

# Functional Preference for Object Sounds and Voices in the Brain of Early Blind and Sighted Individuals

Giulia Dormal<sup>1,2</sup>, Maxime Pelland<sup>1</sup>, Mohamed Rezk<sup>2</sup>, Esther Yakobov<sup>4</sup>, Franco Lepore<sup>1</sup>, and Olivier Collignon<sup>1,2,3</sup>

## Abstract

■ Sounds activate occipital regions in early blind individuals. However, how different sound categories map onto specific regions of the occipital cortex remains a matter of debate. We used fMRI to characterize brain responses of early blind and sighted individuals to familiar object sounds, human voices, and their respective low-level control sounds. In addition, sighted participants were tested while viewing pictures of faces, objects, and phase-scrambled control pictures. In both early blind and sighted, a double dissociation was evidenced in bilateral auditory cortices between responses to voices and object sounds: Voices elicited categorical responses in bilateral superior temporal sulci, whereas object sounds elicited categorical responses along the lateral fissure bilaterally, including the primary auditory cortex and planum temporale. Outside the auditory regions, object sounds also elicited categorical responses in the left lateral and in the ventral occipitotemporal regions in both groups. These regions also showed response preference for images of objects

in the sighted group, thus suggesting a functional specialization that is independent of sensory input and visual experience. Between-group comparisons revealed that, only in the blind group, categorical responses to object sounds extended more posteriorly into the occipital cortex. Functional connectivity analyses evidenced a selective increase in the functional coupling between these reorganized regions and regions of the ventral occipitotemporal cortex in the blind group. In contrast, vocal sounds did not elicit preferential responses in the occipital cortex in either group. Nevertheless, enhanced voice-selective connectivity between the left temporal voice area and the right fusiform gyrus were found in the blind group. Altogether, these findings suggest that, in the absence of developmental vision, separate auditory categories are not equipotent in driving selective auditory recruitment of occipitotemporal regions and highlight the presence of domain-selective constraints on the expression of cross-modal plasticity. ■

## INTRODUCTION

In the visual and auditory areas of the human brain, separate brain clusters show category preferences. For instance, regions in the lateral part of the fusiform gyrus (fusiform face area [FFA]) respond more to faces than to nonface objects (Rossion, Hanseeuw, & Dricot, 2012; Kanwisher, McDermott, & Chun, 1997), whereas nonface objects elicit larger responses in parahippocampal gyri and in the lateroventral aspect of the occipitotemporal cortex (Andrews & Schluppeck, 2004). Although less frequently investigated than in vision, similar categorical preferences have been evidenced for sound processing in temporal cortex in the auditory cortices. It was found that listening to different categories of sounds such as humans voices (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) and artifacts (Lewis, Talkington, Tallaksen, & Frum, 2012; Lewis, Talkington,

Puce, Engel, & Frum, 2011; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005) activates distinct regions of the auditory temporal cortices (for a review, see Brefczynski-Lewis & Lewis, in press). Importantly, a few studies that have directly compared brain responses elicited by these sound categories while also taking into account their low-level characteristics suggest that categorical responses in temporal cortex partially abstract away from differences in basic sensory properties (Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2013; Leaver & Rauschecker, 2010).

Research on how sensory experience shapes these categorical preferences has recently received considerable attention. In people who lack visual experience because of early blindness, auditory and tactile stimulations massively activate the occipital cortex (e.g., Weeks et al., 2000; Sadato et al., 1996). Importantly, this reorganized occipital cortex is thought to follow a division of computational labor that is similar to the one observed in the sighted group (for reviews, see Reich, Maidenbaum, & Amedi, 2012; Voss & Zatorre, 2012; Dormal & Collignon, 2011; Ricciardi & Pietrini, 2011; Collignon, Voss, Lassonde, & Lepore, 2009). For instance, dorsal occipitoparietal

<sup>1</sup>University of Montreal, <sup>2</sup>University of Louvain, <sup>3</sup>University of Trento, <sup>4</sup>McGill University, Montreal, Canada

regions support spatial localization and motion processing in early blind participants (Dormal, Rezk, Yakobov, Lepore, & Collignon, 2016; Collignon et al., 2011; Ricciardi et al., 2007; for reviews, see Dormal, Lepore, & Collignon, 2012; Collignon et al., 2009), whereas occipitotemporal regions respond during tasks that require identification of a nonvisual input such as speech comprehension and semantic processing (Bedny, Pascual-Leone, Dodell-Feder, Fedorenko, & Saxe, 2011; Noppeney, Friston, & Price, 2003; Röder, Stock, Bien, Neville, & Rösler, 2002), Braille reading (Reich, Szwed, Cohen, & Amedi, 2011; Büchel, Price, & Friston, 1998), or the discrimination of shape attributes of objects based on tactile (Amedi, Raz, Azulay, & Malach, 2010; Amedi et al., 2007; Pietrini et al., 2004), auditory (Amedi et al., 2007), or verbal material (He et al., 2013; Peelen et al., 2013). Results from these studies suggest that categorical organization of the ventral occipitotemporal cortex (VOTC) may develop independently of visual experience. Importantly, some authors have also observed similar domain preference in the VOTC of sighted individuals during tasks that require processing of nonvisual material, therefore suggesting that those regions may at least partially abstract from vision (Bi, Wang, & Caramazza, 2016; Heimler, Striem-Amit, & Amedi, 2015; Wang et al., 2015; Ricciardi, Handjaras, & Pietrini, 2014; Reich et al., 2012). In contrast, other studies have failed to show preferential responses to specific auditory categories in the VOTC of either blind (He et al., 2013; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009) or sighted participants (Adam & Noppeney, 2010; Engel, Frum, Puce, Walker, & Lewis, 2009; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Lewis et al., 2005; Tranel, Grabowski, Lyon, & Damasio, 2005).

To date, categorical responses to sounds of objects *per se* (i.e., with nonverbal material and contrasting sounds of objects to another sound category) have only been investigated in sighted participants. These few studies failed to identify a clear categorical selectivity in the VOTC (Adam & Noppeney, 2010; Engel et al., 2009; Doehrmann et al., 2008; Lewis et al., 2005; Tranel et al., 2005). Numerous studies demonstrated that nonvisual information elicits radically distinct patterns of responses in the occipital cortex of individuals without visual experience (Collignon et al., 2013; Laurienti et al., 2002; Sadato, Okada, Honda, & Yonekura, 2002) and that unique patterns of functional specialization (Dormal et al., 2016; Collignon et al., 2011; Bedny, Konkle, Pelphrey, Saxe, & Pascual-Leone, 2010; Weeks et al., 2000) and connectivity (Dormal et al., 2016) exist in early blind people. In light of these findings, specific categorical responses to sounds of objects *per se* could be expected in early blind individuals as a result of cross-modal plasticity (Bavelier & Neville, 2002).

In sighted individuals, person (Mathias & von Kriegstein, 2014), emotion (Collignon et al., 2008), and speech recognition (van Wassenhove, 2013) are part of cognitive operations that benefit from the ability to efficiently bind facial

and vocal stimuli (Yovel & Belin, 2013). It has recently been suggested that face–voice interactions might rely on direct functional and structural links between face-selective regions in the visual cortex (e.g., FFA) and voice-selective regions in the auditory cortex (e.g., temporal voice area [TVA]; Blank, Anwander, & von Kriegstein, 2011). Preferential responses to voices in face-selective regions, and vice versa, have however never been demonstrated in sighted individuals. In blind individuals, the ability to extract crucial social information, such as the speaker's identity, emotional state, and speech, relies almost uniquely on voice perception. It could therefore be hypothesized that regions typically responsive to faces in sighted individuals would display a preferential response to voices in blind individuals due to the enhancement/unmasking of preexisting connections between these two cortical systems (Blank et al., 2011). Recent findings in congenitally deaf individuals support this hypothesis by demonstrating the presence of cross-modal face-selective responses in this population within regions of the temporal cortex that are typically tuned to voices in hearing participants (Benetti et al., 2017).

In the present study, we used fMRI to characterize brain responses to object sounds, voices, and the scrambled version of these stimuli in early blind and sighted individuals. We relied upon a factorial design that allowed us to directly contrast brain responses to familiar object sounds and voices. The voice and object stimuli were controlled for global energy (Belin et al., 2000, 2002), but not for other low-level auditory cues that differ between voices and objects, such as the spectral content of the sounds (Belin et al., 2000, 2002). Control for these low-level cues is provided by the scrambling (see Methods) of the two categories of stimuli. Therefore, by relying on directional contrasts between sound categories and between each category and its scrambled control condition, we assessed categorical preference while controlling for differences in the spectral content of the sounds and their global energy (see Rossion et al., 2012; Andrews, Clarke, Pell, & Hartley, 2010, for similar methodology in the visual literature). Sighted individuals were also tested in a visual experiment involving pictures of faces, objects, and scrambled pictures to assess the spatial correspondence between putative VOTC responses to auditory stimuli, on the one hand, and categorical responses elicited by visual material, on the other hand.

Our goals were threefold: (1) to test the existence of a double dissociation between voice- and object-selective regions in temporal auditory cortices while controlling for differences in low-level properties of these sounds (Giordano et al., 2013; Leaver & Rauschecker, 2010), (2) to investigate the existence of categorical responses to voices and sounds of objects in the VOTC and to elucidate whether these putative responses are unique to blind individuals (due to cross-modal plasticity) or whether they are also observable in sighted individuals, and (3) to explore the differences in the functional

connectivity profile of domain-selective regions between blind and sighted individuals.

## METHODS

### Participants

Thirty-three participants were recruited for this study. Sixteen early blind participants (EB; five women, age range = 23–62 years, mean = 45, *SD* = 12 years; Table 1) and 15 sighted control participants (SC; five women, age range = 22–61 years, mean = 41.9, *SD* = 11.8 years) took part in the auditory experiment. Participants were matched for age, sex, handedness, educational level,

and musical experience. Seventeen sighted participants (including the 15 that participated in the auditory experiment) were also tested in an independent visual experiment (seven women, age range = 22–61 years, mean = 40.7, *SD* = 11.6 years). At the time of testing, the blind participants were either totally blind or had only rudimentary sensitivity for brightness differences and no pattern vision. In all cases, blindness was attributed to peripheral deficits with no neurological impairment (Table 1). All the procedures were approved by the research ethical and scientific boards of the “Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal” and the “Quebec Bio-Imaging Network.” Experiments were undertaken with the consent of each participant.

**Table 1.** Characteristics of the Blind Participants

<i>Participant</i>	<i>Age</i>	<i>Sex</i>	<i>Handedness</i>	<i>Residual Vision</i>	<i>Onset</i>	<i>Etiology</i>	<i>Educational Level</i>	<i>Musical Experience</i>
EB01	48	M	R	None	1 y	Glaucoma	University	Yes
EB02	44	M	R	DL	0	Leber's congenital amaurosis	University	No
EB03	60	F	R	None	0	Retinopathy of prematurity	High school	Yes
EB04	43	M	R	None	0	Retinopathy of prematurity	High school	Yes
EB05	36	F	R	None	10 m (OS)/ 3.5 y (OD)	Retinoblastoma	Cegep	No
EB06	31	M	R	None	0	Leber's congenital amaurosis	University	Yes
EB07	55	M	R	None	2 m	Electrical burn of optic nerves	High School	No
EB08	51	M	R	None	0	Glaucoma	University	Yes
EB09	45	M	R	None	0	Retinopathy of prematurity	University	Yes
EB10	31	F	A(R)	None	0	Retinopathy of prematurity	High School	No
EB11	51	M	A(R)	None	0	Major eye infection	University	Yes
EB12	62	M	R	DL	0	Congenital cataracts	Cegep	Yes
EB13	23	M	R	DL	0	Glaucoma and microphthalmia	University	Yes
EB14	28	M	R	None	0	Retinopathy of prematurity	University	Yes
EB15	57	F	R	None	0	Chorioretinal atrophy (Toxoplasmosis)	Cegep	Yes
EB16	58	F	R	None	0	Retinopathy of prematurity	Cegep	Yes

Handedness was assessed using an adapted version of the Edinburgh inventory. Blind and sighted participants were classified as musicians if they had practiced a musical instrument or had vocal training for at least 2 years on a regular basis (at least 2 hours a week). A = Ambidextrous; M = male; F = female; DL = diffuse light; m = months; y = years; OS = left eye; OD = right eye; Cegep = 2 years of education between high school and university.

## Experimental Design and Stimuli

Participants in both groups were scanned in an auditory run and were blindfolded throughout the fMRI acquisition. Sighted participants were additionally scanned in a visual run on a separate day. To familiarize the participants to the fMRI environment, participants underwent a training session in a mock scanner while listening to recorded scanner noise. To ensure that all object sounds were clearly recognized, participants were familiarized to all stimuli before practicing the tasks in the mock scanner. In the scanner, auditory stimuli were delivered by means of circumaural, fMRI-compatible headphones (Mr Confon, Magdeburg, Germany). Visual stimuli were projected on a screen at the back of the scanner and visualized through a mirror (127 mm × 102 mm) that was mounted at a distance of approximately 12 cm from the eyes of the participants.

### Auditory Stimuli

Auditory stimuli consisted of four different categories: human voices, object sounds, and their respective scrambled versions (hereafter V, O, SV, SO, respectively; Figure 1). All sounds were monophonic, 16-bit, and sampled at 44.1 Hz. Voices and object sounds were cut at 995 msec (5 msec fade-in/fade-out). A 5-msec silence was added at the beginning of the stimuli to prevent clicking.

Human voices consisted of eight exemplars of each of five vowels (“a,” “e,” “i,” “o,” “u”), pronounced by 40 different speakers (half were male; Figure 1A). Object sounds consisted of 40 sounds of man-made artifacts (Figure 1B). In line with previous studies (Lewis et al., 2005, 2012; Lewis, Talkington, et al., 2011), object sounds included a range of nonverbal sounds of nonliving objects, namely, human action sounds (lighting a match, jingling coins, hammering a nail, water flushing in the sink, jigsaw, manual saw, typing on a writing machine (2), dropping ice cubes in a glass, broom falling on the floor, pouring water in a glass, velcro, jingling keys, plate breaking, zipper, cleaning brush), bells and musical instruments (christmas bells, shop bell, door bell, piano, flute, drums, maracas, trumpet, guitar, tom tom, bicycle bell, harp), and automated machinery (car horn, train horn, helicopter, cuckoo clock, phone tone, motorcycle, gun bursts, printer, automatic camera, police car, tractor, hair dryer). These sounds were selected from a larger sample of 80 sounds in a pilot study based on the recognition performance of 10 sighted participants. In this pilot study, participants were asked to name each sound and rate it on a scale from 1 to 10 according to how much the sound was characteristic (representative) of the object. The 40 sounds with the highest rates (all above 7) were selected for the fMRI experiment. Before the actual fMRI experiment and before practicing the repetition task (see Paradigm section) in the simulator, all participants were familiarized to each of the

object sounds: They were asked to name each object after listening to its sound. Recognition accuracy during familiarization was at ceiling and was therefore not monitored.

Scrambled versions of the vocal and object sounds were obtained using MATLAB (The MathWorks, Inc., Natick, MA; Figure 1C and D). Scrambling was inspired by the method of Belin and colleagues (2000, 2002) but differed in that the scrambling of amplitude and phase components was conducted separately within frequency windows (here 700 Hz) instead of time windows. Each vocal and object sound was submitted to a fast Fourier transformation, and the resulting components were separated into frequency windows of ~700 Hz based on their center frequency. Scrambling was then performed by randomly intermixing the magnitude and phase of each Fourier component (Belin et al., 2000, 2002) within each of these frequency windows separately. The inverse Fourier transform was then applied on the resulting signal. The output was a sound of the same length of the original sound with similar energy within each frequency band (Figure 1A–D, power spectrum, and E, decibel level). For scrambled vocal sounds only, the envelope of the original voice was further applied on the output signal (Figure 1C). This was not done for scrambled object sounds because the application of the original envelope in this case led to recognition of many scrambled object sounds despite the scrambling (Figure 1D). Hence, for these sounds, a 5-msec ramp was applied in the beginning and at the end, and a 5-msec silence was added at the beginning. Following standard practice, voices, object sounds, and their scrambled versions were equalized in root mean square level (Giordano et al., 2013; Belin et al., 2000, 2002).

Measures of spectral content (FC and FCSD) and spectral structure (HNR) were extracted for each sound using Praat as described in Leaver and Rauschecker (2010) and are depicted in Figure 1F. FC reflects the center of gravity of the spectrum, an approximation of overall frequency content, and FCSD is its standard deviation across the spectrum. HNR measures the ratio of the strength of the periodic and aperiodic (noisy) components of a signal.

The scrambling method used in this study has the important advantage of altering the perception of the stimuli as object- and voice-like (sound examples are provided as supplemental material) while leaving the frequency spectrum of the original sound relatively unaffected (Figure 1). Temporal structure is relatively preserved only in the case of scrambled voices by application of the original sound envelope (Figure 1C). In contrast, harmonicity, typically higher for vocal stimuli, is altered by scrambling (Figure 1F, HNR).

This factorial design thus allows for control of the frequency spectrum of objects and voices by contrasting these sounds to their scrambled versions. This is crucial, considering recent evidence that occipital regions in

congenitally blind participants respond differently to distinct auditory frequencies (Watkins et al., 2013). Beyond controlling for low-level parameters of the sounds, this paradigm further allows for the assessment of the degree to which low-level parameters contribute to a given categorical response (i.e., as it is the case for instance when a larger categorical response for voices relative to objects is also found when contrasting the corresponding scrambled control sounds; Table 2).

### *Visual Stimuli*

In this experiment, stimuli consisted of four different categories: pictures of faces, objects, and their phase-scrambled version (hereafter F, O, SF, SO, respectively; Rossion et al., 2012). The face category consisted of full front pictures of 50 different faces (half male; between 170 and 210 pixels width and 250 pixels height) that were cropped for external features and embedded in a white rectangle (220 pixels width  $\times$  270 pixels height). Similarly, the objects category consisted of pictures of 50 different objects (170–210 pixels width  $\times$  250 pixels height) inserted in a white rectangle (220 pixels width  $\times$  270 pixels height). The phase-scrambled pictures were used to control spatial frequencies and pixel intensity in each color channel (RGB) in the face and in the object categories. Phase-scrambled pictures were created using a Fourier phase randomization procedure by replacing the phase of each original image by the phase of a uniform noise allowing for amplitude to be conserved in each frequency band (Sadr & Sinha, 2004).

Pictures of objects consisted of the following items: fan, lamp, hat, garbage, coins, bag, balloon, stroller, glass, jeans, pair of boots, jewel, small bell, sofa, door, present, hairdryer, vase, hourglass, frame, headphones, key, clipboard, wine barrel, guitar, mug, toothbrush, tennis racket, alarm clock, tap, wardrobe, gloves, car tire, scissors, adjustable wrench, lens, screw, drum, trumpet, water gallon, light bulb, bucket, rugby ball, padlock, ring, paper bag, pepper, apple, plastic bag, ruby.

### *Paradigm*

Both the auditory and the visual experiments consisted of a single run lasting about 18 min and with 10 repetitions of each of the four conditions that alternated in blocks of 21 sec. Blocks were separated by a 7-sec baseline (silence and white fixation cross on a black background in the auditory and visual experiment, respectively). In each block, 20 items (sounds, pictures) were presented with a 50-msec ISI. Participants were instructed to detect a repetition in the stimuli (the same sound or picture presented twice in a row) by pressing a key with the right index finger. Emphasis was put on accuracy rather than speed. The number of repetitions within each block was unpredictable (i.e., two to four repetitions), thus ensuring

that participants kept attending to the stimuli throughout the block. Within each condition, there were four blocks with one repetition, four blocks with two repetitions, and two blocks with three repetitions, for a total of 18 targets per condition. This design aimed at matching as best as possible attention, arousal, and motor components between conditions.

### **Behavioral Analysis**

Behavioral performance in the auditory experiment was analyzed by submitting accuracy scores (hits – false alarms) to a mixed 2 Group (blind, sighted; between-subject factor)  $\times$  4 Condition (V, O, SV, SO) ANOVA. In the visual experiment, a repeated-measures ANOVA was conducted with Condition (faces, objects, scrambled faces, scrambled objects) as a within-subject factor. A Greenhouse–Geisser correction was applied to the degrees of freedom and significance levels whenever an assumption of sphericity was violated.

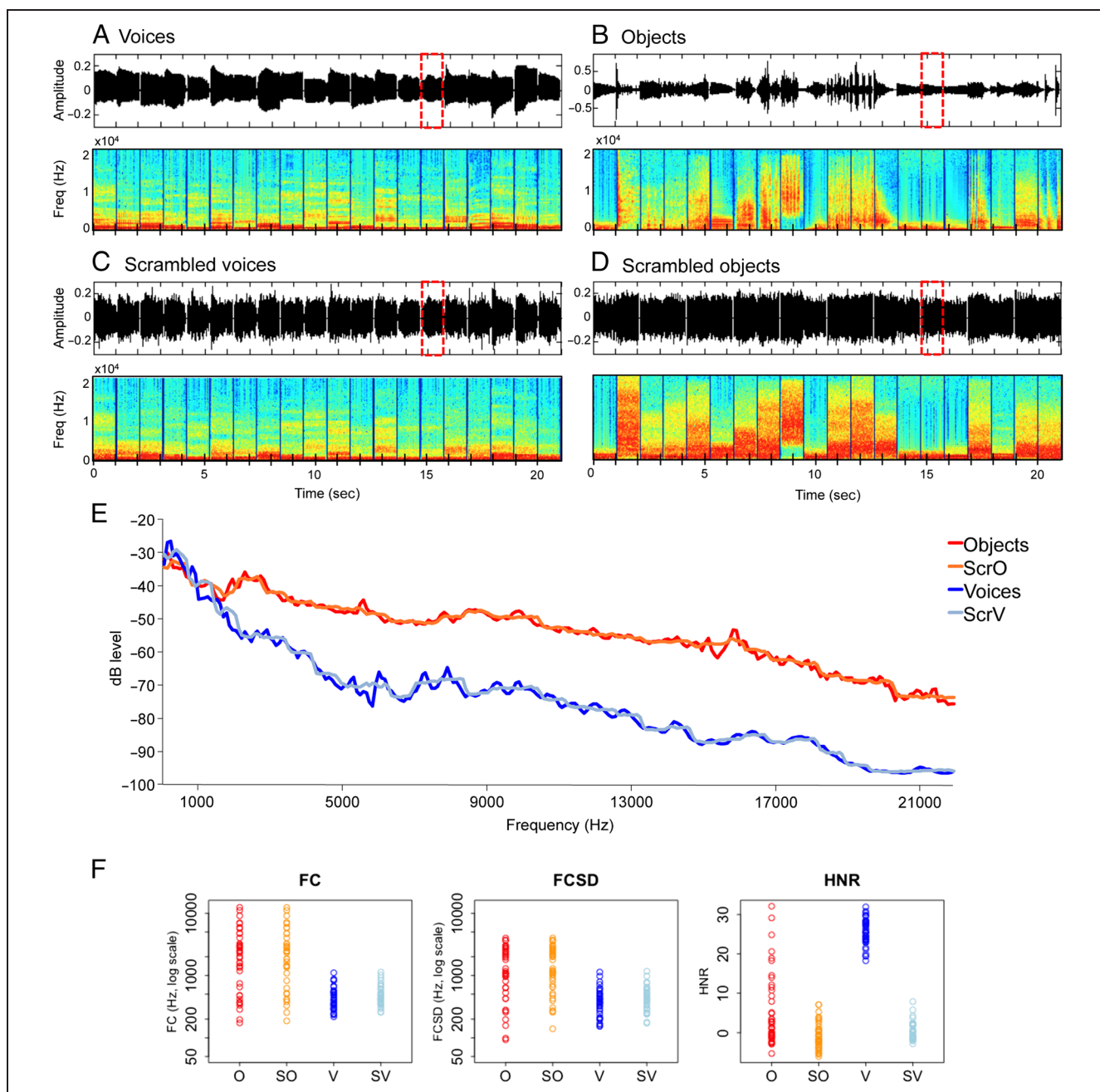
### **fMRI Data Acquisition**

fMRI series were acquired using a 3-T TRIO TIM system (Siemens, Erlangen, Germany), equipped with a 12-channel head coil. Multislice T2\*-weighted fMRI images were obtained with a gradient echo-planar sequence using axial slice orientation (repetition time = 2200 msec, echo time = 30 msec, flip angle = 90°, 35 transverse slices, 3.2 mm slice thickness, 0.8 mm interslice gap, Field of View = 192  $\times$  192 mm<sup>2</sup>, matrix size = 64  $\times$  64  $\times$  35, voxel size = 3  $\times$  3  $\times$  3.2 mm<sup>3</sup>). Slices were sequentially acquired along the z-axis in feet-to-head direction. The four initial scans were discarded to allow for steady state magnetization. The participants' head was immobilized using foam pads. A structural T1-weighted 3-D magnetization-prepared rapid gradient echo sequence (voxel size = 1  $\times$  1  $\times$  1.2 mm<sup>3</sup>, matrix size = 240  $\times$  256, repetition time = 2300 msec, echo time = 2.91 msec, inversion time = 900 msec, Field of View = 256; 160 slices) was also acquired for all participants.

### **fMRI Analysis**

Functional volumes from the auditory and visual experiments were preprocessed and analyzed separately using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK; [www.fil.ion.ucl.ac.uk/spm/software/spm8/](http://www.fil.ion.ucl.ac.uk/spm/software/spm8/)), implemented in MATLAB R2008a (The MathWorks, Inc.).

Preprocessing included slice timing correction of the functional time series (Sladky et al., 2011), realignment of functional time series, coregistration of functional and anatomical data, creation of an anatomical template using DARTEL (a template including participants from both groups in the auditory experiment and a template



**Figure 1.** Stimuli in the auditory experiment. Top: Sound properties in a representative 21-sec block in the (A) voice and (B) object sounds conditions and in their respective (C and D) scrambled conditions. Graphs represent sound amplitude as a function of time and frequency spectrum as a function of time. Red dashed lines indicate the occurrence of a target sound. (E) Bode magnitude plot expressing the magnitude in decibels as a function of frequency for the four blocks depicted in the top part of the figure. (F) Measures of spectral content (FC and FCSD) and spectral structure (HNR) are plotted in color for each stimulus. Scrambling leaves the frequency spectrum relatively unaffected while altering harmonicity.

including sighted participants only in the visual experiment; Ashburner, 2007), spatial normalization of anatomical and functional data to the template, and spatial smoothing (Gaussian kernel, 8 mm FWHM). The creation of a study-specific template using DARTEL was performed to reduce deformation errors that are more likely to arise when registering single-participant images to an unusually shaped template (Ashburner, 2007). This is particularly relevant when comparing blind and sighted participants,

given that blindness is associated with significant changes in the structure of the brain itself, particularly within the occipital cortex (Jiang et al., 2009; Park et al., 2009; Pan et al., 2007).

#### Activation Analyses

The analysis of fMRI data, based on a mixed effects model, was conducted in two serial steps, accounting for fixed

and random effects, respectively. In the auditory experiment, changes in brain regional responses were estimated for each participant by a general linear model including the responses to each of the four conditions (V, O, SV, SO). These regressors consisted of boxcar function convolved with the canonical hemodynamic response function. The movement parameters derived from realignment of the functional volumes (translations in  $x$ ,  $y$ , and  $z$  directions and rotations around  $x$ ,  $y$ , and  $z$  axes) and a constant vector were also included as covariates of no interest. High-pass filtering was implemented in the design matrix using a cutoff period of 128 sec to remove low-frequency noise and signal drift from the time series. Serial correlations in fMRI signal were estimated using an autoregressive (Order 1) plus white noise model and a restricted maximum likelihood algorithm. Linear contrasts tested the main effect of each condition ([V], [O], [SV], [SO]) and the contrasts between conditions ([V > O], [O > V], [V > SV], [O > SO]) and generated statistical parametric maps [SPM (T)]. These summary statistics images were then further spatially smoothed (Gaussian kernel 6 mm FWHM) and entered in a second-level analysis, corresponding to a random effects model, accounting for intersubject variance. For each of the above-mentioned contrasts, one-sample  $t$  tests were performed within each group, and two-sample  $t$  tests were performed to compare effects between groups (EB > SC, SC > EB). Voice-selective voxels were identified by means of an “AND” conjunction contrast of [V > O] and [V > SV] (Nichols, Brett, Andersson, Wager, & Poline, 2005). Object-selective voxels were identified by means of an “AND” conjunction contrast of [O > V] and [O > SO]. These contrasts thus identified voxels responding more to a category of sound relative to the other and for which this difference could not be accounted by differences in global energy or frequency spectrum.

These two conjunction analyses were conducted separately for each group (testing for voxels fulfilling these requirements in each group), jointly between groups (testing for voxels fulfilling these requirements in both groups, that is, independently of visual experience; see Figure 2A and B), and on between-group two-sample  $t$  tests (testing for voxels fulfilling these requirements in one group more than in the other; Figure 4A).

Preprocessing and statistical analyses of the fMRI data in the visual experiment were performed as in the auditory experiment, with the exception that random effects were only calculated based on a one-sample  $t$  test (no group comparison).

Statistical inference was performed at a threshold of  $p < .05$  after correction for multiple comparisons (family-wise error [FWE] method) over either the entire brain volume or over small spherical volumes (15-mm radius) located in structures of interest (see table legends). Significant clusters were anatomically labeled using a brain atlas (Petrides, 2012). Beta-weight extraction was used for visualization in figure charts only, whereas statistical ana-

lyses were performed on the single-voxel data, as per convention.

### *Psychophysiological Analyses*

Psychophysiological interaction (PPI) analyses were computed to identify any brain regions showing a significant change in functional connectivity with seed areas as a function of experimental condition (O, V) and group (EB > SC). Seed areas were selected using a two-step approach. First, all of the regions that were significant in the contrasts of interest, namely, regions showing preferential responses to voices and objects in both groups (Figure 2A and B) and those selectively responding to object sounds in the blind group only (Figure 4A) were selected as potential seed areas. Second, among significantly active regions, seeds for PPI analyses were selected based on previous literature (see the list of selected regions in Table 6).

In each participant, the first eigenvariate was extracted using the single value decomposition of the time series across the voxels in a 10-mm radius sphere centered on the peak of activation reported at the group level. New linear models were generated using three regressors. The first two regressors were modeled as covariates of no interest and represented the condition (i.e., psychological regressor: O > V and V > O) and the raw activity extracted in the seed area (i.e., physiological regressor), respectively. The third, psychophysiological regressor, represented the interaction of interest between the first (psychological) and the second (physiological) regressor. To build this third regressor, the underlying neuronal activity was first estimated by a parametric empirical Bayes formulation combined with the psychological factor and subsequently convolved with the hemodynamic response function (Gitelman, Penny, Ashburner, & Friston, 2003). Thus, variance explained by the psychophysiological regressor is above variance explained solely by the main effects of task (psychological regressor) and physiological correlation (O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012). Movement parameters and a constant vector were also included as covariates of no interest. A significant PPI indicated a change in the regression coefficients between any reported brain area and the seed area related to the experimental condition (O > V, V > O). Next, individual summary statistic images obtained at the first-level (fixed-effects) analysis were spatially smoothed (6-mm FWHM Gaussian kernel) and entered in a second-level (random-effects) analysis using a one-sample  $t$  test. Two-sample  $t$  tests were then performed to compare these effects between groups.

Statistical inference was performed as for the activation analyses with the exception that here we only report those regions showing a functional connectivity change in the blind group compared with the sighted group (EB > SC) and where the effect is driven by the blind group. For this purpose, small volume correction (SVC)

(corrected for multiple comparisons using FWE method at  $p < .05$ ) were performed on the between-group functional connectivity maps (two-sample  $t$  tests thresholded at  $p(\text{unc}) < .001$ ,  $EB > SC$ ) and inclusively masked by the functional connectivity map in the blind group (one-sample  $t$  test,  $p(\text{unc}) < .001$ ).

## RESULTS

### Behavioral Results

#### Auditory Experiment

There was no effect of group ( $p > .15$ ), indicating that overall accuracy (hits – false alarms) did not differ between blind (mean = 91.58%,  $SD = 11.22\%$ ) and sighted participants (mean = 86.2%,  $SD = 9.42\%$ ). There was a significant effect of condition,  $F(2.23, 64.71) = 4.493$ ,  $p = .012$ , but group did not interact with this effect ( $p > .9$ ). Two-tailed paired  $t$  tests, collapsed across groups, revealed that detecting repetitions in the scrambled objects condition (mean = 83.87%,  $SD = 14.79\%$ ) was more challenging than in the other conditions (scrambled voices [SV]: mean = 90.86%,  $SD = 12.79\%$ ,  $t(30) = -3.198$ ,  $p = .003$ ; objects [O]: mean = 92.29%,  $SD = 10.21\%$ ,  $t(30) = -3.591$ ,

$p = .001$ ; voices [V]: mean = 88.89%,  $SD = 15.18\%$ ,  $t(30) = -2.005$ ,  $p = .05$ ).

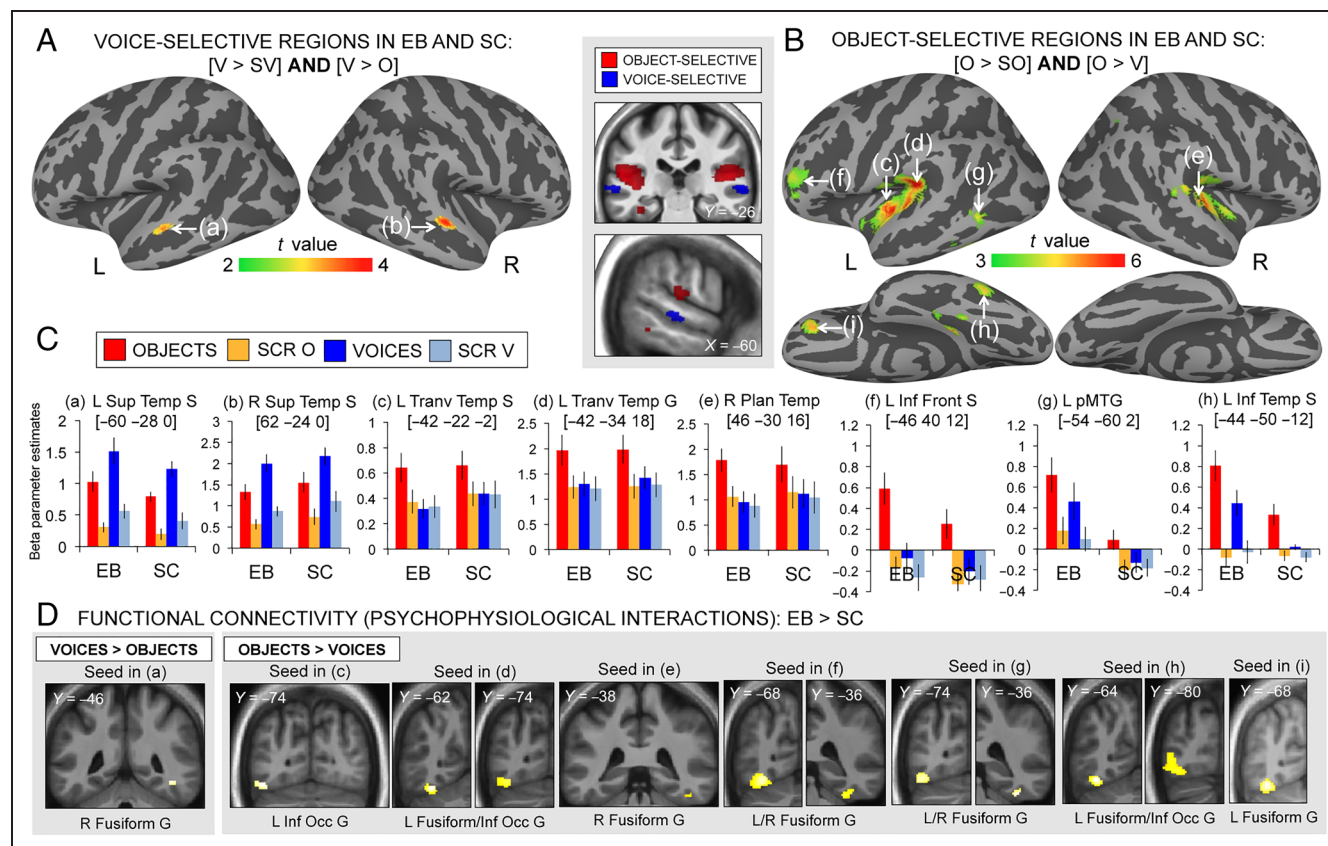
#### Visual Experiment

There was a significant effect of condition ( $F(3, 48) = 9.663$ ,  $p < .001$ ). Two-tailed paired  $t$  tests revealed that accuracy (hits – false alarms) was lower in the scrambled faces condition (mean = 73.2%,  $SD = 21.26\%$ ) compared with the remaining conditions (faces: mean = 85.29%,  $SD = 11.44\%$ ,  $t(16) = -2.766$ ,  $p = .014$ ; objects: mean = 92.16%,  $SD = 9.43\%$ ,  $t(16) = -4.788$ ,  $p < .001$ ; scrambled objects: mean = 87.58%,  $SD = 11.87\%$ ,  $t(16) = -3.507$ ,  $p = .003$ ) and lower in the face than in the object condition ( $t(16) = -2.159$ ,  $p = .046$ ).

### fMRI Results—Activation Analyses

#### Object and Voice Categorical Responses Common to Early Blind and Sighted Participants

Between-group conjunction (AND) analyses identified brain regions commonly responsive in both groups when listening to voices compared with both scrambled voices



**Figure 2.** Categorical responses to (A) voices and (B) object sounds common to blind and sighted participants. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .005$  with  $k > 90$  (A) and  $p(\text{unc}) < .001$  with  $k > 8$  (B). Color bar represents  $t$  values. (C) Mean activity estimates (arbitrary units  $\pm SEM$ ) are plotted for the four auditory conditions in significant peaks depicted in A and B. (D) PPI analyses as a function of group (blind > sighted) and experimental condition ( $V > O$  and  $O > V$ ) based on the peaks of activation depicted in A and B. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .005$  and masked inclusively by the main effect in the blind ( $p(\text{unc}) < .005$ ). EB = early blind; SC = sighted controls; L = left; R = right; S = sulcus; G = gyrus. See Tables 2 and 6 for a list of brain regions depicted in this figure.



and objects and when listening to objects compared with scrambled objects and voices.

Categorical responses to voices common to blind and sighted participants were found in two circumscribed areas within the superior temporal sulci bilaterally (Figure 2A and Table 2). Inspection of the individual data revealed that such responses were present in each single participant (data not shown). Importantly, these areas also strongly responded to low-level characteristics of voices (SV > SO; Table 2).

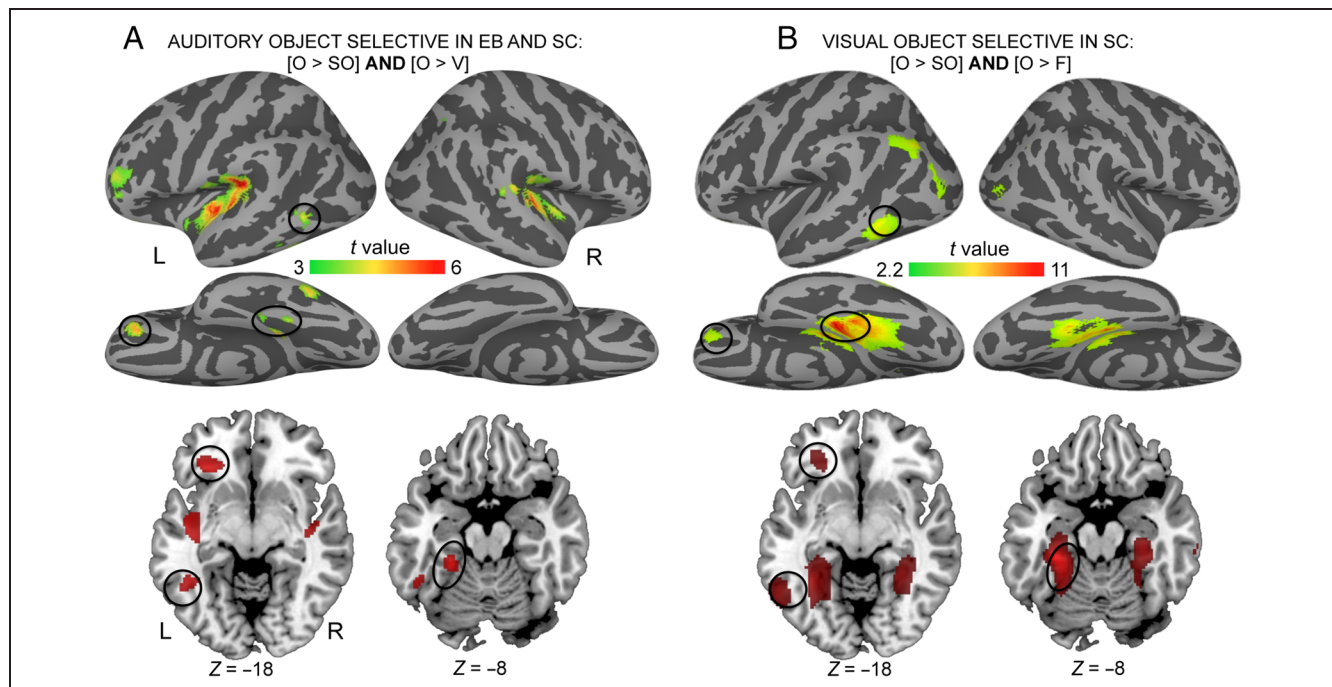
In both blind and sighted participants, object sounds preferentially activated large portions of the auditory cortex bilaterally—although stronger in the left hemisphere—in the medial part of the transverse temporal gyrus (A1), extending laterally along the lateral fissure and posteriorly to the planum temporale (Figure 2B and Table 2). In the left hemisphere, additional clusters of activation were found within the inferior frontal gyrus and sulcus and within the temporal cortex, in the posterior middle temporal gyrus (pMTG) extending to the inferior temporal sulcus and fusiform gyrus. Contrary to voice-responsive regions, there was no contribution of low-level parameters to the response observed in object-responsive regions, neither in the sighted group nor in the blind group (no significant responses in the contrast SO > SV). Object-selective areas common to both groups in the left-lateralized inferior frontal gyrus, pMTG, and fusiform gyrus overlapped with visual-selective areas responsive to

pictures of objects in the sighted group (Figure 3B and Table 3) (as also confirmed by a conjunction analysis, data not shown).

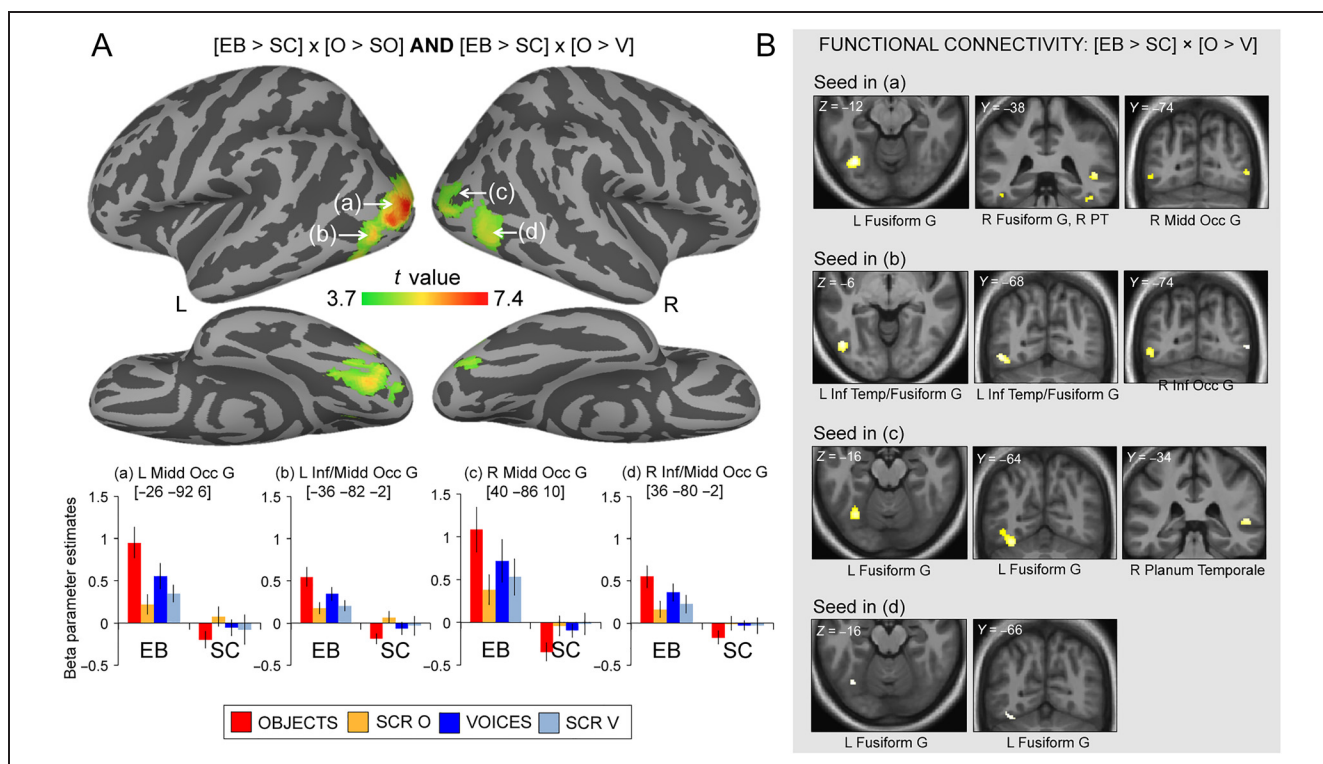
### Object and Voice Categorical Responses Specific to Early Blind Participants

Two-sample *t* tests were then performed to compare these effects between groups. A conjunction (AND) analysis was conducted on the two-sample *t* tests [EB > SC] × [O > V] and [EB > SC] × [O > SO] to identify regions specifically activated in the blind group (relative to sighted) for the processing of object sounds relative to both scrambled objects and voices (Figure 4A and Table 4). This analysis revealed large bilateral activations in the occipital cortex that peaked in the middle and inferior occipital gyri bilaterally. There was no contribution of low-level parameters to the categorical response observed for objects (no significant responses in the contrast SO > SV). These between-group effects [EB > SC] were driven by the blind group (Table 4). The reverse group comparisons [SC > EB] did not reveal any region that was more strongly responsive in the sighted to object sounds relative to voices or scrambled objects.

On the lateral portion of the occipitotemporal cortex, these object-selective responses specific to the blind group partially overlapped with shape-selective visual cortex localized in the sighted group using the contrast [O > SO] (lateral occipital complex [LOC]; Malach et al., 1995;



**Figure 3.** Objects categorical responses across modalities (visual and auditory) and populations (blind and sighted). (A) Categorical responses to object sounds common to blind and sighted participants. (B) Categorical responses to pictures of objects in the sighted. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .001$  with  $k > 8$ . Color bar represents *t* values. Regions in the left inferior frontal gyrus, pMTG, and fusiform gyrus are responsive across groups in the auditory modality (A) and across the auditory and the visual modalities in the sighted group (A and B). EB = early blind; SC = sighted controls; L = left; R = right; S = sulcus; G = gyrus; O = objects; V = voices; SO = scrambled objects; SV = scrambled voices; F = faces; SF = scrambled faces. See Tables 2 and 3 for a list of brain regions depicted in this figure.

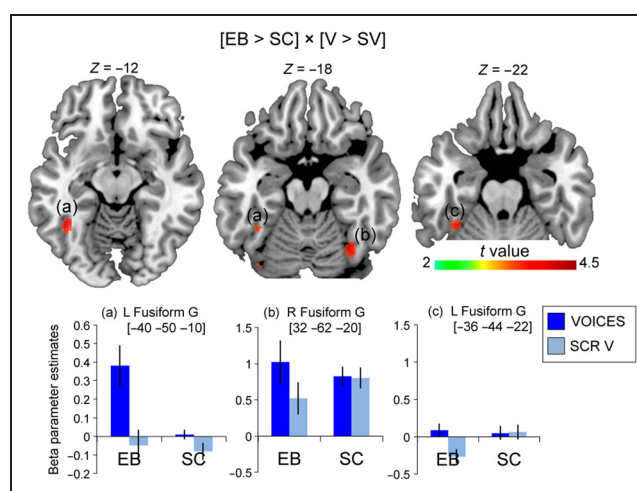


**Figure 4.** (A) Categorical responses to object sounds specific to the blind. Color bar represents  $t$  values. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .0001$ . Mean activity estimates (arbitrary units  $\pm SEM$ ) are plotted for the four auditory conditions in significant peaks depicted in A. (B) PPI analyses as a function of group (blind > sighted) and experimental condition (O > V) based on the peaks of activation depicted in A. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .001$  and masked inclusively by the main effect in blind group ( $p(\text{unc}) < .001$ ). See Tables 4 and 6 for a list of brain regions depicted in this figure.

Table 3), as also confirmed by a conjunction analysis (data not shown).

A conjunction (AND) analysis was conducted on the two-sample  $t$  tests [EB > SC] × [V > O] and [EB > SC] × [V > SV] to identify regions specifically activated in the blind group (relative to the sighted group) for the processing of voices relative to both scrambled voices and objects. This analysis yielded no significant response, even at a very lenient threshold of  $p < .01$  uncorrected. Considering each of these  $t$  tests separately revealed that voices relative to scrambled voices [EB > SC] × [V > SV] elicited higher responses in the blind group in the fusiform gyrus bilaterally (Figure 5 and Table 5). This effect was driven by the blind group (Table 5). In contrast, voices compared with objects [EB > SC] × [V > O] did not elicit any larger activation in the blind group relative to the sighted group. The reverse group comparisons [SC > EB] did not reveal any region that was more strongly responsive in the sighted group for voices relative to objects or scrambled voices.

In summary, voice-selective responses relative to both object sounds and scrambled voices were limited to the superior temporal sulci (auditory cortices) in both groups with no evidence of cross-modal responses in the VOTC in either group (as also accounted by the individual data, data not shown).



**Figure 5.** Regions showing stronger response to voices relative to scrambled voices specifically in the blind. For illustration purposes, activity maps are displayed at  $p(\text{unc}) < .001$  with  $k > 10$ . Mean activity estimates (arbitrary units  $\pm SEM$ ) are plotted for voices and scrambled voices in significant peaks. EB = early blind; SC = sighted controls; L = left; R = right; S = sulcus; G = gyrus. See Table 5 for a list of brain regions depicted in this figure.

## fMRI Results—Psychophysiological Analyses

PPI analyses were computed to identify any brain regions showing a significant change in functional connectivity with specific seed areas as a function of experimental condition ( $O > V$  and  $V > O$ ) and group ( $EB > SC$ ; Table 6).

Among the two regions selectively responsive to voices in both groups (Figure 2A), the left STS displayed an increase in functional connectivity with the right fusiform gyrus in the blind group during voice processing compared with object sounds processing (Figure 2D, a).

Among the regions that selectively responded to object sounds in both groups (Figure 2B), several seed areas located in auditory cortices showed a significant increase in functional connectivity with ventral occipitotemporal

regions during the processing of object sounds relative to processing of voices in blind relative to sighted participants (Figure 2D, c–e). Notably, the left primary auditory cortex showed increased connectivity with the left inferior occipital gyrus (Figure 2D, c), the left transverse temporal gyrus showed increased connectivity with the left inferior occipital gyrus and the left posterior fusiform gyrus (Figure 2D, d), whereas the right planum temporale showed increased connectivity with the right fusiform gyrus (Figure 2D, e). Regions located in the left inferior frontal gyrus and the sulcus and those located in the left temporal cortex (left pMTG and the left inferior temporal sulcus), all showed an increase in functional connectivity with a circumscribed region located in the left posterior fusiform gyrus (Figure 2D, f–i). In addition, the left inferior frontal

**Table 2.** Categorical Responses to Voices and Object Sounds Common to Blind and Sighted Participants and Responses to Low-level Properties of Voices Common to Blind and Sighted Participants

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
<i>Between-group (AND) Conjunction: [V &gt; SV] ∩ [V &gt; O]</i>						
R superior temporal S	39	62	−24	0	3.45	.022
L superior temporal S	9	−60	−28	0	3.23	.04
<i>Between-group (AND) Conjunction: [O &gt; SO] ∩ [O &gt; V]</i>						
L planum temporale	1293	−50	−28	8	4.83	.013*
L transverse temporal G		−42	−34	18	4.80	.014*
L transverse temporal S (A1)		−42	−22	−2	4.79	.015*
R Heschl's G (A1)	956	50	−26	10	4.65	.026*
R planum temporale		46	−30	16	4.61	.030*
R planum temporale		60	−36	16	4.02	.003
L inferior frontal G (orbital part)	176	−32	30	−8	4.28	.001
L inferior frontal G (triangular part)	367	−46	40	12	4.18	.002
L inferior temporal S	283	−44	−50	−12	4.02	.003
L pMTG		−54	−60	2	3.96	.004
L collateral S (fusiform G)	81	−26	−38	−18	4.35	.001
L fusiform G	22	−36	−26	−22	3.36	.025
<i>Between-group (AND) Conjunction: [SV &gt; SO]</i>						
R superior temporal S	135	60	−24	0	3.47	.016
L superior temporal S	197	−60	−14	2	3.70	.008
		−66	−22	4	3.51	.014

Coordinates reported in this table are significant ( $p < .05$  FWE) after correction over small spherical volumes (SVC) or over (\*) the whole brain. *k* represents the number of voxels when displayed at  $p(\text{unc}) < .001$ . V = voices; O = objects; SV = scrambled voices; SO = scrambled objects; L = left; R = right; G = gyrus; S = sulcus. Coordinates used for SVC are as follows (in MNI space): R superior temporal S: [60 −32 4] (Gougoux et al., 2009); L superior temporal S: [−64 −28 2] (Gougoux et al., 2009); R planum temporale: [52 −44 10] (Lewis, Talkington, et al., 2011); L collateral S: [−28 −26 −26] (He et al., 2013); L inferior frontal G (triangular part): [−51 30 3] (Noppeney et al., 2003); L inferior frontal G (orbital part): [−28 34 −6] (Bar et al., 2001); L pMTG/ITG: [−52 −58 −6] (Peelen et al., 2013). Voice-selective regions common to blind and sighted participants are depicted in Figure 2A. Object-selective regions common to blind and sighted participants are depicted in Figures 2B and 3A.

**Table 3.** Visually Responsive Regions in the Sighted

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
<i>Shape-selective Regions in Vision: [O &gt; SO]</i>						
L collateral S	13013	-34	-34	-18	5.76	<.001*
L fusiform G		-46	-54	-18	5.41	.002*
L fusiform G		-38	-46	-10	4.96	.015*
L inferior occipital G		-46	-82	-6	5.64	.001*
L inferior occipital G		-56	-66	-14	4.98	.013*
L angular G		-48	-70	24	4.77	.030*
R inferior temporal G	8510	50	-70	-10	5.53	.001*
R parahippocampal G		32	-24	-22	5.52	.001*
R middle temporal G		52	-74	4	5.24	.005*
R middle temporal G		58	-14	-18	5.30	.004*
L superior frontal G	5739	-18	-66	6	5.42	.002*
L inferior frontal G (orbital part)		-48	42	-18	4.98	.013*
L inferior frontal G (orbital part)		-38	34	-14	4.89	.019*
L inferior frontal G (orbital part)		-30	30	-10	4.66	.046*
L superior frontal G		-4	42	36	4.86	.021*
<i>Object-selective Regions in Vision: [O &gt; SO] ∩ [F &gt; SF]</i>						
L collateral S	1295	-32	-34	-18	5.75	.000*
L fusiform G		-30	-50	-14	4.67	.044*
R collateral S	809	28	-26	-22	5.20	.005*
R fusiform G		30	-46	-12	4.75	.032*
L angular G	633	-38	-76	38	3.99	.005
L angular G		-42	-82	22	3.51	.032
L pMTG	360	-58	-58	-6	4.29	.002
L inferior temporal G		-56	-60	-10	4.16	.003
L inferior frontal G (orbital part)	148	-30	38	-10	3.77	.010

Coordinates reported in this table are significant ( $p < .05$  FWE) after correction over small spherical volumes (SVC) or over (\*) the whole brain. *k* represents the number of voxels when displayed at  $p(\text{unc}) < .001$ . F = faces; O = objects; SF = scrambled faces; SO = scrambled objects; L = left; R = right; G = gyrus; S = sulcus. Coordinates used for SVC are as follows (in MNI space): L inferior frontal G (orbital part): [-28 34 -6] (Bar et al., 2001); L pMTG/inferior temporal G: [-52 -58 -6] (Peelen et al., 2013), L angular G: [-48 -70 31] (Fairhall & Caramazza, 2013). Object-selective regions in vision are depicted in Figure 3B.

sulcus and pMTG showed an increase with the right anterior fusiform gyrus (Figure 2D, f–g), whereas the left inferior temporal sulcus showed an increase with the left inferior occipital gyrus (Figure 2D, h).

All reorganized occipital regions that showed a categorical response to object sounds only in the blind group (Figure 4A) showed increased connectivity with a circumscribed region in the left posterior fusiform gyrus (Figure 4B, a–d). In addition, the left middle occipital gyrus showed increased connectivity with the right anterior fusiform gyrus, the right middle occipital gyrus, and the

right planum temporale (Figure 4B, a); the left inferior occipital gyrus showed increased functional connectivity with the right inferior occipital gyrus (Figure 4B, b), and the right middle occipital gyrus showed an increase in connectivity with the right planum temporale (Figure 4B, c).

## DISCUSSION

This study investigated how visual experience impacts on the neural basis of object sounds and voice processing.

We used scrambled control sounds to control for low-level differences in the frequency spectrum between these categories of sounds and assess the contribution of low-level parameters to the categorical responses observed for object sounds and voices.

### Double Dissociation for Object Sounds and Voices in the Auditory Temporal Cortices

In both blind and sighted groups, a double dissociation was identified in the temporal cortex between separate regions that showed categorical responses to either object sounds or voices. These findings suggest that the cortical networks for processing these two auditory categories are at least partially separate (Figure 2 and Table 2). In line with previous work, categorical responses to sounds

of objects were observed along the lateral fissure bilaterally (Giordano et al., 2013; Lewis et al., 2005, 2012; Lewis, Talkington, et al., 2011), whereas categorical responses to voices were observed within bilateral superior temporal sulci (Belin, Fecteau, & Bédard, 2004; Belin et al., 2000, 2002). These findings support the notion that the auditory system—like the visual one—hosts a domain-specific organization where distinct areas preferentially respond to different categories of complex environmental sounds such as voices, animal vocalizations, tools, or musical instruments (Engel et al., 2009; Lewis et al., 2005, 2009; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002). These results are also in line with neuropsychological evidence demonstrating that lesions to portions of the temporal or temporoparietal cortex can lead to auditory agnosia, an impaired capacity to recognize complex natural

**Table 4.** Categorical Responses to Object Sounds Specific to the Blind

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
<i>Between-group Effects: [EB &gt; SC]</i>						<i>Main Effect in EB</i>					
<i>[O &gt; SO]</i>						<i>[O &gt; SO]</i>					
L inferior occipital G	7443	−26	−92	6	5.46	.001*	−20	−94	6	5.35	.001*
L middle/inferior occipital G		−36	−82	0	5.45	0.001*	−32	−84	−4	5.14	0.003*
L fusiform G		−36	−68	−14	4.86	.011*	−34	−68	−18	5.82	<.001*
R inferior OTC	6075	38	−66	−4	4.75	.018*	38	−66	−6	4.54	.039*
R inferior occipital G		36	−84	0	4.57	.035*	34	−82	−4	4.52	.043*
R fusiform S		34	−54	−16	4.52	.042*	38	−56	−20	5.31	.002*
<i>[O &gt; V]</i>						<i>[O &gt; V]</i>					
L middle occipital G	15494	−26	−92	4	5.57	<.001*	−24	−94	8	5.39	.001*
L middle/inferior occipital G		−28	−78	−4	5.32	0.001*	−26	−78	−4	5.54	0.001*
L superior occipital G		−26	−94	26	5.22	.002*	−26	−94	24	5.34	.001*
<i>[O &gt; SO] ∩ [O &gt; V]</i>						<i>[O &gt; SO] ∩ [O &gt; V]</i>					
L middle occipital G	5582	−26	−92	6	5.46	.001*	−24	−94	6	5.31	.001*
L middle/inferior occipital G		−36	−82	−2	5.21	0.002*	−30	−80	−6	5.14	0.003*
L lingual G		−20	−74	0	4.49	.047*	−20	−72	2	4.83	.013*
R inferior OTC	4821	38	−66	−4	4.75	.018*	38	−72	4	4.60	.031*
R inferior occipital G		36	−80	−2	4.40	.001	36	−76	−4	4.15	.002
R middle occipital G		40	−86	10	4.29	.001	40	−82	10	3.75	.008

Coordinates reported in this table are significant ( $p < .05$  FWE) after correction over small spherical volumes (SVC) or over (\*) the whole brain. *k* represents the number of voxels when displayed at  $p(\text{unc}) < .001$ . EB = early blind; SC = sighted controls; V = voices; O = objects; SV = scrambled voices; SO = scrambled objects; L = left; R = right; G = gyrus; S = sulcus; OTC = occipitotemporal cortex. For each region significant in the between-group contrasts (left-hand table), corresponding coordinates significant in the main effect in the blind are listed in the right-hand table. None of these regions were activated in the sighted group, indicating that the between-group effects (blind > sighted) are driven by these regions being responsive only in the blind group. Two regions (underlined in the left-hand table) showed selective deactivation in the sighted group, thus contributing to the between-group effects observed in the R inferior OTC [34 −68 4] ( $z = 3.21$ ) and in the L middle occipital G [−24 −84 6] ( $z = 3.25$ ). Coordinates used for SVC are as follows (in MNI space): R middle occipital G: [44 −74 8] (Gougoux et al., 2009). Regions listed showing specific responses to objects (relative to both scrambled objects and voices) in blind compared with sighted are depicted in Figure 4A.

**Table 5.** Voice-selective Regions Specific to the Blind

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
<i>Between-group Effects: [EB &gt; SC]</i>						<i>Main Effect in EB</i>					
<i>[V &gt; SV]</i>						<i>[V &gt; ScrV]</i>					
L lateral occipitotemporal S	147	-40	-50	-10	3.89	.006	-40	-50	-10	5.96	<.001*
L fusiform G		-36	-44	-22	3.40	.025	-40	-50	-10	5.96	<.001*
R fusiform G	75	32	-62	-20	3.50	.019	32	-62	-22	4.54	<.001*
<i>[V &gt; O]</i>						<i>[V &gt; O]</i>					
No significant voxels						No significant voxels					

Coordinates reported in this table are significant ( $p < .05$  FWE) after correction over small spherical volumes (SVC) or over (\*) the whole brain. *k* represents the number of voxels when displayed at  $p(\text{unc}) < .001$ . EB = early blind; SC = sighted controls; V = voices; O = objects; SV = scrambled voices; SO = scrambled objects; L = left; R = right; G = gyrus; S = sulcus. For each region significant in the between-group contrast (left-hand table), corresponding coordinates significant in the main effect in the blind group are listed in the right-hand table. None of these regions were activated or deactivated in the sighted group, indicating that the between-group effects (blind > sighted) are driven by these regions being responsive only in the blind group. Coordinates used for SVC are as follows (in MNI space): L fusiform/inferior temporal G: [-46 -48 -16] (Gougoux et al., 2009); R fusiform G: [34 -52 -16] (Gougoux et al., 2009). Regions more responsive to voices than scrambled voices in the blind group compared with the sighted group are depicted in Figure 5.

sounds despite preserved speech comprehension and visual object recognition (for a review, see Goll, Crutch, & Warren, 2010).

Our paradigm further allowed us to investigate the contribution of low-level parameters to these categorical responses. Our scrambling technique preserved the frequency content of the sounds (Figure 1) while altering harmonic and phase-coupling content that is known to contribute to the response in voice-selective regions (Lewis et al., 2009). In this study, higher responses in voice-selective areas were observed when contrasting scrambled voices and objects (SV > SO; Table 2). This finding suggests that the spectral frequency content of voices participates to the signal attributes that preferentially activate these regions. In other words, the preference observed for voices compared with objects in bilateral STS may emerge, at least partly, from the differential processing of low-level features that are typical of these two categories of sounds (for a similar interpretation in vision, see Andrews et al., 2010). In contrast, low-level parameters did not contribute to object categorical responses, as no area in the brain was more largely responsive to scrambled objects than to scrambled voices.

In contrast to previous findings, no between-group differences were observed for these domain-preferential responses in temporal “auditory” cortex (Hölig, Föcker, Best, Röder, & Büchel, 2014; Lewis, Frum, et al., 2011; Gougoux et al., 2009).

### Multimodal Object Representations in the Left Lateral and Ventral Occipitotemporal Cortex

Beyond the auditory cortex, preferential responses to object sounds common to both groups were found in left-lateralized inferior frontal and occipitotemporal regions

including the pMTG, inferior temporal gyrus, and fusiform gyrus (Figure 2B and Table 2). These left frontal and temporal regions have been associated with auditory object recognition (Lewis et al., 2004) and with semantic processing of concrete objects (Gold et al., 2006; Gough, Nobre, & Devlin, 2005; Wheatley, Weisberg, Beauchamp, & Martin, 2005; Sharp, Scott, & Wise, 2004; for a review, see Martin, 2007). Of note, in this study, these regions also responded selectively when sighted participants viewed pictures of objects (compared with both faces and scrambled objects; Figure 3B and Table 3).

In previous studies, similar left frontotemporal regions were found to be responsive in both early blind and sighted participants on tasks of action-related semantics (left inferior frontal [-51 30 3] and left posterior MTG [-63 -51 -6]; Noppeney et al., 2003), sounds of tools (left pMTG [-51 -57 3]; Lewis et al., 2005), heard names of tools (left pMTG [-50 -52 -3]; Peelen et al., 2013) and places (left parahippocampal gyrus/fusiform [-28 -26 -21]; He et al., 2013), as well as viewing pictures of corresponding objects in sighted participants (He et al., 2013; Peelen et al., 2013).

The finding of preferential responses to “objects” independent of the input modality (visual and auditory) and of visual experience in left occipitotemporal regions may suggest that these regions support a multimodal organization of object representations (Bi et al., 2016; Fairhall & Caramazza, 2013). Because all sounds of objects in this study were highly recognizable, we speculate that these abstract representations were automatically activated when participants were listening to these familiar environmental sound sources (see Lewis et al., 2004, for similar interpretation). In line with previous studies, it is possible that the left occipitotemporal regions that showed object-selective responses in this study contain

**Table 6.** Regions Showing Increased Functional Connectivity with Specific Seed Areas as a Function of Experimental Condition ( $O > V$  and  $V > O$ ) and Group (Blind  $>$  Sighted)

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
<i>PPI [EB &gt; SC] × [V &gt; O]</i>						
Seed areas in voice-selective regions common to EB and SC						
L superior temporal sulcus [-60 -28 0]						
R anterior fusiform G	2	42	-46	-8	3.24	.041
R superior temporal sulcus [62 -24 0]						
No significant voxels						
<i>PPI [EB &gt; SC] × [O &gt; V]</i>						
Seed areas in object-selective regions common to EB and SC						
L transverse temporal sulcus (A1) [-42 -22 -2]						
L inferior occipital G	1	-44	-74	-14	3.11	.056#
L transverse temporal G [-42 -34 18]						
L fusiform	12	-34	-62	-16	3.88	.006
L inferior occipital G	9	-44	-76	-14	3.34	.032
R planum temporale [46 -30 16]						
R anterior fusiform G	1	46	-38	-20	3.18	.052#
L inferior frontal G [-32 30 -8]						
L posterior fusiform G	101	-38	-66	-14	3.79	.007
L inferior frontal S [-46 40 12]						
L posterior fusiform G	160	-34	-68	-12	3.71	.011
R anterior fusiform G	5	36	-34	-24	3.16	.05#
L pMTG [-54 -60 2]						
R anterior fusiform G	7	40	-36	-22	3.8	.008
L posterior fusiform G	19	-38	-66	-14	3.57	.017
L inferior temporal S [-44 -50 -12]						
L posterior fusiform G	37	-38	-64	-14	3.87	.006
L inferior occipital G	6	-38	-80	-10	3.22	.044
<i>PPI [EB &gt; SC] × [O &gt; V]</i>						
Seed areas in object-selective regions specific to EB						
L middle occipital G [-26 -92 6]						
L posterior fusiform G	178	-36	-64	-12	4.07	.003
R anterior fusiform G	14	44	-38	-20	3.48	.021
R planum temporale	30	50	-36	4	3.80	.008
R middle occipital G	19	50	-74	0	3.46	.022
L inferior/middle occipital G [-36 -82 -2]						
R inferior occipital G	14	50	-74	2	3.94	.005
L inferior temporal/fusiform G	112	-46	-68	-6	3.82	.008

**Table 6.** (continued)

Area	<i>k</i>	<i>x</i> (mm)	<i>y</i> (mm)	<i>z</i> (mm)	<i>Z</i>	<i>p</i>
R middle occipital G [40 -86 10]						
L posterior fusiform G	202	-34	-64	-16	4.12	.002
R planum temporale	84	52	-34	8	4.11	.002
R middle/inferior occipital G [36 -80 -2]						
L posterior fusiform G	13	-36	-66	-16	3.27	.038

Seed areas are the ones resulting from the activation analyses (depicted in Figures 2A and B, 3A, and 4A). Regions showing increased connectivity with these seed areas in the blind group compared with the sighted group are listed in this table and are depicted in Figures 2D and 4B. Coordinates reported in this table are significant ( $p < .05$  FWE) after correction over small spherical volumes (SVC). Marginally significant clusters are indicated with (#). EB = early blind; SC = sighted controls; V = voices; O = objects; L = left; R = right; G = gyrus; S = sulcus. Coordinates used for correction over small spherical volumes are as follows (in MNI space): R fusiform G: [40 -36 -10] (Hölig et al., 2014); L fusiform G: [-36 -63 -18] (Noppeney et al., 2003); L inferior occipital G: [-36 -81 -15] (Noppeney et al., 2003); R planum temporale: [52 -44 10] (Lewis, Talkington, et al., 2011); R middle occipital G: [44 -74 8] (Gougoux et al., 2009).

an abstract representation of objects—such as the object’s meaning and the semantics knowledge associated with it (Bi et al., 2016; Fairhall & Caramazza, 2013; Bracci, Cavina-Pratesi, Ietswaart, Caramazza, & Peelen, 2012; Kassuba et al., 2011; Lewis et al., 2004)—which may develop independently of visual experience.

An alternative interpretation to multimodality in left occipitotemporal regions is visual mental imagery: The latter could have driven responses to auditory stimuli in the sighted group, as previously demonstrated for the tactile exploration of objects (Lacey, Flueckiger, Stilla, Lava, & Sathian, 2010). In other words, it is possible that similar activation patterns in the blind and sighted groups are related to different cognitive processes. Although no study to date can conclusively rule out visual imagery, we attempted to minimize this potential confound in our task by focusing the participant’s attention on the acoustical properties of the sounds. Future studies may test more directly whether the format of the representation in this region is identical between auditory and visual stimuli or whether this region still maintains separate representational formats for each modality input despite coding for both.

Functional connectivity analyses performed on the left inferior frontal cortex and pMTG showed a unique connectivity pattern in the blind group, namely, an increased task-related coupling with a circumscribed region of the left fusiform gyrus. These findings are remarkably similar to the ones reported by Noppeney et al. (2003). These authors found that the left inferior frontal cortex and left pMTG were activated in both early blind and sighted participants on a verbal semantic retrieval task. Yet, functional connectivity analyses on both of these regions revealed increased coupling with left-lateralized occipitotemporal areas only in the blind group. Altogether, these findings suggest that early visual deprivation, although preserving the responsiveness of the inferior frontal cortex and the left occipitotemporal cortex to object sounds (Figure 2B), impacts on these regions at the network level (Figure 2D). Worth noting, between-group differ-

ences in the connectivity profile of regions that show similar task-dependent activity level could support the notion that the cognitive processes underlying the recruitment of those regions partially differ between the blind and sighted groups.

### Cross-modal Categorical Responses to Object Sounds in Posterior Occipital Cortex of the Blind Group

A unique pattern of categorical responses to object sounds were found within large portions of the occipital cortex in the blind group, peaking in the middle and in the inferior occipital gyri bilaterally (Figure 4A and Table 4). This suggests a posterior expansion of cortical function related to the representations of object’s sounds in the blind group. These unique object-selective responses in the blind group partially overlapped with portions of shape-selective visual cortex localized visually in the sighted group (Malach et al., 1995; Table 3). This runs counter to the notion that object-related responses in the occipital cortex of the blind group rely solely on the processing of shape information conveyed by objects (either via touch or sensory substitution devices; Amedi et al., 2007, 2010) because our task did not involve shape processing. In line with the present findings, another study reported a trend for responses to object sounds in LOC in two congenitally blind participants when no imagery of shape was involved (Amedi et al., 2007). Together, these findings suggest that at least portions of LOC in early blind individuals contain representations of object sounds that are not related to shape and that these regions reorganize due to the lack of developmental vision as they do not activate in sighted individuals. In this study, cross-modal responses to object sounds in the blind group were most pronounced outside visual regions with preferential responses to either shape (LOC, object pictures > scrambled objects) or object (objects > face) in the sighted group: They extended more posteriorly in the occipital cortex (compare Figure 3B and Figure 4A).



Similar activation patterns with cross-modal responses extending posteriorly were reported in a previous study when congenitally blind participants performed a tactile recognition task (Amedi et al., 2010).

In this study, an important question pertains to the cognitive processes or representational format that supports the categorical responses to sounds of objects observed in the blind group. It has been proposed that environmental sounds that are perceived as “object-like,” such as those produced by automated machinery and man-made objects (as in this study), share common acoustical features, which may serve as low-level cues for their identification in a complex acoustic environment (Lewis et al., 2012). In this study, none of the reorganized occipital regions showed stronger responses to scrambled objects compared with scrambled voices, running counter to the assumption that categorical responses to object sounds are driven by low-level acoustic features that differentiate object sounds from voices (i.e., frequency spectrum). Instead, we argue that these occipital regions are an extension of more anterior occipitotemporal regions that commonly respond to sounds of objects in both sighted and blind participants (Figure 3) and support a more abstract representation of an object’s meaning. Several arguments account for this assumption. Object-selective cross-modal responses in the blind group were strongest in the left hemisphere and in the vicinity of regions previously reported as being responsive when early blind participants (compared with sighted participants) process meaningful speech (sentences and word lists compared with nonsemantic sentences and nonword lists; Bedny et al., 2011; Röder et al., 2002), generate semantically related verb to heard nouns (Amedi, Raz, Pianka, Malach, & Zohary, 2003; Burton, Diamond, & McDermott, 2003), and perform semantic decisions on heard nouns (Noppeney et al., 2003). Moreover, the functional connectivity pattern of these reorganized occipital regions in the blind group resembles to the one observed in the left pMTG and in the inferior frontal cortex, that is, a systematic increased coupling with ventral occipitotemporal regions (inferior temporal/fusiform gyrus), mainly in the left hemisphere. Hence, we propose that the left pMTG showing object’s sound selectivity in both blind and sighted participants and more posterior occipital regions showing preferential response to object’s sounds only in the blind group support similar functions, namely, an abstract representation of object semantics. Although such representations are shared across modalities and populations in more anterior occipitotemporal regions, posterior occipital regions might support similar functions only in the early blind due to cross-modal plasticity. Future studies may include speech material to further investigate the complex hierarchy from sounds to words (Perlovsky, 2011) that may shed light on the mechanisms that drive the occipital responses to object sounds observed in early blind participants in this study.

### **Lack of Cross-modal Categorical Responses to Voices in Early Blind or Sighted Participants**

In contrast to our observation of categorical responses to object’s sounds in the occipitotemporal cortex of the sighted group and, to a much larger extent, of the blind group, no such categorical responses to voices were observed outside the temporal auditory cortices in either group. This is unlikely to be related to a lack of sensitivity of our paradigm to detect voice-selective responses, as preferential responses to voices compared with both object sounds and scrambled voices were successfully identified in bilateral superior temporal sulci in every participant (data not shown) in both blind and sighted groups (Figure 2A for group-level statistics).

This lack of reorganization for voices in the VOTC of the blind group contrasts with recent evidence of cross-modal face-selective regions in TVAs in congenitally deaf individuals (Benetti et al., 2017), suggesting that compensatory brain plasticity in case of sensory deprivation follows principles of reorganization that are specific to the deprived sense. Genetic influence, developmental trajectory, susceptibility to plasticity, and the need for behavioral compensation are all factors that may potentially influence specific differences between the cross-modal plasticity observed in blind and deaf individuals (Frasnelli, Collignon, Voss, & Lepore, 2011).

Our findings suggest that different auditory functions are not equally susceptible to be supported by the occipital cortex in early visual deprivation. Similarly, we have previously shown that the spatial processing of sounds preferentially activates right dorsal regions of the occipital cortex in early blind participants, whereas pitch processing of sounds does not (Collignon et al., 2011; Collignon, Lassonde, Lepore, Bastien, & Veraart, 2007). We conclude that preferential responses to voices over nonvocal auditory objects are confined to the areas of the superior temporal sulci in early blind participants. Nevertheless, given that functional connectivity analyses identified unique patterns of connectivity between the left TVA and the right fusiform gyrus in the blind group (Figure 2D, a), it appears that early visual deprivation affects these regions at the network level. This, however, does not exclude the possibility that the VOTC supports identification of auditory objects in general—vocal and nonvocal—in the blind group. For instance, in a recent fMRI study, Hölig et al. (2014) reported a voice (speaker) congruency effect in the right anterior fusiform gyrus of congenitally blind participants, such that this region may have reorganized to support person identification through the auditory modality in case of early visual deprivation (Hölig et al., 2014). However, the absence of another category of sounds prevents from concluding that this effect represents a categorical preference for voices, because a similar congruency effect could have been observed in the same region for other nonvocal sounds. In our study, selective responses to voices over scrambled voices were

found in bilateral fusiform gyri of the blind group (Figure 5), about 3 cm more posteriorly than the region reported by Hölig et al. (2014). However, responses in these regions were also significantly larger for object sounds compared with their scrambled counterpart and, if anything, significantly larger for object sounds than for voices. Future studies should further investigate whether tasks involving the extraction of speaker's identity from voices triggers enhanced fusiform activations in early blind participants when compared with other type of vocal and nonvocal processing. This may relate to the suggestion that a specific link exists between the facial and vocal neural networks for speaker's identity recognition (von Kriegstein et al., 2005) and that sensory deprivation could trigger functionally selective recruitment of the deprived system through the remaining senses (Benetti et al., 2017; Hölig et al., 2014).

Similar conclusions about a lack of cross-modal reorganization of the face processing system in early blind participants arise from a previous study that investigated patterns of response elicited during tactile exploration of face masks and man-made objects in VOTC (Pietrini et al., 2004; see also Goyal, Hansen, & Blakemore, 2006). Category-related patterns of response in VOTC were found in sighted and blind participants for man-made objects (shoes and bottles), but not for face masks (Pietrini et al., 2004). Moreover, in the sighted group, category-related patterns correlated across the visual and the tactile modality for man-made objects, but not for faces. On the basis of these observations, the authors concluded that, although objects' representations might be supramodal in the VOTC, face representations are specific to vision. Similarly, more recent studies reported overlapping responses to names of nonliving objects in the VOTC of blind and sighted participants (He et al., 2013; Peelen et al., 2013), whereas category-related responses to animals in the VOTC were only observed in the sighted group and only with visually presented material (He et al., 2013). It has thus been proposed that selectivity for nonliving stimuli is multimodal and independent of visual experience whereas selectivity for living items, particularly in the lateral fusiform gyrus, is driven by visual stimulation only (Bi et al., 2016). In this study, the lack of categorical responses to voices combined with preferential responses to objects' sound in VOTC of the blind and sighted groups are in agreement with this theoretical framework. These findings also suggest that regions supporting the representation of faces in the sighted individuals' brain do not transfer their preferential tuning to human voices in early blind participants.

This lack of plasticity of the face recognition system is in line with the high degree of specialization (domain specificity or modularity) of this system in typically developed individuals. Studies on the ontogeny of face recognition demonstrate impressive face recognition skills in newborns within a few days of birth (Johnson, Dziurawiec, Ellis, & Morton, 1991) and in monkeys raised without any

exposure to faces (Sugita, 2008). A recent study even showed a visual preference in response to face-like stimulation in human fetuses (Reid et al., 2017). Moreover, categorical neural responses to faces embedded among various nonface objects were recently identified in 4-month-old babies (de Heering & Rossion, 2015). In the nonhuman primate brain, face-responsive areas contain neurons that respond selectively to faces (Tsao, Freiwald, Tootell, & Livingstone, 2006; Desimone, 1991; Gross, Rocha-Miranda, & Bender, 1972), and these areas have been demonstrated to be strongly interconnected and isolated from the rest of the visual recognition system (Moeller, Freiwald, & Tsao, 2008). Together, these characteristics of the face recognition system could come at the expense of generalization (to other domains) and plasticity. Some researchers have proposed that the development of face recognition may be under high genetic control (Kanwisher, 2010). This assumption is supported by studies on families with hereditary prosopagnosia (Grüter, Grüter, & Carbon, 2008; Schmalzl, Palermo, & Coltheart, 2008; Duchaine, Germine, & Nakayama, 2007) and performance of monozygotic relative to dizygotic twins on a face memory task (Wilmer et al., 2010). In addition, Polk, Park, Smith, and Park (2007) found that genetics may play a larger role on neural activity patterns evoked by faces (Polk et al., 2007) compared with the ones evoked by written pseudowords (Park, Park, & Polk, 2012; Polk et al., 2007; but see Pinel et al., 2014). Hence, different functional areas in the cortex may result from different neurodevelopmental mechanisms (Kanwisher, 2010). For example, it could be that the selectivity for word strings of the visual word form area emerges through learning-dependent mechanisms (Dehaene et al., 2010; He, Liu, Jiang, Chen, & Gong, 2009) whereas selectivity for faces in the FFA arises because "the specific instructions for constructing the critical circuits for face perception are in the genome" (Kanwisher, 2010). These different developmental mechanisms for defining functional areas might interact with sensory deprivation and therefore influence and constrain the process of cross-modal plasticity. In summary, the finding of cross-modal categorical responses to objects but not voices in the occipital cortex of early blind individuals suggests that cross-modal compensation in the case of early visual deprivation depends on the neural systems investigated and on the neurodevelopmental mechanisms that underlie the emergence of these systems.

## Acknowledgments

This work was supported by the Canada Research Chair Program (F. L.), the Canadian Institutes of Health Research (F. L.), the Belgian National Fund for Scientific Research (G. D.), and a European Research Council starting grant (MADVIS Grant 337573) attributed to O. C.

Reprint requests should be sent to Olivier Collignon, Université catholique de Louvain, 10, Place du Cardinal Mercier, 1348 Louvain-La-Neuve, Belgium, or via e-mail: [olivier.collignon@uclouvain.be](mailto:olivier.collignon@uclouvain.be).

## REFERENCES

- Adam, R., & Noppeney, U. (2010). Prior auditory information shapes visual category-selectivity in ventral occipitotemporal cortex. *Neuroimage*, *52*, 1592–1602.
- Amedi, A., Raz, N., Azulay, H., & Malach, R. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorative Neurology and Neuroscience*, *28*, 143–156.
- Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early “visual” cortex activation correlates with superior verbal memory performance in the blind. *Nature Neuroscience*, *6*, 758–766.
- Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*, 687–689.
- Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage*, *49*, 703–711.
- Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *Neuroimage*, *21*, 91–98.
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*, 95–113.
- Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., et al. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, *29*, 529–535.
- Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nature Reviews Neuroscience*, *3*, 443–452.
- Bedny, M., Konkle, T., Pelphrey, K., Saxe, R., & Pascual-Leone, A. (2010). Sensitive period for a multimodal response in human visual motion area MT/MST. *Current Biology*, *20*, 1900–1906.
- Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 4429–4434.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research, Cognitive Brain Research*, *13*, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Benetti, S., Van Ackeren, M. J., Rabini, G., Zonca, J., Foa, V., Baruffaldi, F., et al. (2017). Functional selectivity for face processing in the temporal voice area of early deaf individuals. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, e6437–e6446.
- Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences*, *20*, 282–290.
- Blank, H., Anwander, A., & von Kriegstein, K. (2011). Direct structural connections between voice- and face-recognition areas. *Journal of Neuroscience*, *31*, 12906–12915.
- Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *Journal of Neurophysiology*, *107*, 1443–1456.
- Brefczynski-Lewis, J. A., & Lewis, J. W. (in press). Auditory object perception: A neurobiological model and prospective review. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2017.04.034.
- Büchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature*, *394*, 274–277.
- Burton, H., Diamond, J. B., & McDermott, K. B. (2003). Dissociating cortical regions activated by semantic and phonological tasks: A fMRI study in blind and sighted people. *Journal of Neurophysiology*, *90*, 1965–1982.
- Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., et al. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain*, *136*, 2769–2783.
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135.
- Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cerebral Cortex*, *17*, 457–465.
- Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., et al. (2011). Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 4435–4440.
- Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, *192*, 343–358.
- de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, *4*, e06564.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*, 1359–1364.
- Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*, 1–8.
- 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.
- Dormal, G., & Collignon, O. (2011). Functional selectivity in sensory-deprived cortices. *Journal of Neurophysiology*, *105*, 2627–2630.
- Dormal, G., Lepore, F., & Collignon, O. (2012). Plasticity of the dorsal “spatial” stream in visually deprived individuals. *Neural Plasticity*, *2012*, 687659.
- Dormal, G., Rezk, M., Yakobov, E., Lepore, F., & Collignon, O. (2016). Auditory motion in the sighted and blind: Early visual deprivation triggers a large-scale imbalance between auditory and “visual” brain regions. *Neuroimage*, *134*, 630–644.
- Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*, 419–430.
- Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage*, *47*, 1778–1791.
- Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, *33*, 10552–10558.
- Frasnelli, J., Collignon, O., Voss, P., & Lepore, F. (2011). Crossmodal plasticity in sensory loss. *Progress in Brain Research*, *191*, 233–249.
- 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.
- Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiological interactions

- in fMRI: The importance of hemodynamic deconvolution. *Neuroimage*, *19*, 200–207.
- Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *The Journal of Neuroscience*, *26*, 6523–6532.
- Goll, J. C., Crutch, S. J., & Warren, J. D. (2010). Central auditory disorders: Toward a neuropsychology of auditory objects. *Current Opinion in Neurology*, *23*, 617–627.
- Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *The Journal of Neuroscience*, *25*, 8010–8016.
- Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: A functional magnetic resonance imaging study. *Neuropsychologia*, *47*, 2967–2974.
- Goyal, M. S., Hansen, P. J., & Blakemore, C. B. (2006). Tactile perception recruits functionally related visual areas in the late-blind. *NeuroReport*, *17*, 1381–1384.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.
- Grüter, T., Grüter, M., & Carbon, C.-C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *Journal of Neuropsychology*, *2*, 79–97.
- He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage*, *79*, 1–9.
- He, S., Liu, H., Jiang, Y., Chen, C., & Gong, Q. (2009). Transforming a left lateral fusiform region into VWFA through training in illiterate adults. *Journal of Vision*, *9*, 853.
- Heimler, B., Striem-Amit, E., & Amedi, A. (2015). Origins of task-specific sensory-independent organization in the visual and auditory brain: Neuroscience evidence, open questions and clinical implications. *Current Opinion in Neurobiology*, *35*, 169–177.
- Hölig, C., Föcker, J., Best, A., Röder, B., & Büchel, C. (2014). Brain systems mediating voice identity processing in blind humans. *Human Brain Mapping*, *35*, 4607–4619.
- Jiang, J., Zhu, W., Shi, F., Liu, Y., Li, J., & Qin, W. (2009). Thick visual cortex in the early blind. *Journal of Neuroscience*, *29*, 2205–2211.
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19.
- Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences, U.S.A.*, *107*, 11163–11170.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kassuba, T., Klinge, C., Hölig, C., Menz, M. M., Ptito, M., Röder, B., et al. (2011). The left fusiform gyrus hosts trisensory representations of manipulable objects. *Neuroimage*, *56*, 1566–1577.
- Lacey, S., Flueckiger, P., Stilla, R., Lava, M., & Sathian, K. (2010). Object familiarity modulates the relationship between visual object imagery and haptic shape perception. *Neuroimage*, *49*, 1977–1990.
- Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y.-F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14*, 420–429.
- 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.
- Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience*, *25*, 5148–5158.
- Lewis, J. W., Frum, C., Brefczynski-Lewis, J. A., Talkington, W. J., Walker, N. A., Rapuano, K. M., et al. (2011). Cortical network differences in the sighted versus early blind for recognition of human-produced action sounds. *Human Brain Mapping*, *32*, 2241–2255.
- Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of Cognitive Neuroscience*, *23*, 2079–2101.
- Lewis, J. W., Talkington, W. J., Tallaksen, K. C., & Frum, C. A. (2012). Auditory object salience: Human cortical processing of non-biological action sounds and their acoustic signal attributes. *Frontiers in Systems Neuroscience*, *6*, 27.
- Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., et al. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience*, *29*, 2283–2296.
- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, *14*, 1008–1021.
- Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron*, *63*, 397–405.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 8135–8139.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.
- Mathias, S. R., & von Kriegstein, K. (2014). How do we recognise who is speaking? *Frontiers in Bioscience (Scholar Edition)*, *6*, 92–109.
- Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe. *Science*, *320*, 1355–1359.
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–660.
- Noppeney, U., Friston, K. J., & Price, C. J. (2003). Effects of visual deprivation on the organization of the semantic system. *Brain*, *126*, 1620–1627.
- O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*, 604–609.
- Pan, W.-J., Wu, G., Li, C.-X., Lin, F., Sun, J., & Lei, H. (2007). Progressive atrophy in the optic pathway and visual cortex of early blind Chinese adults: A voxel-based morphometry magnetic resonance imaging study. *Neuroimage*, *37*, 212–220.
- Park, H.-J., Lee, J. D., Kim, E. Y., Park, B., Oh, M.-K., Lee, S., et al. (2009). Morphological alterations in the congenital blind based on the analysis of cortical thickness and surface area. *Neuroimage*, *47*, 98–106.
- Park, J., Park, D. C., & Polk, T. A. (2012). Investigating unique environmental contributions to the neural representation

- of written words: A monozygotic twin study. *PLoS One*, 7, e31512.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.
- Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool selectivity in left occipitotemporal cortex develops without vision. *Journal of Cognitive Neuroscience*, 25, 1225–1234.
- Perlovsky, L. (2011). Language and cognition interaction: Neural mechanisms. *Computational Intelligence and Neuroscience*, 2011, 454587.
- Petrides, M. (2012). *The human cerebral cortex: An MRI atlas of the sulci and gyri in MNI stereotaxic space*. New York: Academic Press.
- Pietrini, P., Furey, M. L., Ricciardi, E., Gobbin, M. I., Wu, W.-H. C., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences, U.S.A.*, 101, 5658–5663.
- Pinel, P., Lalanne, C., Bourgeron, T., Fauchereau, F., Poupon, C., Artiges, E., et al. (2014). Genetic and environmental influences on the visual word form and fusiform face areas. *Cerebral Cortex*, 25, 2478–2493.
- Polk, T. A., Park, J., Smith, M. R., & Park, D. C. (2007). Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *Journal of Neuroscience*, 27, 13921–13925.
- Reich, L., Maidenbaum, S., & Amedi, A. (2012). The brain as a flexible task machine: Implications for visual rehabilitation using noninvasive vs. invasive approaches. *Current Opinion in Neurology*, 25, 86–95.
- Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology*, 21, 363–368.
- Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology*, 27, 1825–1828.
- Ricciardi, E., Handjaras, G., & Pietrini, P. (2014). The blind brain: How (lack of) vision shapes the morphological and functional architecture of the human brain. *Experimental Biology and Medicine (Maywood, N.J.)*, 239, 1414–1420.
- Ricciardi, E., & Pietrini, P. (2011). New light from the dark: What blindness can teach us about brain function. *Current Opinion in Neurology*, 24, 357–363.
- Ricciardi, E., Vanello, N., Sani, L., Gentili, C., Scilingo, E. P., Landini, L., et al. (2007). The effect of visual experience on the development of functional architecture in hMT+. *Cerebral Cortex*, 17, 2933–2939.
- Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience*, 16, 930–936.
- Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, 79, 138–157.
- Sadato, N., Okada, T., Honda, M., & Yonekura, Y. (2002). Critical period for cross-modal plasticity in blind humans: A functional MRI study. *Neuroimage*, 16, 389–400.
- Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., et al. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, 380, 526–528.
- Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, 28, 259–287.
- Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology*, 2, 99–117.
- Sharp, D. J., Scott, S. K., & Wise, R. J. S. (2004). Retrieving meaning after temporal lobe infarction: The role of the basal language area. *Annals of Neurology*, 56, 836–846.
- Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *Neuroimage*, 58, 588–594.
- Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences, U.S.A.*, 105, 394–398.
- Tranel, D., Grabowski, T. J., Lyon, J., & Damasio, H. (2005). Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *Journal of Cognitive Neuroscience*, 17, 1293–1305.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311, 670–674.
- van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology*, 4, 388.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, 17, 367–376.
- Voss, P., & Zatorre, R. J. (2012). Organization and reorganization of sensory-deprived cortex. *Current Biology*, 22, R168–R173.
- Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How visual is the visual cortex? Comparing connective and functional fingerprints between congenitally blind and sighted individuals. *Journal of Neuroscience*, 35, 12545–12559.
- Watkins, K. E., Shakespeare, T. J., O'Donoghue, M. C., Alexander, I., Ragge, N., Cowey, A., et al. (2013). Early auditory processing in area V5/MT+ of the congenitally blind brain. *Journal of Neuroscience*, 33, 18242–18246.
- Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., et al. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *Journal of Neuroscience*, 20, 2664–2672.
- Wheatley, T., Weisberg, J., Beauchamp, M. S., & Martin, A. (2005). Automatic priming of semantically related words reduces activity in the fusiform gyrus. *Journal of Cognitive Neuroscience*, 17, 1871–1885.
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences, U.S.A.*, 107, 5238–5241.
- Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, 17, 263–271.