domain-independent role of this executive network (see for instance Shomstein and Yantis 2006 and Falkenberg et al. 2011). The medial superior parietal lobe in particular is thought to be a cortical hub for the initiation of attention and task shifts in multiple domains (Greenberg et al. 2010).

In sum, our hypothesis was that selective auditory attention to object or action could potentially be reflected in the following cortical regions: Transverse temporal gyrus and surrounding regions in STG (processing of acoustic features); left IFG (access to conceptual representations of objects); left premotor cortex (PMC) and in particular its ventral part (VPMC linking sensory inputs to motor repertoire, i.e., frontoparietal network of action representations); IPL, the temporoparietal junction (TPj), and in particular IPS in left hemisphere (multimodal integration, action replay, frontoparietal network of action representations); bilateral pSTS/MTG complex (representation of biological motion); bilateral parahippocampal regions (anterior cingulate gyrus in particular: access to semantic information, visual recognition); bilateral primary occipital regions, lingual and fusiform gyri (visual processing of surface properties, potentially also activated by auditory inputs). In addition, we also expected the attention shifts to be reflected in the executive dorsal attention network.

We used 2 different types of analyses: a general linear model analysis (GLM), and a multivoxel pattern analysis (MVPA). The GLM analysis highlighted clusters of voxels whose average cortical activity (percent change of blood oxygenation level) was significantly different between the different tasks. In contrast, the MVPA identified distributed patterns of activity across voxels that were different across the different tasks (Norman et al. 2006; Davis et al. 2014).

Materials and Methods

All participants provided written consent to participate in this study. This study was approved by the Carnegie Mellon University Institutional Review Board (HS14-339: Neural Bases of Sound Event Identification).

Participants

A total of 23 participants (13 females and 10 males) participated in the study. The data for 6 participants were excluded from the study due to excessive movement in the fMRI scanner related to the procedure for fitting the headcoil. All participants were right-handed (participants filled out the Edinburgh questionnaire, Oldfield 1971), were not highly musically trained (median 0 years, between 0 and 8 years of formal musical training—no formal training during the last 5 years) and also had normal hearing verified via an audiogram. Participants were selected on the basis of a screening test, during which they performed the same experimental procedure as in the scanner (but with fewer sounds; see the description below). Only participants who made a minimal number of mistakes (i.e., <20%), identified the actions and objects equivalently well, and responded with equivalent reaction times to both tasks were selected for the main experiment. The resulting selection of 17 participants (6 males and 11 females) ranged from 18 to 34 years old (median 20 years old).

Stimuli

As shown in previous studies (Lemaitre and Heller 2012), action is fairly easy to identify, whereas this is not necessarily the case for material. Therefore, we conducted 2 separate behavioral experiments for the specific purpose of creating sounds with easily discriminable materials and actions. In the first step, we manually generated sounds by performing a variety of actions on hollow cylinders of different materials to determine the combinations of materials and actions that would be easiest to identify by their sound. Participants in a pilot experiment attempted to identify and discriminate actions and materials. This procedure selected copper and ABS plastic as the most-easily discriminable materials, and scraping and tapping the cylinders on a heavy wooden board as the most-easily discriminable actions. Overall, 10 exemplars were created for each combination of the 2 actions and 2 materials (tap-metal, tap-plastic, scrape-metal, scrape-plastic), resulting in 40 sounds. The actions of the tapping and scraping sounds were altered to produce different exemplars; for example, the speed of the scraping action was varied. In the second step, participants attempted to discriminate and identify the action and material from listening to the sound. They listened to different pairs of sounds and judged whether each pair was caused by 2 different actions (or materials) using a 6-point scale ranging from "Very sure" to "Not sure." Responses were considered accurate when the 2 actions (or materials) were different and the participant used the left-hand-side of the scale (and vice versa). We selected the 24 sounds (6 sets of 4 sounds) that resulted in high accuracy for both action and material discrimination and equivalent average accuracy between action and material discrimination.

The sounds were recorded with a 96-kHz sampling rate and a 24-bit resolution with a Tucker Davis Technologies MA3 microphone preamplifier, a Sound Devices USBPre soundboard, and an Earthworks QTC 30 microphone in an IAC double-walled sound-attenuated booth.

In addition to the cylinder sounds, we created meaningless versions of the same 24 sounds by scrambling the phase of the sounds, following the procedure described by Gazzola et al. (2006). Phase scrambling consists of replacing the phase of each coefficient of the Fourier transform of the sound by the phase of another coefficient, randomly chosen. The resulting sounds have the same long-term amplitude spectrum as the original sounds, but a completely different temporal envelope and are thus unidentifiable.

The stimuli used in the experiment consisted of pairs of sounds. For each of the 6 sets of sounds we selected 8 pairs. The correct answer to each pair depended on the task (judging action or material). For instance, the pair "tap-metal/scrape-metal" corresponds to different actions but same materials. The selection of these 8 pairs insured that there was always the same number of same and different answers for the 2 tasks. Thus, there were a total of 48 pairs of sounds (i.e., 48 stimuli) in the experiment, and 8 types of pairs reported in Table 1 (We did not cross every sound with every other sound in every possible order in order to keep the test runs short enough.). The stimulus duration (for each pair) ranged from 1.5 to 3.4 s (average 2.1 s).

Prior to the main experiment, sounds were resampled to 44.1 kHz, 16 bits, and their levels were adjusted to produce the same loudness: the experimenters listened to each sound and adjusted their level so that they perceived them at the same loudness and ensured that they were clearly audible against the background noise while being at a comfortable level.

Procedures

During the main experiment, participants lay supine in the bore of the MRI scanner and responded with an MR-compatible button box affixed to their left hand with tape, using their third finger to respond "same," or index finger to respond "different." Reaction times (i.e., the asynchrony between the offset of the sound and participant's response) and accuracy of the responses were recorded. Using the left hand ensured than any activation observed in left motor or premotor areas could not be attributed to participants' motor response.

Auditory stimuli were delivered using Matlab 7.10.0 and Psychophysics toolbox 3.0.9 (Brainard 1997) on an Apple MacBook Pro running Mac OS 10.6.8, through Sensimetrics S14 insert earphones. Visual stimuli were projected onto a screen located behind the participant. Participants viewed the screen through a mirror located above the head coil.

Three tasks occurred in the main experiment: Noise, Action, and Material. In the Noise task, participants heard the scrambled versions of the stimuli and judged if the 2 sounds were the same or different. In the Action task, participants listened to intact stimuli (pairs of sounds) and judged whether actions were the same or different, and in the Material task they judged whether the materials were the same or different. Material and Action tasks are therefore "meaningful tasks" whereas the Noise task is a purely "acoustic" task. Each stimulus was presented twice for each task, resulting in a total of 288 trials (48 stimuli times 2 repetitions times 3 tasks).

The experiment used a slow event-related design. It was divided into 8 runs. Each run consisted of 6 blocks. Each block corresponded to 1 of the 3 tasks and consisted of 6 trials. In each run, the 3 tasks were presented twice each, and the order of the tasks was randomized for each participant (with the constraint that the same task could not occur twice in a row). The 288 trials were randomized for each participant, with the constraint that the same pair could not occur twice in a row.

Each run began with a fixation cross, displayed for 10 s, followed by the 6 blocks. Each block began with the name of the task displayed for 5 s, followed by the fixation symbol for 5 s, and then 6 trials. The fixation symbol consisted of the black letter N for the Noise task, the blue letter A for the Action task, and the green letter M for the Material task. The fixation symbol was displayed for the whole duration of the block (70 s). Each trial began with the pair of sounds (the maximum stimulus duration was 3.5 s) followed by a period in which the participants could answer. The inter-trial interval was fixed to 10 s (Figure A in supplementary material). Because the stimulus duration was different for each type of pair (Table 1), the interstimulus interval was variable, ranging from 6.5 to 8.4 s (average 7.8 s).

Before proceeding to the main experiment in the scanner, participants performed a 30-min practice session. They listened to each sound individually and were presented with its label (e.g., "metal-scrape"). Then, they made same-different judgments in 3 blocks (Action, Material, Noise), randomly selected from the main experiment. In the rare cases where the participants made more than a few mistakes, they did another set of 3 blocks. The goal of this practice session was to further train participants and to ensure that they could discriminate actions and materials with near-perfect performance.

Imaging and Analysis Parameters

Magnetic Resonance Imaging Parameters

Whole brain functional magnetic resonance imaging (fMRI) data were collected from each participant on a 3 T Siemens Verio MR scanner at the Scientific Imaging & Brain Research Center at Carnegie Mellon University using a phased array 32-channel head coil. Functional images were acquired using a T2*-weighted echoplanar imaging pulse sequence (36 oblique axial slices, in plane resolution 3 mm × 3 mm, 3 mm slice thickness, no gap, repetition time TR = 2000 ms, echo time TE = 29 ms, flip angle = 79°, GRAPPA = 2, matrix size = 64 × 64). High-resolution anatomical scans were acquired for each participant using a T1-weighted MPRAGE sequence (1 mm × 1 mm × 1 mm, 176 sagittal slices, TR = 2300 ms, T1 = 900 ms, FA = 9°, GRAPPA = 2). Anatomical scans were acquired between runs 4 and 5.

fMRI data were analyzed with AFNI (version 2011_12_21_1014). Preprocessing included linear trend removal, spatial smoothing (with a Gaussian kernel of 6 mm full-width at half maximum), slice-timing correction, and 3D motion correction. Trials with large motions were excluded from analyses and motion parameters were used as regressors of no interest. Incorrect and missed trials were also removed from analysis. Table 2 reports the percentage of incorrect and missed trials in each experimental condition.

Table 1 The 8 types of pairs of sounds used in the experiment. There were 6 different pairs for each type of pair, resulting in a total of 48 pairs

	TP	TM	SP	SM
TP	TP–TP (1.8 s)			
TM	TM–TP (2.1 s)			
SP	SP–TP (1.8 s)	SP–TM (2.2 s)		
SM	SM–TP (2.1 s)	SM–TM (2.5 s)	SM–SP (2.2 s)	SM–SM (2.5 s)

TP is "tap-plastic", TM is "tap-metal", SP is "scrape-plastic", SM is "scrape-metal". The average duration of each type of pair is reported in parenthesis.

Table 2 Percentage of incorrect and missed trials for each experimental condition

Task/ type of pair	TPTP	TMTP	SPTP	SPTM	SMTP	SMTM	SMSP	SMSM	Average
Noise	0.0	0.7	0.0	0.7	11.2	8.0	0.0	2.2	2.9
Action	0.0	0.4	1.1	6.2	2.2	1.8	1.1	1.1	1.7
Material	0.0	0.0	2.9	3.6	4.7	12.3	0.7	1.1	3.2
Average	0.0	0.4	1.3	3.5	6.0	7.4	0.6	1.4	2.6

Anatomical scans volumes were transformed into the common space of Talairach and Tournoux using a 12-parameter affine transformation. Functional volumes were coregistered to the anatomical volumes thus transforming them into the common stereotactic space.

GLM Analysis

Individual BOLD responses were modeled using a separate explanatory variable for each combination of the 2 factors (the 3 Tasks and 8 Types of pairs), as well as another explanatory variable for the first 10 s at the beginning of each block where the instructions changed to indicate the different tasks (factor Instruction change with 3 levels: Noise, Action, Material). Design matrices were constructed from predictors generated based on the timing of the protocol for the placement of a canonical double-gamma function (3.5 s, the duration of the longest stimulus).

In a first group-level analysis, the amplitudes of the canonical responses fitted to the BOLD signal were averaged for each participant and combination of the 2 factors (Types of pairs and Tasks) for the experimental trials and fed to a general linear model analysis (GLM) with the participants as a random factor, the Task (3 levels: Noise, Action, Material) and the Type of pair (8 levels, Table 1) as fixed factors, and including the interactions between the factors. A second, separate group-level analysis focused only on the first 10 s of each block by treating the averaged canonical response to GLM with participants as a random factor, and the Instruction changes as a fixed effect.

All whole brain contrasts were thresholded using random field theory. The number of contiguous voxels required providing an alpha-value of 0.05 cluster-wise was estimated using 3dClustSim, based on an estimation of spatial smoothness. Spatial smoothness was estimated to be 9.0 mm × 9.2 mm × 8.0 mm. Therefore, random field theory indicates that a cluster-wise P-value of .05 can be achieved by using a voxel-wise P-value of .01 and a minimal cluster-size of 80 voxels.

Multivoxel Pattern Analysis

Data for the MVPA analyses were left in native space. Deobliquing, slice time correction, linear detrending, motion correction, and co-registration were still applied to the data. MVPA searchlight analysis was performed by walking a sphere with a radius of 5 voxels through the brain images and then classifying each sphere using a linear support vector machine (SVM, Princeton Multi-Voxel Pattern Analysis (MVPA) Toolbox (https://code.google.com/p/princeton-mvpa-toolbox/, last retrieved on 2 December 2015.), Chang and Lin 2011). Functional responses within the searchlight for each trial were classified based on Action or Material trials. Seven runs of data were used to train the classifier and then tested against one run of data left out. This training and testing was repeated until each run had been left out once (each run contained 2 blocks of each task, and each block 6 trials). The centroid voxel of each sphere was assigned the classification accuracy of the surrounding sphere. The temporal window of the time series

submitted to the classifier was chosen to maximize the hemodynamic response function (for details of this searchlight process see Nestor et al. 2011). The classifier's ability to distinguish between the 2 tasks at each voxel should reflect the encoding of that information in the surrounding cortex.

Group level accuracy maps were made by warping the individual accuracy maps to Talairach space, subtracting chance accuracy from them (to normalize around zero), and then performing single sample t-tests at each voxel against the null for each voxel. Only voxels that had data from every participant were tested in this way. The P-values from these t-tests were then corrected for multiple comparisons using a false discovery rate (FDR, Benjamini and Hochberg 1995) with a $q < = 0.05$ ($P < 0.000376$). The t-test maps were thresholded at the critical P value identified by this FDR and a mask was created showing voxels that had significant classification in the surrounding sphere of voxels. This mask was used to select voxels in a group level accuracy map created by averaging accuracy across all participants in each voxel.

Behavioral Results

Behavioral data were first averaged across the 2 repetitions for each participant, sound, and task. Data were discarded for trials on which participants did not answer. We computed the accuracy of the response by averaging correct answers across the 6 sounds for each of the 8 different types of pair. Reaction times were averaged across correct answers only.

The experiment had a full-factorial within-participant design, with the Tasks and the Types of pairs as the within-participant factors. The factor Task had 3 levels (Noise, Action, Material). The factor Types of pairs had 8 levels (Table 1).

Response accuracy and response times were submitted to a repeated-measure analysis of variance (ANOVA). Here and in the following, all P-values are reported after using the Geisser-Greenhouse correction for violation of sphericity when necessary. Planned contrasts used Pillai's test. The ANOVA showed a significant effect of the Task on the accuracy of the responses ($F(2,44) = 4.00$, $P < 0.05$), a significant effect of the Type of pair ($F(7,154) = 20.4$, $P < 0.01$), and a significant interaction between the 2 factors ($F(14,308) = 7.39$, $P < 0.01$).

Planned contrasts revealed that accuracy was best for the Action task (98.2%) compared with the Material task (96.8%, $P < 0.01$) but was not significantly lower for the Noise task (97.1%, $P = 0.07$). Accuracy was not significantly different between the Material and Noise tasks ($P = 0.54$). Contrasts also showed that accuracy was best for the pairs made of 2 identical sounds (SM–SM and TP–TP: 99.8%) and for the pairs in which both material and action were different (i.e., the sounds that were maximally different SP–TM and SM–TP: 99.0%). Accuracy was worst for the pairs in which only action (SM–TM and SP–TP: 93.3%) or material (TM–TP and SM–SP: 97.6%) was different.

Reaction time was analyzed in a similar ANOVA, which showed a significant effect of the Task on the reaction times

$(F(2,44) = 35.6, P < 0.01)$, a significant effect of the Type of pair $(F(7,154) = 69.3, P < 0.01)$, and a significant interaction between the 2 factors $(F(14,308) = 14.0, P < 0.01)$. Planned contrasts showed that participants responded significantly faster to the Noise task (515 ms) than to the Action task (676 ms, $P < 0.01$) and to the Material task (765 ms, $P < 0.01$), and that they responded significantly faster for the Action than the Material task. RTs ranged from 337 ms for the pair SM–SM to 954 ms for the pair SP–TP. Across tasks, RTs were faster on average for the pairs of identical sounds (SM–SM and TP–TP: 487 ms) and significantly slower (with an alpha-value of 0.01) for the other pairs (SM–TM and SP–TP: 760 ms; TM–TP and SM–SP: 727 ms, SP–TM and SM–SP: 727 ms).

Results: General Linear Model Analysis of fMRI Data (GLM)

For the fMRI BOLD response data, the main effect of both factors (attentional Task and Type of Pair) resulted in significant differences of activity in several cortical areas. Interaction between the 2 factors did not produce any significantly activated region. Next paragraphs therefore report the effects of each factor separately. Table 3 reports the stereotactic coordinates of the regions significantly activated in the different analyses.

Effects of the Different Types of Pairs

The 8 different Types of pairs possess distinct acoustic characteristics: tapping sounds are shorter and have a much sharper onset than the scraping sounds, and metal sounds have a longer decay and more energy in high frequencies than plastic sounds. We therefore expected that the different Types of pairs would modulate activity in early auditory areas. This was confirmed by the results of the GLM analysis. Table 3 shows that cortical activity elicited by the different Types of pairs peaked in both Heschl's gyri (HG) and spread to surrounding regions in STG and the thalamus. Significant activity in the thalamus may reflect earlier auditory processing and relaying. The nonsignificant interaction between the Types of pairs and the Task suggests that this activity in early perceptual regions was not modulated by attention.

Effects of the Attentional Tasks

Altogether, the different attentional tasks resulted in significant changes of the BOLD response in several cortical regions reported in Table 3. The main effect of the Tasks first appears in a large bilateral cluster encompassing bilateral HG, surrounding area in STG, and spreading to inferior frontal gyri and superior frontal gyri (b-aSFG), a bilateral cluster at the junction of the inferior precuneus and posterior cingulate gyri (b-iPC/pCG), a bilateral cluster located on the anterior portion of the cingulate gyri (b-aCG), and 4 other areas in left hemisphere, centered on posterior inferior frontal regions (l-IF), and the medial and posterior part of the left superior frontal gyrus (l-MedSFG), on the IPS (l-IPS), and straddling through the junction of the temporal and parietal lobes (l-TPj): posterior parts of STG and MTG, supramarginal, angular gyri, and anterior part of the superior occipital gyrus.

Planned orthogonal contrasts between the 3 tasks (A vs. M and N vs. (A + M)/2) detail this effect (see the top panel of Fig. 1). First, the contrast between the Noise and the 2 meaningful tasks (i.e., N vs. (A + M)/2) reveals that the Noise task elicited higher activation than the meaningful tasks in b-HG, b-PC/pCG, and l-TPj. The 2 meaningful tasks (A and M) elicited higher activation than the Noise task in the 3 main areas centered on l-IFS, l-IPS, b-MedSFG, as well as b-aCG. Next, a contrast between tasks A and M showed that the Material task resulted in higher activation than the Action task in an area located in the posterior part of the left inferior frontal sulcus (l-IFS), extending to the junction with the inferior part of the precentral sulcus, which is part of the frontoparietal network of action sound identification (see the middle panel of Fig. 1). The Material task also elicited higher activation in a bilateral area located in the inferior part of the occipital lobe (b-IO). This area overlaps with the most posterior parts of the inferior occipital, middle occipital, lingual, and fusiform gyri, and the cunei (i.e., visual areas). The Action task resulted in higher activation than the Material task in a cluster straddling the posterior parts of STG and MTG, and angular and supramarginal gyri (i.e., the junction of the temporal and parietal lobes l-TPj, part of the frontoparietal network of action representation), and a bilateral area in the anterior section of the superior and medial frontal gyri (b-aMedSFG).

Figure 2 represents the BOLD response averaged across participants, blocks, and the voxels of the 4 significant clusters in the A versus M contrast. This figure shows a striking difference between the different clusters. On the one hand, the BOLD response in l-IFS and b-IO clusters is clearly synchronized with the timing of the stimuli, and the amplitude of the response is greater for the Material task than for the Action and Noise tasks, as highlighted by the GLM analysis. On the other hand, the response in MedSFG and l-TPj clusters is also synchronized with the stimulus timing, but in antiphase with the response of the 2 former clusters. Whereas the peak of the BOLD response occurs 3–4 TRs (6–8 s) after the stimulus onset for the l-IFS and b-IO clusters (i.e., consistent with the canonical HRF), the response for the MedSFG and l-TPJ peaks earlier in the trial (i.e., 1–2 TRs after stimulus onset) and decreases between 3 and 5 s after stimulus onset (i.e., when it increases in the other clusters). This antiphase is reflected in the negative coefficients applied to the canonical HRF when modeling the BOLD response (see the IRFs on the right hand side of Fig. 2). Such a negative BOLD response (NBR) has been reported in several studies and interpreted as reflecting deactivation from the brain's default mode (Shmuel et al. 2002; Raichle 2015). In such a framework, the significant GLM would be interpreted as resulting from a greater deactivation in the Material task, rather than a greater activation in the Action task. Such an interpretation is consistent with the fact that l-TPj and medial SFG has considered as important components of the brain's default network (see the general discussion below).

In addition, correlation analyses across the 8 types of pairs and the 3 tasks (i.e., 24 data points) showed that the percentage of signal change (compared with baseline) in l-TPj and medial SFG was negatively correlated with the reaction times $(r(22) = 22120.68, P < 0.01$ for both clusters) across all participants. This further confirms that a decrease of activity in these regions was modulated by task demand. In comparison, percent change was positively correlated with reaction times in l-IFS $(r(22) = 0.59, P < 0.01)$, and not significantly correlated with reaction times in the bilateral IO clusters (left: $r(22) = 0.25$; right: $r(22) = 0.15$) (To further investigate the influence of the reaction times, we conducted a linear mixed-effect analysis of the data, with the 2 experiment factors as within-participant factors, the reaction times as a random covariate, and the participants as a random effect, using the method proposed by Chen et al. 2013

Table 3 Stereotactic coordinates of the clusters found in the GLM analysis of the imaging data. Clusters with the same shortcuts overlap

Effect	Brain region label	Coordinates (MNI)	Number of voxels	BA	Shortcut
Type of pair	Right Heschl's gyrus, STG	47, −19, 8	677	41, 22	b-HG
	Left Heschl's gyrus	−43 −17, 5	596	41, 22	
	Thalamus	8, −56, −4	161	−	
Main effect of attention condition	Bilateral Heschl's gyri, anterior superior and middle temporal gyri, anterior superior frontal and medial gyri	−43, −20, 8	5621	41, 22	b-HG, b-aS/MTG, b-aSFG, b-aSFG, b-MedFG
	Bilateral precuneus, posterior cingulate gyrus	−10, −52, 32	597	31	b-iPC/pCG
	Left angular, supramarginal, and superior occipital gyri	−43, −80, 27	497	19, 39	l-Ang/Supr/SO
	Bilateral anterior cingulate gyri	−1, −4, 23	234		b-aCG
	Left middle frontal gyrus, inferior frontal gyrus, precentral gyrus	−43, 13, 32	856	9, 44	l-IF
	Left supramarginal gyrus/intraparietal sulcus	−28, −49, 25	342		l-IPS
	Left posterior part of medial frontal gyrus	−4, 9, 55	143	6	l-pMedFG
	Right lingual gyrus	20, −94, −7	234	17	b-IO
	Left lingual gyrus	−16, −90, −10	226	17	
	Left fusiform gyrus	−46, −49, −13	80	37	l-fusi
N > (A + M)/2	Bilateral Heschl's gyrus, connected through inferior frontal gyrus, and extending to superior frontal and medial gyri	−43, −19, 8	5220	41, 22	b-HG, bSFG
	Bilateral Precuneus, posterior cingulate gyrus	−10, −52, 32	593	31	b-PC, bCG
	Left posterior parts of superior and middle temporal, angular and suprarmarginal gyri, as well as anterior parts of superior occipital gyrus	−37, −83, 24	401	39	l-TPj
(A + M)/2 > N	Left middle frontal gyrus, Inferior frontal gyrus, Precentral gyrus	−46, 22, 29	789	9, 46	l-IF
	Bilateral ant. cingulate gyrus	6, −2, 24	387	24	b-aG
	Left Supramarginal gyrus/Intraparietal sulcus	−28, −49, 35	320		l-IPS
	Left posterior part of medial frontal gyrus	−4, 9, 55	105	6	l-pMedFG
M > A	Bilateral inferior/middle/lingual/fusiform occipital gyri	14, −97, 4	635	17, 18	b-IO
	Left post. Inferior frontal sulcus	−46, 26, 23	113	44	l-IFS
A > M	Left posterior parts of superior and middle temporal, angular and suprarmarginal gyri. Temporoparietal junction. Also includes the post. part of the left middle temporal gyrus	−49, −65, 35	177	39	l-TPj
	Bilateral anterior portion of Superior/ Medial FG	14, 49, 43	81	9	b-aMedSFG
Instruction Change	Bilateral fusiform gyrus	29, −68, −16	125	19	b-FG

(The analysis is reported in Supplementary material.) Including the reaction times as a covariate had very little influence on the clusters discussed here: the size of the b-IO, l-IFS, and l-TPj clusters was reduced by about 5%, and the location of the center of gravity varied by no more than 5 mm. This rules out the possibility that the results were only driven by the significant differences of reaction times between the tasks.).

Effects of the Instruction Changes

The BOLD response of the IO clusters has another notable aspect: there is a peak at the beginning of the block, when no stimulus is presented, but the instructions displayed on the monitor change from one task to another. This suggests that the significant differences in activation found in the occipital regions between the 2 meaningful tasks may be partially accounted for by the differences between their visually presented instructions and fixation symbols. The second, separate GLM analysis therefore focused only on the changes of visual display at the beginning of each block (Instruction Changes) to disentangle potential visual effects from auditory processing effects.

Table 3 reports the results of the main effect of this GLM analysis with the Instruction Changes as the experimental factor (thus focusing on the initial 10 s of each block). These instruction changes resulted in significant differences of activity in the ventral part of the occipital lobe, including parts of
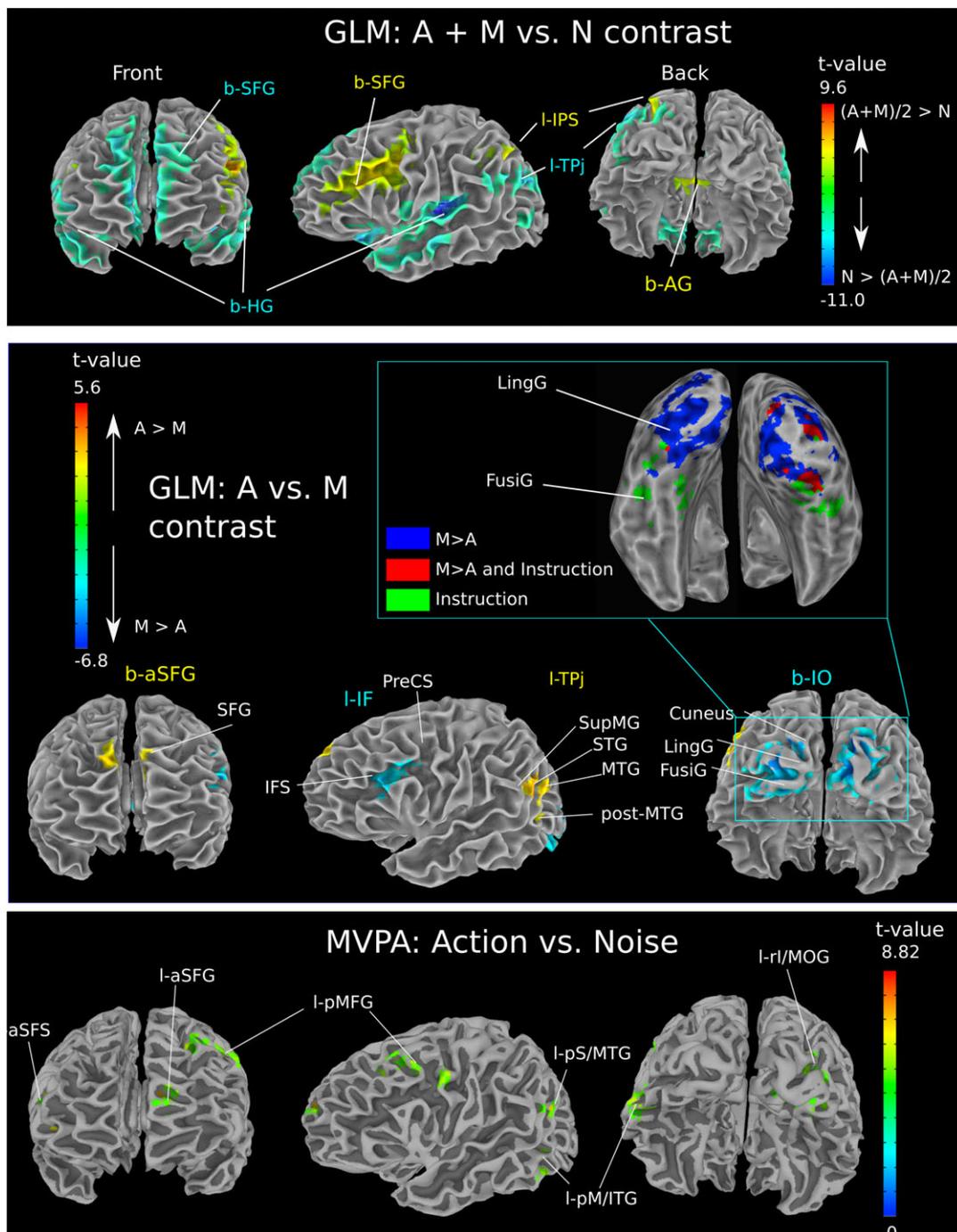
Figure 1. Top panel: Results of the A + M – 2 N contrast in the GLM analysis. Regions in yellow correspond to A + M > 2 N, regions in blue correspond to 2 N > A + M. Middle panel: results of the A – M contrast in the GLM analysis. Regions in yellow correspond to A > M. Regions in blue correspond to M > A. Bottom panel: results of the MVPA (A vs. M). SFG = superior frontal gyrus. IFS = inferior frontal sulcus. PreCS = precentral sulcus. SupMG = supramarginal gyrus. STG = superior temporal gyrus. MTG = middle temporal gyrus. LingG = lingual gyrus. FusiG = fusiform gyrus. b-aSFG, l-IF, l-TPj, and b-IO are the acronyms used in Table 3.

lingual and fusiform gyri (i.e., ventral visual pathway involved in the processing identity of visually presented objects). The insert in the middle panel of Figure 1 magnifies the overlap between the 2 analyses (M > A and main effect of Instruction Changes). This shows that there is actually very little overlap between the 2 effects and suggest that the greater activation found in occipital regions for the (auditory) material task genuinely reflects the effect of the auditory stimuli and not the differences of visual instructions between the tasks.

Results: MVPA of MRI data (MVPA)

A whole brain MVPA searchlight was conducted on the full dataset to find regions classifying Material versus Action tasks. When the resulting t-statistic maps were thresholded we revealed areas that had spatial patterns of activity supporting the classification of the 2 tasks ($q < 0.05$). A total of 9 clusters of voxels supporting classification were identified with more than 10 significant voxels (see Table 4 and the bottom of Fig. 1).
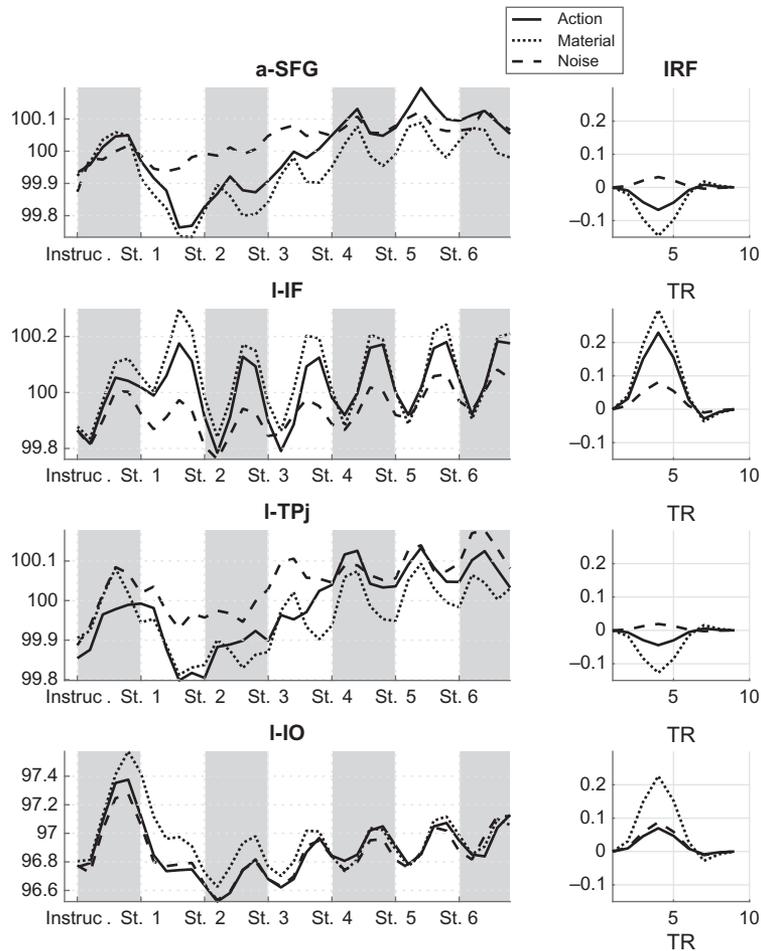
Figure 2. Left panel: BOLD response (in percent change) averaged across participants, blocks, and the voxels of the 4 clusters found in the A versus M contrast. The first 5 TRs are silent and correspond to the moment where the instructions are changed. The next TRs correspond to the 6 trials of each block. The right panel represents the average Impulse Response Function found in the GLM analysis for the stimulus trials in each task.

Reported coordinates correspond to the peak of the clusters in Talairach space.

The largest significant cluster found by the MVPA is located in the inferior part of right occipital lobe, straddling middle occipital, inferior occipital, lingual and fusiform gyri (r-I/MOG). Another large area comprises 2 clusters in the left frontal lobe, including posterior portions of the middle frontal gyrus and the middle section of the precentral gyrus (l-pMFG). A smaller area in the right hemisphere includes 2 smaller clusters in the middle and inferior frontal sulci (r-aMFS and r-IFS). There are also 2 clusters in the left temporal lobe, straddling l-TPj and middle and inferior temporal gyri (l-pM/ITG). Other significant clusters are located in the right hemisphere, in the posterior part of superior temporal sulcus (r-pSTS) and the anterior part of the superior frontal gyrus (r-aSFG).

Figure 3 overlays the results of the MVPA, 2 orthogonal contrasts in the GLM analysis on attentional tasks ((A + M)/2 > N and A vs. M), and the overlap of the 2 analyses with a color code. The description and location of the overlapping areas are also reported in Table 4. The A vs. M contrast and the MVPA overlap in 3 areas. One such area is located at the l-TPj, where the GLM found more deactivation for the Material than the Action tasks. Another area is located in the inferior part of the right occipital lobe, on the lingual gyrus (r-LG), and another one on the right inferior occipital gyrus (r-IOG). Finally, the clusters

found by the 2 analyses on medial part of the superior frontal gyrus (aSFG) are also very close (more deactivation for the Material task). The conjunction of the (A + M)/2 > N contrast and the MVPA highlights a region located on the inferior part of the left precentral gyrus and the posterior middle frontal gyrus (l-PCG). Interestingly, this region is very close to a region found by the GLM analysis where the Material task results in higher activations than the Action task (l-IFS). The conjunction also reveals a much smaller area in the right inferior frontal sulcus.

Overall, the comparison of the MVPA and GLM analyses confirm that the 4 clusters previously identified (temporoparietal junction, left inferior frontal, medial superior frontal gyrus, and occipital regions) play an important role when participants are attending to the material and the action producing the sounds.

Again, the classification of 2 tasks in inferior occipital regions could result from activity in visual areas processing shape and color elicited by the differences of visual instructions between the 2 tasks. To disentangle these effects, the bottom right part of Figure 3 represents the overlap of the results of the MVPA and the GLM analyses of the Instruction Changes. There is clear overlap of the 2 analyses in the right lingual and fusiform gyri, indicating that the successful classification of the 2 tasks based on patterns of activity in these regions resulted from the different visual displays. However, there is no overlap between the 2 analyses for a cluster found on the right inferior

Table 4. Stereotactic coordinates of the clusters found in the MVPA analysis, and the conjunction of the GLM contrasts (A vs. M) and the MVPA.

Effect	Brain region label	Coordinates (MNI)	Number of voxels	BA	Shortcut
MVPA: action versus material	Right inferior/middle occipital gyrus, lingual and fusiform gyri	29, −81, −13	212	17, 18	r-I/MOG
	Left posterior part of middle frontal gyrus straddling to middle section of precentral gyrus	−34, 15, 49	65	6	l-pMFG
	Left posterior part of middle frontal gyrus	−40, 12, 52	56	6	l-pMFG
	Right anterior part of inferior frontal sulcus/gyrus	38, 45, 11	56	46	r-aIFS
	Right anterior part of middle frontal sulcus	38, 31, 33	22	9, 44	r-aMFS
	Left anterior part of superior frontal gyrus	−9, 62, 22	56	10	l-aSFG
	Left posterior part of inferior and middle temporal gyri	−56, −54, −7	40	37	l-pM/ITG
	Left posterior parts of superior and middle temporal gyri	−47, −67, 21	29	39	l-pS/MTG
	Right posterior part of superior temporal sulcus	51, −57, 9	36	39	r-pSTS
Conjunction (M + A)/2 > N and MVPA	Left precentral gyrus and posterior part of middle frontal gyrus	−46, 0, 38	16	7	l-PreCG
	Right anterior inferior frontal sulcus	38, 29, 30	4	9, 44	r-aIFS
Conjunction: (M > A or A > M) and MVPA	Right lingual gyrus, inferior occipital gyrus	32, −81, −20	39	18, 19	r-LG
	Right middle occipital gyrus	29, −91, 4	4	18, 19	r-MOG
	Left posterior parts of superior and middle temporal gyri, temporoparietal junction	−49, −64, 18	22	39	l-TPj

For the conjunction, stereotactic coordinates correspond to the center of mass of the overlapping areas.
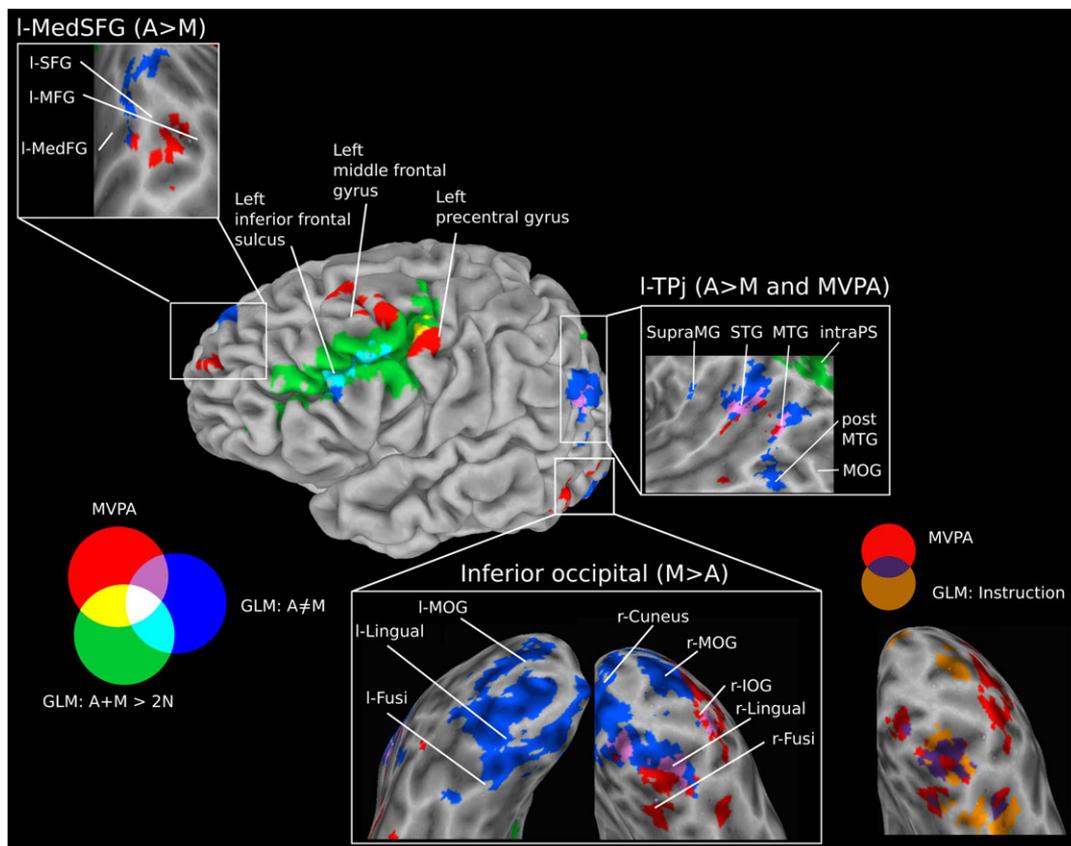
Figure 3. Conjunction of the results of the GLM analysis and the MVPA. Cluster names are described in Table 4. The overlap between the 2 analyses is represented by different shades.

occipital gyrus. This suggests that more superior regions in the right occipital lobe were actually responding differently to auditory information in the 2 tasks.

General Discussion

Participants made only a few errors across all tasks and types of pairs, confirming that they were actively focusing on the task, and that the sounds were easily discriminable. In more detail, the slightly longer reaction times and poorer accuracy within the Material task suggests that discriminating the materials involved deeper processing of the stimuli than discriminating the actions. Another possibility for the longer reaction times is that the acoustic cues specifying the material are usually located at the offset of the sounds (spectrotemporal decay) whereas the acoustic cues specifying are usually found at the onset of the sounds. The different Types of Pairs also influenced accuracy in the Material task (the task was easier when the sounds were identical or maximally different), but not in the Action task. This further suggests that the Material Task was slightly more challenging for the listeners. This was echoed by the comments of several participants.

The results of the brain imaging analyses showed that attending to material activated a network of areas involved in accessing semantic, motor, and visual representations, and probably caused a deactivation of the brain's default network. In contrast, attending to action did not increase activity in any specific cortical area. The following paragraphs discuss these observations.

Task Effect in the Action-Sound Network

The first and somewhat expected result is that the meaningful tasks overall (A and M) elicited more activation than the Noise task in the left frontoparietal action-sound network (centered on the left-IPS (l-IPS) and encompassing a large part of the left Inferior Frontal lobe l-IF, see Lewis 2006). Importantly, the Action and Material Tasks activated most of this network equivalently. We therefore suggest that these regions are for the most part not under the control of focused auditory attention and may identify both basic properties of sound objects even when listeners focus on only one property. In particular, the left intraparietal sulcus (i.e., dorsal pathway, integration of sensory and motor representations) was activated equivalently by both meaningful tasks (whereas one of our prior expectation was that identifying action would activate more strongly this area). This is consistent with an interpretation of the dorsal pathway as processing auditory stimuli to guide actions (in particular for visual orientation, Arnott and Alain 2011): the action task required the listeners to identify actions, not to perform an action.

Contrast analyses offered more insights into the effect of the meaningful tasks. It revealed significant differences in activation between the Action and Material tasks in 4 cortical areas: the medial aspect of the Superior Frontal Gyrus (MedSFG), the left temporoparietal junction (l-TPj: posterior part of the left superior and middle temporal gyri, IPL), the left inferior frontal sulcus (l-IFS), and the bilateral inferior occipital lobes (b-IO).

Material Task Targets the Default Brain Mode

Analysis of the time course of the BOLD signals showed that the significant differences found in MedSFG and l-TPj actually resulted from negative BOLD responses: responses were in antiphase with stimulus timing and with the BOLD response in other brain areas. One interpretation is that the phase differences were simply caused by small differences in the timing of the BOLD response in different parts of the brain. However, the near-perfect antiphase we observed is striking and suggests instead that focusing attention to the different properties of the sound sources actually caused deactivation of these regions (i.e., decrease of activity from baseline). Decreased cerebral blood flow and negative bold responses have been previously reported in several PET and fMRI studies (Shulman et al. 1997; Raichle et al. 2001; Shmuel et al. 2002). Deactivation is typically found in medial prefrontal regions, posterior cingulates, and the temporoparietal junction (BA 39), and occurs in variety of nonself-referential goal-directed tasks when compared with a resting state baseline. The location of the 2 regions highlighted by our GLM analysis (medial SFG and temporoparietal junction) is therefore consistent with these regions, identified in the literature as the "default brain mode." It has been shown to correspond to neural inhibition (Wade 2002; Shmuel et al. 2002), and these results have been interpreted in some cases as cross-modal inhibition of sensory areas (Laurienti 2004; Amedi et al. 2005), or as the existence of an organized, baseline default mode of brain function that is suspended during specific goal-directed behaviors (Raichle 2015). In this context, we interpret the significant contrast (Action > Material) as resulting from a greater deactivation in the Material task, rather than from a greater activation in the Action task. This is consistent with the idea that the Material task was more demanding than the Action task as it required a larger inhibition of the brain default mode; this was confirmed by the significant correlation between the amount of deactivation and the reaction times. However, this interpretation is speculative because the experimental design did not include resting state periods specifically designed to assess baseline activity. It thus difficult to exclude the possibility of artifacts such as those caused by global signal normalization (Laurienti 2004).

In addition, the results of the MVPA showed that it is possible to predict the M and A tasks from the patterns of activity in l-SFG and l-TPJ, in areas close or overlapping the regions highlighted by the GLM analysis. This confirms that these regions are consistently responding to the 2 meaningful tasks in 2 different ways and again suggests that their activity is sensitive to the focus of attention and to task demand.

Discriminating Material Recruits Semantic and Motor Association Areas

In contrast, the significant differences in l-IFS were caused by a greater activation elicited by the Material task, with activity positively correlated with the task demand (as indexed by the differences in reaction times, despite very high overall accuracy). Both meaningful tasks resulted in higher activation than the Noise task in a large region centered on l-IFS. This contrast thus shows that the center of this region (the sulcus proper) is more activated by the Material task and higher task demand. The l-IFS corresponds to the projection of the ventral pathway of auditory processing (object identification), also belongs to the frontoparietal action-sound network, and elaborates a primary mapping between prerepresentational stages of auditory analyses and semantic representations (Lewis 2006). Whereas identifying both properties of sound sources involves access to semantic representations, the higher activation of IFS therefore suggest that identifying material relied more heavily on these representations.

Because of the correlations between reaction times and activation, we explored whether the significant differences in activation between material and action tasks were driven by the reaction

time differences. However, adding the reaction times as a random covariate in the analysis had very little influence on the activation of this region (the analysis is reported in Supplementary material). The neural differences seen between tasks are not explained by difficulty (as indexed by reaction times), which allows us to conclude that they reflect different cognitive processes involved in performing the 2 tasks.

The MVPA identified a region similarly located in the posterior part of the left posterior frontal lobe, overlapping the results of the GLM analysis (A + M > 2 N). More precisely, the 2 analyses overlapped on the ventral precentral gyrus, a region corresponding to the VPMC. VPMC receives projections of both ventral and dorsal pathways and is posited to be the human equivalent to audiovisual mirror neurons (Aglioti and Pazzaglia 2010), mapping sensory inputs with motor programs. Altogether, these results suggest a functional organization of the left inferior frontal region involved in sound source identification, with the inferior frontal sulcus mapping sensory inputs to semantic representations and being sensitive to task demand (activity in VMPC was correlated with reaction times), and VPMC (specifically mapping sensory inputs to motor representations) under the control of the focus of attention (both analyses concluded that focusing attention modulated activity in this area).

Discriminating Material Recruits Bilateral Visual Areas

The results of both analyses also converged in occipital regions. The Material task elicited significantly more (positive) activation than the Action task in bilateral occipital region, but this activation was not correlated with task demand. Activity driven by auditory stimuli in occipital regions (usually considered as visual areas) is more surprising (though not unique) and should therefore be considered with care. The MVPA also successfully predicted the meaningful tasks based on the patterns of activity in occipital regions. Comparisons with the different GLM analyses (effects of visual display and effects of experiment factors) revealed that some of these regions located on lingual and fusiform gyri were actually reflecting the changes of visual displays between the 2 tasks (color and shape of the fixation symbols), consistent with the role of the ventral visual pathway. More interestingly, these comparisons also singled out a region located on the right inferior and middle occipital gyri whose patterns of activity could successfully classify the 2 meaningful tasks and was not driven by the changes of visual instructions. This suggests that focusing auditory attention on the material of the objects causing the sounds actually recruited visual areas previously found to process material identification (Cant and Goodale 2007). The result of the MVPA is therefore consistent with the findings of Vetter et al. (2014) who showed that the classification of sound categories can be successfully predicted from the cortical activity in V1, V2, and V3 (see also Lehet et al. submitted). One interpretation is that identifying material required access to visual representations or memories. Such an idea is consistent with the results of Merabet et al. (2004), who hypothesized that the visual cortex is well suited for precise spatial judgments, even for nonvisual stimuli.

Discriminating Action Did not Uniquely Activate the Human Motion Complex (p-STS/MTG)

Finally, it is important to note that neither analysis found any significant results in some regions where we expected to see them. For instance, changing the focus of attention was not reflected in primary auditory regions (areas surrounding STG),

which would have been the case if focusing attention toward material or action had modulated the processing of the corresponding basic acoustic properties. In addition, we expected that focusing attention to the actions that caused the sounds would elicit activity in regions close to the p-STS/MTG complex, which has been shown to be sensitive to human motion in both the visual and auditory modalities. A detailed analysis of the A > M contrast shows that the l-TPj cluster actually includes the junction of the posterior MTG and the anterior middle occipital gyrus, and thus corresponds pMTG. This area consists of a number of voxels that would have not survived thresholding if they had not been connected to the TPj cluster. One possible interpretation is that our stimuli evoked human motion only moderately and thus that the statistical power of our whole-brain analyses was not sufficient to reveal small changes of activity in this area.

Conclusion

In conclusion, the results of this study are in line with our previous findings (Lemaitre and Heller 2012, 2013) that suggest that auditory perception is better suited to comprehend the actions creating sounds in the listeners' environment than the properties of the objects radiating these sounds. These results are consistent with embodied cognition theories (Barsalou 2008) and ideomotor theories (Shin et al. 2010) specifying that sensory stimuli are stored together with the motor representations of the actions causing them. However, this research did not specifically test between them. The results indicate that recovering action is an inherent part of listening to sounds, but it does not suggest that a motor simulation is the exclusive means by which this is done. More importantly, these results elucidate the neural bases of this behavior: The most striking result is that shifting the listeners' attention toward the material resulted in greater activation in a network of areas involved in accessing semantic and visual representations, and probably caused a greater deactivation of the brain's default network (when compared with shifting their attention to the action). Together with the behavioral results, this provides evidence that identifying the action causing the sounds is a default way of listening to sounds, whereas identifying the material requires a deeper processing, access to long-term knowledge, and access to visual properties of the objects potentially causing the sounds. Future work using other types of sounds will be useful to confirm whether the cortical networks found in other studies on sound identification depend on the properties elicited by the experimental task.

Supplementary Material

Supplementary material are available at *Cerebral Cortex* online.

Funding

Notes

References

Adams RB, Janata P. 2002. A comparison of neural circuits underlying auditory and visual object categorization. Neuroimage. 16:361–377.

Aglioti SM, Pazzaglia M. 2010. Representing actions through their sound. Exp Brain Res. 206 (2):141–151.

Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW. 2006. Task-modulated "what" and "where" pathways in human auditory cortex. Proc Natl Acad Sci U S A. 103 (39):14608–14613.

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. 2001. "What" and "where" in the human auditory system. Proc Natl Acad Sci U S A. 98 (21):12301–12306.

Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A. 2007. Shape conveyed by visual-to-auditory substitution activates the lateral occipital complex. Nat Neurosci. 10:687–689.

Amedi A, von Kriegstein K, Atteveldt NM, Beauchamp MS, Naumer MJ. 2005. Functional imaging of human crossmodal identification and object recognition. Exp Brain Res. 166:559–571.

Arnott SR, Binns MA, Grady CL, Alain C. 2004. Assessing the auditory dual-pathway model in humans. Neuroimage. 22 (1):401–408.

Arnott SR, Cant JS, Dutton GN, Goodale MA. 2008. Crinkling and crumpling: an auditory fMRI study of material properties. Neuroimage. 8:368–378.

Arnott SR, Alain C. 2011. The auditory dorsal pathway: orienting vision. Neurosci Biobehav Rev. 35 (10):2162–2173.

Barsalou LW. 2008. Grounded cognition. Annu Rev Psychol. 59: 617–645.

Beauchamp MS, Argall D, Borduka J, Duyn JH, Martin A. 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 7 (11): 1190–1992.

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron. 34:149–159.

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2003. fMRI responses to video and point-light displays of moving humans and manipulable objects. J Cogn Neurosci. 15 (7):991–1001.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 57 (1):289–300.

Bidelman GM, Moreno S, Alain C. 2013. Tracing the emergence of categorical speech perception in the human auditory system. Neuroimage. 79:201–212.

Bidet-Caulet A, Voisin J, Bertrand O, Fonlupt P. 2005. Listening to a walking human activates the temporal biological motion area. Neuroimage. 28:132–139.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and non-speech sounds. Cereb Cortex. 10:512–528.

Borst G, Thompson WL, Kosslyn SM. 2011. Understanding the dorsal and ventral systems of the human cerebral cortex. Am Psychol. 66 (7):624–632.

Brainard DH. 1997. The psychophysics toolbox. Spat Vis. 10: 433–436.

Cant JS, Goodale MA. 2007. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. Cereb Cortex. 17 (3):713–731.

Caporello Bluvas E, Gentner TQ. 2013. Attention to natural auditory signals. Hear Res. 305:10–18.

Carello C, Wagman JB, Turvey MT. 2005. Acoustic specification of object property. In: Anderson JD, Fisher Anderson B, editors.. Moving Image theory: ecological considerations. Carbondale, IL: Southern Illinois University Press. p. 79–104.

Chang CC, Lin CJ. 2011. LIBSVM: a library for support vector machines. ACM Trans Intel Syst Tech. 2 (3):27.

Chen G, Saad ZS, Britton JC, Pine DS, Cow RW. 2013. Linear mixed-effects modeling approach to fMRI group analysis. Neuroimage. 73:176–190.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 3: 201–215.

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage. 97:271–283.

Doehrmann O, Naumer MJ, Wolz S, Kaiser J, Altmann CF. 2008. Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia. 46:2776–2786.

Engl LR, Frum C, Puce A, Walker NA, Lewis JW. 2009. Different categories of living and non-living sound-sources activate different cortical networks. Neuroimage. 47:1778–1791.

Falkenberg LE, Specht K, Westerhausen R. 2011. Attention and cognitive control networks assessed in a dichotic listening study. Brain Cogn. 76:276–280.

Fritz JB, Elhilali M, David SV, Shamma SA. 2007. Auditory attention. Focusing the searchlight of sound. Curr Opin Neurobiol. 17 (4):437–455.

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

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P. 2012. Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex. 23 (9):2025–2037.

Giordano BL, McDonnell J, McAdams S. 2010. Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds. Brain Cogn. 73:7–19.

Goll JC, Kim LG, Hailstone JC, Lehmann M, Buckley A, Crutch SJ, Warren JD. 2011. Auditory object cognition in dementia. Neuropsychologia. 49:2755–2765.

Goodale MA, Milner D. 1992. Separate visual pathways for perception and action. Trends Neurosci. 15 (1):20–25.

Greenberg AS, Esterman M, Wilson D, Serences JT, Yantis S. 2010. Control of spatial and feature-based attention in frontoparietal cortex. J Neurosci. 30 (43):14330–14339.

Grossman ED, Battelli L, Pascual-Leone A. 2005. Repetitive TMS over posterior STS disrupts perception of biological motion. Vision Res. 45 (22):2847–2853.

Houix O, Lemaitre G, Misdariis N, Susini P, Urdapilleta I. 2012. A lexical analysis of environmental sound categories. J Exp Psychol Appl. 18 (1):52–80.

James TW, Stevenson RA, Kim S, VanDerKlok RM, James KH. 2011. Shape from sound: evidence for a shape operator in the lateral occipital cortex. Neuropsychologia. 49:1807–1815.

Kaplan JT, Iacoboni M. 2007. Multimodal action representation in human left ventral premotor cortex. Cogn Process. 8: 103–113.

Keysers C, Kohler E, Ulmita MA, Nanetti L, Fogassi L, Gallese V. 2003. Audiovisual neurons and action recognition. Exp Brain Res. 153:628–636.

Kiefer M, Trumpp N, Herrnberger B, Sim EJ, Hoenig K, Pulvermuller F. 2012. Dissociating the representation of action- and sound-related concepts in middle temporal cortex. Brain Lang. 122:120–125.

Klatzky RL, Pai DK, Krotkov EP. 2000. Perception of material from contact sounds. Presence. 9 (4):399–410.

Kunkler-Peck AJ, Turvey MT. 2000. Hearing shape. J Exp Psychol Hum Percept Perform. 26 (1):279–294.

Lakatos S, McAdams S, Caussé R. 1997. The representation of auditory source characteristics: simple geometric forms. Percept Psychophys. 59 (8):1180–1190.

Laurienti PJ. 2004. Deactivations, global signal, the default mode of brain function. J Cogn Neurosci. 16 (9):1481–1483.

Lehet M, Navolio N, Pyles JA, Lemaitre G, Heller LM. A neural bias for processing environmental sounds based on action-related information rather than concurrent material information. Manuscript submitted for publication.

Lemaitre G, Heller LM. 2012. Auditory perception of material is fragile, while action is strikingly robust. J Acoust Soc Am. 131 (2):1337–1348.

Lemaitre G, Houix O, Misdariis N, Susini P. 2010. Listener expertise and sound identification influence the categorization of environmental sounds. J Exp Psychol Appl. 16 (1):16–32.

Lemaitre G, Heller LM. 2013. Evidence for a basic level in a taxonomy of everyday action sounds. Exp Brain Res. 226 (2): 253–264.

Lewis JW. 2006. Cortical networks related to human use of tools. Neuroscientist. 12 (3):211–231.

Lewis JW. 2010. Audio-visual perception of everyday natural objects—hemo-dynamic studies in humans. In: Naumer MJ, Kaiser J, editors.. Multisensory object perception in the primate brain. Berlin: Springer. p. 155–190.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. 2005. Distinct cortical pathways for processing tools versus animal sounds. J Neurosci. 25 (21):5148–5158.

Lewis JW, Talkington WJ, Puce A, Engel LR, Frum C. 2011. Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. J Cogn Neurosci. 23 (8):2079–2101.

Lewis JW, Talkington WJ, Tallaksen KC, Frum CA. 2012. Auditory object salience: human cortical processing of non-biological action sounds and their acoustical signal attributes. Front Syst Neurosci. 6. Article 27, 15 pages.

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. 2004. Human brain regions involved in recognizing environmental sounds. Cereb Cortex. 14:1008–1021.

Lipschutz B, Kolinsky R, Damhaut P, Wikler D, Goldman S. 2002. Attention-dependent changes of activation and connectivity in dichotic listening. Neuroimage. 17:643–656.

McAdams S, Roussarie V, Chaigne A, Giordano BL. 2010. The psychomechanics of simulated sound sources: material properties of impacted thin plates. J Acoust Soci Am. 128: 1401–1413.

McIntosh RD, Schenk T. 2009. Two visual streams for perception and action: current trends. Neuropsychologia. 47 (6):1391–1396.

Merabet L, Thut G, Murray B, Rews J, Hsiao S, Pascual-Leone A. 2004. Feeling by sight or seeing by touch? Neuron. 42 (1): 173–179.

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 485 (7397):233–236.

Nestor A, Plaut DC, Behrmann M. 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proc Natl Acad Sci. 108 (24):9998–10003.

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10 (9):424–430.

Paltoglou AE, Sumner CH, Hall DA. 2011. Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. Eur J Neurosci. 33:1733–1741.

Petersen SE, Posner MI. 2012. The attention system of the human brain: 20 years after. Annu Rev Neurosci. 35:73.

Pugh KR, Shaywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler DP, Katz L, Constables RT, Fletcher J, Lacadie C, Marchione K, Gore JC. 1996. Auditory selective attention: an fMRI investigation. Neuroimage. 4:159–173.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. Proc Natl Acad of Sci. 98 (2):676–682.

Raichle ME. 2015. The brain's default mode network. Annu Rev Neurosci. 38:433–447.

Shin YK, Proctor RW, Capaldi EJ. 2010. A review of contemporary ideomotor theory. Psychol Bull. 136 (6):943.

Shmuel A, Yacoub E, Pfeuffer J, Van de Moortele PF, Adriany G, Hu X, Ugurbil K. 2002. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron. 36 (6):1195–1210.

Shomstein S, Yantis S. 2006. Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. J Neurosci. 26 (21): 435–439.

Shulman GL, Fiez JA, Corbetta M, Buckner RL, Miezin FM, Raichle ME, Petersen SE. 1997. Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. J Cogn Neurosci. 9 (5):648–663.

Vetter P, Smith FW, Muckli L. 2014. Decoding sound and imagery content in early visual cortex. Current Biol. 24 (11): 1256–1262.

Wade AR. 2002. The negative BOLD signal unmasked. Neuron. 36 (6):993–995.

Zatorre RJ, Bouffard M, Belin P. 2004. Sensitivity to auditory object features in human temporal neocortex. J Neurosci. 24 (14):3637–3642.

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory-motor interactions in music perception and cognition. Nat Rev Neurosci. 8:547–558.