OXFORD

Cerebral Cortex, 2022, **32**, 4913–4933
https://doi.org/10.1093/cercor/bhab524
Advance access publication date 21 January 2022
Original Article

Preference for animate domain sounds in the fusiform gyrus of blind individuals is modulated by shape–action mapping

Łukasz Bola [1,2,†,*], Huichao Yang[3,4,5,†], Alfonso Caramazza[1,6], Yanchao Bi[3,4,7,*]

[1]Department of Psychology, Harvard University, Cambridge, MA 02138, USA,
[2]Institute of Psychology, Polish Academy of Sciences, Warsaw 00-378, Poland,
[3]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China,
[4]Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China,
[5]School of Systems Science, Beijing Normal University, Beijing 100875, China,
[6]Center for Mind/Brain Sciences (CIMeC), University of Trento, Trento 38122, Italy,
[7]Chinese Institute for Brain Research, Beijing 102206, China

*Address correspondence to Łukasz Bola, Institute of Psychology, Polish Academy of Sciences, 1 Jaracza Street, Warsaw 00-378, Poland. Email: lbola@psych.pan.pl and Yanchao Bi, National Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China. Email: ybi@bnu.edu.cn

†Łukasz Bola and Huichao Yang share first authorship of this work

In high-level visual areas in the human brain, preference for inanimate objects is observed regardless of stimulation modality (visual/auditory/tactile) and individual's visual experience (sighted/blind) whereas preference for animate entities seems robust mainly in the visual modality. Here, we test a hypothesis explaining this domain difference: Object representations can be activated through nonvisual stimulation when their shapes are systematically related to action system representations, a quality typical of most inanimate objects but of only specific animate entities. We studied functional magnetic resonance imaging activations in congenitally blind and sighted individuals listening to animal, object, and human sounds. In blind individuals, the typical location of the fusiform face area preferentially responded to human facial expression sounds clearly related to specific facial actions and resulting face shapes but not to speech or animal sounds. No univariate preference for any sound category was observed in the fusiform gyrus in sighted individuals, but the expected multivoxel effects were present. We conclude that nonvisual signals can activate shape representations of those stimuli—inanimate or animate—for which shape and action computations are transparently related. However, absence of potentially competing visual inputs seems necessary for this effect to be clearly detectable in the case of animate representation.

Key words: face perception; facial expressions; fusiform face area; shape; ventral occipitotemporal cortex.

Introduction

The human occipitotemporal cortex (OTC) is a high-level visual region that hosts areas preferentially responding to stimuli from specific domains, such as manipulable inanimate objects that can be used as tools, stable inanimate objects relevant for navigation, and animate entities relevant for social interactions (Konkle and Caramazza 2013; Magri et al. 2021). Lesions in this region result in a variety of visual agnosias—an inability to recognize objects visually, often coupled with relatively unimpaired low-level vision, object knowledge, and object recognition through other senses (e.g., Goodale et al. 1994; Wada and Yamamoto 2001; James et al. 2003). These observations suggest that the OTC is critical for the processing of visual shape (Goodale and Milner 1992). The existence of shape representation in the OTC was further confirmed in a number of functional magnetic resonance imaging (fMRI) studies (Kourtzi and Kanwisher 2000; Haushofer et al. 2008; Peelen et al. 2014).

Interestingly, studies have shown that the OTC responds not only to visual stimulation but also to certain kinds of auditory or tactile stimulation (e.g., Amedi et al. 2001; Amedi et al. 2004; Pietrini et al. 2004; Wolbers et al. 2011; He et al. 2013; Handjaras et al. 2016, 2017; Mattioni et al. 2020; Murty et al. 2020). A number of these studies have suggested that nonvisual responsiveness in the OTC is not uniform across stimulus types and OTC areas; instead, these studies have hinted at the existence of domain by sensory modality interaction in the OTC. That is, functional preference for inanimate objects in this region has been repeatedly observed regardless of stimulation modality (visual/auditory/tactile) and individual's visual experience (sighted/blind) (e.g., Mahon et al. 2009;

Received: February 2, 2021. **Revised:** December 16, 2021. **Accepted:** December 17, 2021

Wolbers et al. 2011; He et al. 2013; Peelen et al. 2013; Dormal et al. 2018); in contrast, functional preference for animate entities seems considerably more tied to the visual modality, with an approximately equal number of nonvisual studies reporting null results (Goyal et al. 2006; He et al. 2013; Kitada et al. 2013; Fairhall et al. 2014; Plaza et al. 2015; Dormal et al. 2018) and positive findings for this domain (Pietrini et al. 2004; Kitada et al. 2009; Handjaras et al. 2016, 2017; Fairhall et al. 2017; van den Hurk et al. 2017; Mattioni et al. 2020; Murty et al. 2020). Recently, it has been proposed that this difference across domains might reflect a difference in mapping between visual shape and action representations for inanimate and animate objects (Bi et al. 2016). This conjecture originates from a theoretical framework that assumes that the domain organization in the OTC is the result of evolutionary pressures to accommodate efficient mapping between locally computed shape representations and appropriate downstream computations (Mahon et al. 2007; Mahon and Caramazza 2011). In the case of inanimate objects, mapping between shape or texture and appropriate action is usually systematic and transparent (e.g., elongated shape and action of gripping). This strong link might make OTC representations of inanimate objects more accessible through nonvisual means—for example, one can expect that the recognition of an inanimate object through audition or touch is followed by the activation of appropriate action representations that, in turn, activate specific shape representation in the OTC "inanimate areas." In the case of animate entities, the link between shape or texture and action representations is usually much less articulated. Lack of a systematic relationship between shape and action representations may have resulted in OTC representations of animate entities being generally less accessible through nonvisual means—one can assume that the recognition of animate entities and activation of appropriate action representations (or other types of downstream representations, such as a person's identity) do not necessarily result in the activation of shape representations in the OTC "animate areas."

Can the shape–action mapping principle be used to specify types of OTC animate object representations that are actually reliably accessible through nonvisual means? In this study, we address this question by investigating responses of the fusiform face area (FFA; Kanwisher et al. 1997; Kanwisher and Yovel 2006) to face information conveyed by sounds. Several previous studies have already investigated whether the FFA shows functional preference for information about the human face in the absence of actual, visual face stimuli. When sighted individuals are asked to imagine faces that they know or that were shown to them before the experiment, the preferential response of the FFA, relative to other imagery conditions, has usually been found (Ishai et al. 2000; O'Craven and Kanwisher 2000; Goyal et al. 2006). However, surprisingly mixed results have been observed in sighted individuals not engaged in explicit imagery tasks and in congenitally and early blind participants. Thus, no preferential FFA activation was reported when sighted and early blind participants were listening to human voices pronouncing vowels (Dormal et al. 2018) and when sighted and congenitally blind participants were perceiving faces through vision-to-auditory sensory substitution devices (Plaza et al. 2015; no effects in the right hemisphere and a positive effect for blind participants in the left counterpart of the FFA), or when blind participants were touching 3D models of faces (Goyal et al. 2006; a positive effect was found for late blind participants; Kitada et al. 2013). Furthermore, a null result for the FFA was reported when sighted participants were making semantic decisions about famous people based on a word cue (Fairhall et al. 2014). In apparent contrast to these reports, other studies showed preferential activation of the FFA when sighted and congenitally blind participants were tactually identifying and comparing 3D models of faces (Kitada et al. 2009; Murty et al. 2020). In the auditory modality, preferential FFA response was found in congenitally blind participants listening to sounds of human facial actions (van den Hurk et al. 2017; repliacted in Murty et al. 2020) or to verbal phrases conveying different emotional states (Fairhall et al. 2017). Although differences between the tactile studies might perhaps be explained by differences in statistical power (a small participant group in Goyal et al. 2006 where an effect in the FFA of congenitally blind and sighted individuals was not reported; a very subtle effect in the visual counterpart of the tactile experiment in Kitada et al. 2013), this methodological argument is not easily applicable to other studies discussed above.

Here, we propose an explanation of these apparently conflicting results. Specifically, in an extension of the shape–action mapping conjecture, we hypothesize that FFA responses to nonvisual stimulation is modulated by the relationship between face shape and facial motor action representations. This view predicts that dynamic representations of face shapes, which are systematically and transparently related teTJ -s7-418(a)-566(models)84a193(r)32(P

Harry et al. 2013; Bernstein and Yovel 2015). Furthermore, previous auditory studies reported functional preference of the typical location of the FFA in blind participants for sounds of facial actions (van den Hurk et al. 2017) or phrases being indicative of the emotional state of a speaker (Fairhall et al. 2017). In contrast, the nonvisual studies which did not observe functional preference of the FFA for face-related information used stimuli for which a relationship between action and shape representations might be less clear, that is, speech sounds produced in a neutral tone (Dormal et al. 2018) or names of famous people (Fairhall et al. 2014).

To experimentally test our hypothesis, we designed an fMRI study in which congenitally blind and sighted individuals listened to object sounds and four categories of sounds produced by animate entities. The first two animate categories included human sounds associated with either emotional or nonemotional facial expressions (sounds of laughing and crying vs. sounds of yawning and sneezing). The other two animate categories included speech sounds produced in a neutral tone and animal sounds. Based on our hypothesis, we expected to find that the FFA shows functional preference for both emotional and nonemotional facial expression sounds—a result that would allow us to distinguish between the FFA sensitivity to specific types of face-related information and sensitivity to emotional content, particularly in blind individuals (e.g., Fairhall et al. 2017). In contrast, we did not expect to observe robust functional preference, neither in the FFA nor in the other ventral OTC (vOTC) animate areas, for speech and animal sounds.

Materials and Methods

Participants

Twenty congenitally blind individuals (13 male, 7 female, mean age ± SD = 46.25 ± 13.34 years, average length of education ± SD = 9.4 ± 3.9 years) and 22 sighted individuals (15 male, 7 female, mean age ± SD = 45.91 ± 10.41 years, average length of education ± SD = 9.41 ± 2.09 years) participated in the study. The groups were matched for age, gender, handedness, and years of education (all $P > 0.25$). All participants were native Mandarin speakers and had no history of neurological disorders. All sighted participants had normal or corrected-to-normal vision. Detailed characteristics of each blind participant are provided in Table 1. All experimental protocols were approved by the institutional review board of the Department of Psychology Peking University, China (4 May 2015) of Harvard University (IRB15-2149) in accordance with the Declaration of Helsinki. Informed consent was obtained from each participant prior to the study.

Functional Magnetic Resonance Imaging Experiment

In the fMRI experiment, participants were presented with sounds belonging to five categories: emotional facial expressions (laughing and crying), nonemotional facial expressions (sneezing and yawning), speech sounds (two pinyin syllables—"bo" and "de"; the syllables conveyed no clear semantic information and were repeated three times in a neutral tone), object sounds (church bells, sleigh bells, sound of a car driving, sound of traffic) and animal sounds (a dog, a horse, a rooster and a cow). Each human sound was produced by two actors (man and woman). Furthermore, each sound subcategory

Table 1. Characteristics of blind participants

Subject	Age	Gender	Years of education	Handedness	Cause of blindness	Light perception
CB1	47	F	12	R	Congenital glaucoma	None
CB2	30	M	0	R	Congenital anophthalmia	None
CB3	40	M	12	L	Congenital microphthalmia	None
CB4	54	F	12	R	Congenital cataracts and eyeball dysplasia	Faint
CB5	52	F	9	R	Congenital glaucoma	None
CB6	61	M	12	R	Congenital eyeball dysplasia	None
CB7	67	M	9	R	Congenital glaucoma and leukoma	None
CB8	52	M	12	L	Congenital microphthalmia	None
CB9	44	M	12	R	Congenital leukoma	Faint
CB10	52	M	12	R	Congenital glaucoma	None
CB11	27	M	12	R	Congenital glaucoma	None
CB12	52	M	12	L	Congenital microphthalmia, cataracts and leukoma	None
CB13	34	F	12	R	Congenital microphthalmia	None
CB14	28	M	0	R	Congenital dysplasia, amotio retinae, aphakia and cataract	Faint
CB15	60	F	9	R	Congenital microphthalmia and anophthalmia	None
CB16	59	F	9	R	Congenital eye tumor disease and dysplasia	None
CB17	54	F	9	R	Congenital microphthalmia	None
CB18	59	M	9	R	Congenital anophthalmia and cataract	None
CB19	23	M	11	R	Unknown	None
CB20	30	M	3	R	Congenital glaucoma and cataract	None

a lesser extent than in the case of facial expressions (i.e., in the case of speech sounds, this mapping is present only for the mouth region). In summary, speech sounds (in combination with nonemotional facial expression sounds, in the case of the emotional dimension) control for most of the potentially confounding dimensions in our study in a better way than continuous vocalizations.

The timing of the sound presentation in the fMRI scanner was kept the same across conditions and participant groups. Stimuli were presented in short blocks that included three sound exemplars belonging to the same condition and subcategory (e.g., three exemplars of church bell sounds). Presentation of each exemplar was followed by a 500 ms of silence period resulting in a total duration of each block equal to 7.5 s. The order of specific sound exemplar presentation within each block was randomized in order to modulate purely auditory properties across blocks belonging to a given subcategory. The participants' task was to respond to each sound and answer whether it was produced by a human or nonhuman entity. Our aim was to use a relatively simple task which would nevertheless focus the participants' attention on categorical distinctions present in the experiment. We reasoned that this would allow us, first, to control for participants' alertness in the scanner and second, to boost the magnitude of responses in the category-preferring areas in the OTC. Critically, if the task demands drove the reported results, one would expect to find activation differences primarily between

[illegible] forawhches20 20 acti2ss differ primari2ss

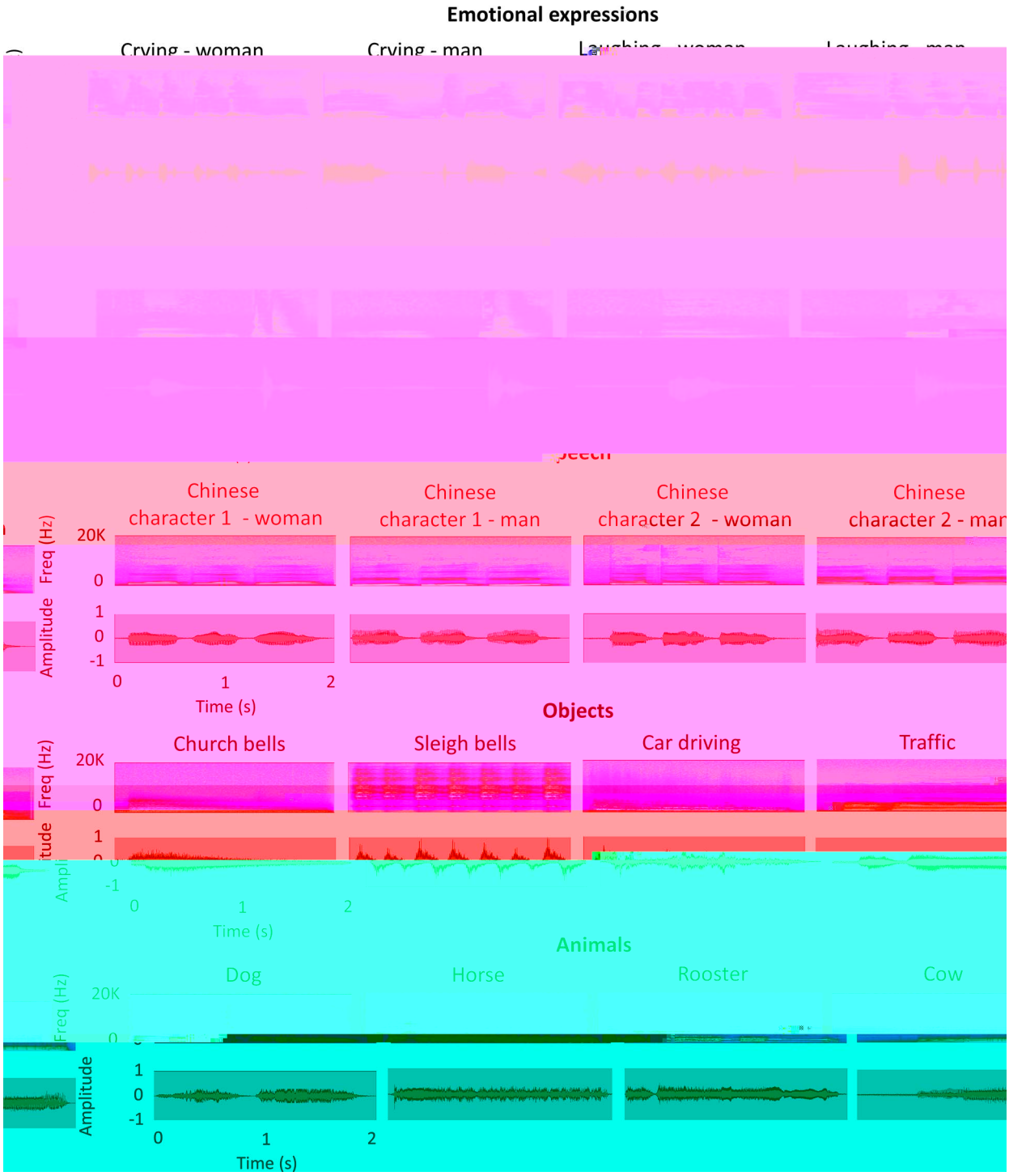

Figure 1. Spectrograms (upper panels) and waveforms (lower panels) of sounds used in the study. Each sound category included four subcategories, which, in turn, included three unique sound exemplars (e.g., three exemplars of a male actor laughing or three different exemplars of a horse neighing). The figure presents a representative exemplar of a sound from each subcategory. The spectrograms and the waveforms were produced using audacity (https://www.audacityteam.org/).

Behavioral Rating Experiment

To verify whether the facial expression sounds induce a stronger impression of the facial motor actions than the speech sounds, we performed an additional behavioral experiment on an independent group of 20 sighted participants (8 males, mean age ± SD = 22.2 ± 2.33). The participants listened to the facial expression sounds and the speech sounds used in the fMRI study and rated on a seven-point scale if those sounds induce a clear image of the face movement in their minds.

Magnetic Resonance Imaging Data Acquisition

Data were acquired on a Siemens Prisma 3-T scanner with a 20-channel phase-array head coil at the Imaging

Center for MRI Research, Peking University. Functional data were acquired using simultaneous multislice sequence supplied by Siemens: slice planes scanned along the rectal gyrus, 62 slices, phase encoding direction from posterior to anterior; 2 mm thickness; 0.2 mm gap; multiband factor = 2; TR = 2000 ms; TE = 30 ms; FA = 90°; matrix size = 112 × 112; FOV = 224 × 224 mm; voxel size = 2 × 2 × 2 mm. Right before the start of the collection of the first functional run, MRI field mapping was performed using the phase–magnitude sequence with the following parameters: slice planes scanned along the rectal gyrus, 62 slices, phase encoding direction from posterior to anterior; 2 mm thickness; 0.2 mm gap; TR = 620 ms; TE^1 = 4.92 ms; TE^2 = 7.38 ms FA = 60°; matrix size = 112 × 112; FOV = 224 × 224 mm; voxel size = 2 × 2 × 2 mm. T1-weighted anatomical images were acquired using a 3D MPRAGE sequence: 192 sagittal slices; 1 mm thickness; TR = 2530 ms; TE = 2.98 ms; inversion time = 1100 ms; FA = 7°; FOV = 256 × 224 mm; voxel size = 0.5 × 0.5 × 1 mm; matrix size = 512 × 448.

Behavioral Data Analysis

Accuracy obtained from the categorization task performed in the MRI scanner were entered into the group (blind participants, sighted participants) × sound category (emotional facial expressions, nonemotional facial expressions, speech, objects, animals) ANOVA. Reaction time was not analyzed as 1) the primary goal of the behavioral task was to verify whether the participants were attentive in the scanner, for which the accuracy is a better measure; and 2) the time needed to categorize a given sound as belonging to a specific category is likely to depend not only on a participant's level of effort or mental processes, but also on characteristics of the sound itself (e.g., its temporal dynamics). Thus, the reaction time is a rather uninformative measure in the context of our experiment.

The data from the additional behavioral rating experiment were entered into a repeated-measures ANOVA with sound category (emotional facial expressions, nonemotional facial expressions, speech sounds) as a factor. The pair-wise comparisons were corrected for multiple comparisons using FDR. All behavioral analyses were performed in SPSS 25 (IBM).

Magnetic Resonance Imaging Data Analysis

Data Preprocessing

Before the actual preprocessing, the MRI data were converted to NIfTI format using dcm2niix (https://github.com/rordenlab/dcm2niix). Furthermore, anatomical images were deidentified using the mri_deface script (https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface; Bischoff-Grethe et al. 2007) The preprocessing was performed using the SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) and CONN 18b toolbox (www.nitrc.org/projects/conn; Whitfield-Gabrieli and Nieto-Castanon 2012) running on MATLAB R2018b (MathWorks). Data from each participant were preprocessed using a standard CONN pipeline for volume-based analyses, which includes the following steps: 1) realignment of all functional images to the first collected image; 2) distortion correction and unwarping of functional images using a voxel-displacement map, created based on the field mapping sequence; 3) slice-timing correction of functional images; 4) detection of outliers in functional time series using the functions of the Artifact Detection Tools toolbox (standard CONN "conservative" settings were used: all volumes showing, relative to the previous image, global signal change of $Z = 3$ or participant's head motion larger than 0.5 mm were marked as an outlier); 5) coregistration of the anatomical image to the mean functional image; 6) segmentation of the coregistered anatomical image and its normalization to Montreal Neurological Institute (MNI) space; 7) normalization of all functional images to MNI space, using a deformation field obtained for the anatomical image; and 8) spatial smoothing with Gaussian kernel at two levels: 8 mm full width at half maximum (FWHM) for the univariate analysis and 2 mm FWHM for the multivariate analyses (see Gardumi et al. 2016 for evidence that mild-to-moderate smoothing might improve the sensitivity of the decoding analysis).

Two first-level statistical models were created for each participant. For the univariate analysis, the data smoothed with 8 mm FWHM kernel were modeled at the level of sound categories (five experimental conditions: emotional expression sounds, nonemotional expression sounds, speech sounds, object sounds, animal sounds). For the multivoxel pattern analysis (MVPA), the data smoothed with 2 mm FWHM kernel were modeled at the level of specific subcategories (20 experimental conditions: five categories × four subcategories). In both models, the signal time course was modeled within a general linear model (Friston et al. 1995) derived by convolving a canonical hemodynamic response function with the time series of the experimental conditions. Additionally, six movement parameter regressors and a separate regressor for each volume identified as an outlier in the preprocessing (see above) were added to both models. An inclusive high-pass filter was used (400 s) to remove drifts from the signal while ensuring that effects specific to each subcategory are not filtered out from the data (note that only two blocks per subcategory were presented in each run and the order of presentation was randomized). Finally, individual contrast images (i.e., images containing the output of the subtraction of beta estimates for the compared conditions) were computed for each experimental condition (i.e., each category in the first model, each subcategory in the second model), relative to rest periods.

Univariate Analysis

For the univariate whole-brain analysis, the contrast images for five sound categories versus rest periods, for both groups, were entered into a random-effect

2 (group) × 5 (sound category) ANOVA model, created using the statistical parametric mapping (SPM) "flexible factorial" functionality. This model was used to perform all whole-brain univariate analyses reported, except for testing for the main effect of the group—as error terms calculated within the SPM flexible factorial design are generally not suitable for testing for main effects of the group, this analysis was performed separately by entering individual average activations for all sound categories versus rest periods into a simple two-group model (the SPM "two-sample *t*-test" functionality). *F*-tests were used to test for main effects of group and sound category and for the omnibus interaction between these factors. *t*-Tests were used to perform a detailed pair-wise comparisons. For each comparison, the group *F*-stat or *t*-stat map was converted into a threshold-free cluster enhancement (TFCE) map (Smith and Nichols 2009). TFCE values are voxel-wise statistics alternative to the classic statistics. An important advantage of using the TFCE values in second-level analyses is that this measure takes into account not only the strength of an effect at a given voxel but also the extent of a given cluster of activation, reflecting a reasonable assumption that the fMRI activation is more likely to be a true positive when it is strong and when it is clustered. Thus, using TFCE values in a second-level analysis is a way to keep increased sensitivity of cluster-based thresholding while avoiding problems encountered by cluster-based methods, such as the need for setting the arbitrary cluster-forming threshold or lower sensitivity to focal effects (Smith and Nichols 2009). In our analysis, the TFCE values were calculated with standard parameters ($E = 0.5$, $H = 2$; see Smith and Nichols 2009) for each comparison, for every voxel within a broad gray matter mask, created based on canonical brain tissue probability maps implemented in SPM (all voxels in the brain image with a 20% or higher probability of belonging to the gray matter were included). The obtained TFCE values were then converted into voxel-wise *P*-values using permutation testing. Specifically, for each voxel included in the analysis, the TFCE value obtained for the actual comparison was compared with the null distribution of 10 000 TFCE values obtained in the comparisons with labels of compared conditions and/or compared groups randomized (thus, minimal uncorrected *P*-value that could be obtained = 0.0001). The resulting maps of *P*-values were corrected across all voxels included in the analysis using the false discovery rate (FDR; Benjamini and Hochberg 1995). The calculation of the TFCE-values and *P*-values was performed using TFCE SPM toolbox (v. 1.77; http://www.neuro.uni-jena.de/tfce/). The results of group comparisons were visualized on a standard MNI152 template using MRIcroGL (https://www.nitrc.org/projects/mricrogl).

To visualize regions showing functional preference for both types of facial expression sounds relative to all control sound categories, we performed a conjunction (i.e., logical AND) analysis. We implemented the analysis by intersecting six fully corrected (as discussed) simple contrasts between the facial expression sounds and the control conditions (two facial expression sound categories × three control sound categories) with each other. In other words, the voxels highlighted in the conjunction analysis had to be significant, at the fully corrected level, in every simple contrast included in the analysis. To visualize local peak maxima of the overall difference between the facial expression sounds and the control sound categories, we overlaid the mean FDR-corrected *P*-values of the six contrasts included on the conjunction maps. However, those values were calculated only for voxels that passed the conjunction test described above.

The univariate region-of-interest (ROI) analysis was performed 1) to confirm the spatial correspondence between the effects observed in the fusiform gyrus in the blind participants and the typical anatomical location of the FFA; and 2) to further investigate the similarity and differences of activation patterns in the fusiform gyrus in the blind group and the activation patterns in the auditory areas and other face processing areas. Toward these aims, the FFA and the control ROIs—the primary auditory cortex (A1), the superior temporal sulcus (STS), the occipital face area (OFA; Gauthier et al. 2000), and the part of the right posterior superior temporal sulcus responsive to facial expressions (right pSTS; Allison et al. 2000; Andrews and Ewbank 2004; Pitcher et al. 2014)—were defined using a statistical association map produced by an automatic Neurosynth meta-analyses (v. 0.7; https://neurosynth.org/). The logic behind Neurosynth association maps is to search the Neurosynth database and highlight brain areas that are reported as activated in articles using a target term significantly more often than in articles that do not contain this term (Yarkoni et al. 2011). Thus, using Neurosynth-defined ROIs allowed us to quantify on a large body of face perception and auditory studies, abstract away from various quirks of individual experiments, and consequently, better visualize the typical location of the FFA and other ROIs. The FFA ROI and the OFA ROI were defined using a target term "face." The meta-analysis included 896 studies and the association map was thresholded at *Z*-value = 15 for the FFA ROI definition and *Z*-value = 11 for the OFA ROI definition. This procedure resulted in 1) a cluster in the right fusiform gyrus consisting of 227 voxels (the FFA ROI; functional data resolution; cluster center of mass MNI coordinates: 41 −49 −20); and a cluster in the inferior occipital gyrus consisting of 145 voxels (the OFA ROI; cluster center of mass MNI coordinates: 43 −77 −12). The A1, STS, and right pSTS ROIs were defined using target terms "primary auditory" (114 studies; $Z = 10$; 672 voxels), "STS" (203 studies; $Z = 7$; 2234 voxels), and "facial expressions" (250 studies; $Z = 4.8$; 233 voxels), respectively. MarsBaR SPM toolbox (http://marsbar.sourceforge.net/) was used to extract an average contrast estimate for all voxels in a given ROI for each participant and experimental condition. This procedure resulted in five values for each ROI and each

participant, representing the contrast estimates for each condition compared with rest periods (i.e., positive values represented above-rest fMRI activation level and negative values represented below-rest fMRI activation level). Analogously to the univariate whole-brain analysis, those values were then entered into a random-effect 2 (group) × 5 (sound category) ANOVA model, which was created in SPSS 25 (IBM). The results of pairwise comparisons were corrected for multiple comparisons using FDR. The FDR correction was calculated separately for the three analysis steps—that is, for pairwise comparisons within each group (10 comparisons across sound categories per group) as well as for pairwise comparisons between groups (five comparisons). The results were visualized with error bars representing standard error of the mean, adjusted to adequately represent within-participant variability across conditions using a method proposed by Cousineau (2005).

To test whether the activation in the typical location of the FFA in the blind group is merely a duplicate of the activation of the right pSTS or right frontal regions, which also showed the functional preference for facial expression sounds, the above-described ANOVA for the FFA ROI was repeated with either frontal or right pSTS activations for all sound categories as covariates. The five values per participant included as covariates (each sound category > rest) were extracted from 8 mm spheres centered on a peak of either frontal or right pSTS effect in all facial expression sounds greater than all other sound categories contrast. In other words, covariates were extracted from areas showing the strongest functional preference for facial expression sounds. The search for the above-described peaks was performed in the right frontal and the right temporal cluster identified as significant in the whole-brain conjunction analysis testing for functional preference for facial expression sounds, relative to all other sound categories (see the description above). This procedure resulted in the following ROI centers in the MNI coordinates: 46 −38 −12 for the right pSTS and 48 −12 −22 for the frontal ROI.

To further investigate the pattern of functional preference for specific sound categories in the OTC, winner-take-all maps were calculated for this region and for each group, following the convention described in recent studies (van den Hurk et al. 2017; Mattioni et al. 2020). We used the broad OTC masks created for the blind and for the sighted participants by Mattioni et al. (2020) as our ROIs for the blind and the sighted group, respectively. Within these ROIs, we assigned each voxel to a condition with the highest contrast estimate, in comparison with rest, in a given group of participants. The contrast estimate values were extracted from the group ANOVA model created for the univariate analysis. This procedure resulted in one OTC map for each group, with each voxel within a map assigned to the experimental condition (one out of five) that showed the highest fMRI activation in this voxel.

Multivoxel Pattern Analysis

The MVPA included both the classification (decoding) analysis and the representational similarity analysis (RSA). Due to our a priori hypothesis about the effects of interest being present in the typical location of the FFA and the richness of our design, making the presentation of the whole-brain results not practical, the MVPA analyses were implemented in the ROI methodology, similarly to other recent and condition-rich studies (e.g., van den Hurk et al. 2017; Mattioni et al. 2020). The FFA ROI was defined in a Neurosynth meta-analysis with a target word "face," threshold at Z-value = 10 (896 studies included). This resulted in a bilateral ROI consisting of 936 functional voxels (cluster center of mass MNI coordinates, right hemisphere: 42 −57 −18; anterior cluster in the left hemisphere: −41 −50 −19; posterior cluster in the left hemisphere: −39 −84 −13). The reason behind broader, compared to the univariate analysis, and bilateral ROI definition was that the MVPA relies on the disperse and subthreshold activation and deactivation patterns, which might be well represented also by cross-talk between hemispheres (e.g., a certain subcategory might be represented by activation of the right FFA and deactivation of the left analog of this area). Performing MVPA in the FFA ROI used in the univariate analysis did not change the results in the blind group and yielded qualitatively similar, but statistically weaker results in the sighted group. The control ROIs in MVPA were the auditory cortex (A1 and STS univariate ROIs combined), the right pSTS (the Neurosynth meta-analysis with a target term "facial expressions"; 250 studies; $Z = 2$, 440 voxels), and the parahippocampal place area (PPA; Epstein and Kanwisher 1998; the Neurosynth meta-analysis with a target term "place"; 189 studies; $Z = 3$, 790 voxels).

The MVPA was performed at a single-participant level using CoSMoMVPA (v. 1.1.0; Oosterhof et al. 2016). All classification analyses were performed on single-participant and single-run contrast estimates for each sound subcategory versus rest periods (20 values—one for each sound subcategory—per run, 160 values per participant) using a support vector machine, as implemented in LIBSVM (v. 3.23; Chang and Lin 2011). Cross-validation was performed using a leave-one-run-out procedure. A standard LIBSVM data normalization procedure (i.e., Z-scoring beta estimates for each voxel in a test set and applying output values to the test set) was applied to the data before the classification. In the first analysis, we tested whether, based on the activation patterns in the FFA and in the control ROIs, the facial expression sounds could be successfully distinguished from the sounds belonging to control conditions. Toward this aim, the classification was performed pair wise for all pairs of sound categories included in the experiment. In the second analysis, we tested whether, based on the activation of the FFA and the control ROIs, one can distinguish between specific facial expressions

(laughing, crying, sneezing, and yawning). As a control analysis, we also tried to classify specific speech sounds and the gender of the actor producing the sounds. For each classification analysis performed, the classification accuracies obtained for single participants were averaged to obtain an average classification accuracy for a given analysis and a given participant group. Then, those average classification accuracies were tested against chance level in a permutation procedure. Specifically, each classification analysis was rerun 1000 times for each participant with the labels of compared conditions randomly reassigned in each iteration. Null distributions created in this way were averaged across participants, within each group, and compared with the actual average classification accuracies. The obtained *P*-values were corrected for multiple comparisons using FDR.

For the RSA, the obtained contrast estimates (see above) were averaged across sound subcategories and runs, which resulted in one contrast estimate for each of the five sound categories, for a given participant. The theoretical model of the dissimilarity of FFA response patterns to the five sound categories was created based on the FFA response patterns to visual stimuli (e.g., Kanwisher et al. 1999) and our hypothesis (the category model). The model assumes that the FFA responses to both types of facial expression sounds are the most similar to each other (face-related sounds with a clear mapping between a facial motor action and a face shape), somewhat similar to speech sounds (face-related sounds for which a mapping between a facial motor action and a face shape is decodable only for the mouth region, and therefore is less salient) and animal sounds (nonhuman animate sound), and the least similar to the object sounds (inanimate sounds). We also created three control models. Two represented divisions between either 1) human and nonhuman sounds or 2) animate and inanimate sounds, and were used as benchmarks for assessing the specificity of categorical representations in the FFA. In other words, if the typical location of the FFA truly represents face-related information, even when this information is conveyed through nonvisual modality, then one can expect the stronger correlation of neural responses in this area with our main model than with more general benchmark models. The third control model indexed average pitch similarity of sounds in each sound category and were used as a further control for the impact of low-level acoustic features on our results. The pitch was calculated in Praat (https://www.fon.hum.uva.nl/praat/) following procedure described in Mattioni et al. (2020), who showed that, compared with other low-level properties, pitch is maximally efficient in capturing encoding of sounds based on the acoustic features. Similarly to that study, we used the autocorrelation method and the "pitch floor" of 60 Hz. We obtained the pitch values for each sound used in the study and averaged those values within each sound category. Next, we created the pitch dissimilarity matrix by calculating the absolute value of a difference between each category pair. The dissimilarity values in the main model and the pitch model were not positively correlated with each other ($r = -0.24$).

For each participant, the four created models were correlated with the actual response patterns in the FFA ROI (that is, with the neural similarity matrix for the five sound categories) using Spearman's rank correlation coefficient. First, a simple correlation was performed for each model. Then, a partial correlation procedure was applied, in which each model was correlated with the fMRI activation for the FFA ROI and the three remaining models were included in the analysis as covariates. The resulting correlation values were averaged across participants in each group. The average correlation values were tested against the chance level in a permutation procedure. As in the classification analysis, the correlation was recalculated 1000 times for each participant with the condition labels in the neural similarity matrix randomized. Null distributions created in this way were averaged across participants, within each group, and compared with the mean correlation values obtained in the actual analysis. The obtained *P*-values were corrected for multiple comparisons using FDR.

Results

Behavioral Rating Experiment

When choosing sounds to be used in the fMRI experiment, we ensured that all used sounds are easily recognizable as produced by human, animal, or an artificial object, and that only sounds included in the emotional expression sound category were considered emotional; all other sound categories were considered neutral (Supplementary Table 1). Furthermore, the results of the behavioral rating experiment indicated that, as we expected, the human sounds included in the fMRI experiment induce a sense of face movements to different extents (ANOVA main effect of condition: $F(2, 38) = 19.91$, $P < 0.001$). Pair-wise comparisons showed that both emotional facial expression sounds (mean face movement rating ± SD = 5.49 ± 0.64) and nonemotional expression sounds (mean face movement rating ± SD = 5.58 ± 0.52) induce a stronger sense of face movement than speech sounds (mean face movement rating ± SD = 3.93 ± 1.57; $P_{FDR} < 0.001$ for both comparisons). In contrast, the ratings for both types of facial expression sounds were not significantly different from each other ($P_{FDR} > 0.25$).

Functional Magnetic Resonance Imaging Experiment: Behavioral Results

To keep the participants attentive in the MRI scanner, we asked them to categorize the presented stimuli into human and nonhuman sounds. Both groups of participants were highly accurate in performing this task (mean accuracy ± SD, blind participants: 95% ± 6%; sighted participants: 92% ± 10%). No difference between groups or sound categories was detected; the interaction between

these two factors was also not significant (group × sound category ANOVA, all $P > 0.18$).

Functional Magnetic Resonance Imaging Experiment: Univariate Analysis

We started the analysis of the fMRI data by performing omnibus *F*-tests, testing for main effects of sound category, group, and for interaction between these two factors. The main effect of sound category was observed in a wide network of regions, including frontoparietal regions, the temporal lobe, the OTC, and the early visual cortex (Supplementary Table 2). When we arbitrarily increased the statistical threshold, in order to explore the data and achieve a higher degree of spatial specificity, we detected the main effect of condition primarily in the auditory cortex, right frontal areas and the left cerebellum (Supplementary Fig. 1; all other analyses reported were performed with the a priori threshold, see Materials and Methods). Testing for the main effect of group did not yield significant results. A group by condition interaction was detected primarily in visual areas, including the OTC, as well as in the left and the right pSTS (Supplementary Fig. 1 and Supplementary Table 3). To disentangle effects induced in the OTC by specific sound categories, in each group, we proceeded with detailed pairwise comparisons.

Congenitally Blind Group

Animate sound categories versus object sounds

We first contrasted activation for sounds of every animate category with activation for object sounds in the blind group. As expected, sounds of both emotional and nonemotional expressions induced stronger response, relative to object sounds, in classic face processing areas—the right pSTS, the right lateral fusiform gyrus (the typical anatomical locus of the FFA) and the right inferior occipital gyrus (the typical anatomical locus of the OFA) (Fig. 2A,B; see also Supplementary Tables 4, 5). Furthermore, sounds of emotional expressions only induced stronger activation in the amygdala (Fig. 2A; peak MNI coordinates, left hemisphere = −26 −8 −20; right hemisphere = 22 −2 −24). Outside the canonical face perception network, stronger responses to both types of facial expression sounds were observed in the primary visual cortex, dorsolateral visual areas, the auditory cortex and frontal regions (primarily the precentral gyrus, the inferior frontal gyrus and the middle frontal gyrus) (Fig. 2A,B; Supplementary Tables 4, 5). Contrasting the other animate categories—that is, speech sounds and animal sounds—with object sounds yielded significant effects in the auditory cortex, the precuneus, frontal regions and, in the case of animal sounds, in some parts of the primary visual cortex (Supplementary Fig. 2; Supplementary Tables 6, 7). However, in line with our hypothesis, no reliable activation in the vOTC animate areas was observed for these animate sound categories (Supplementary Fig. 2; Supplementary Tables 6, 7).

Facial expression sounds versus other animate sound categories

We then directly compared activation produced by facial expression sounds with activation for other animate sound categories in the blind group. These comparisons confirmed that, relative to speech sounds and animal sounds, both types of the facial expression sounds induced stronger activation in the vOTC—particularly, in the typical anatomical location of the FFA and of the OFA (Fig. 2A,B; Supplementary Tables 8–11). Stronger activation for facial expression sounds was also observed in dorsolateral visual areas, the right pSTS, the auditory cortex and frontal regions (Fig. 2A,B; Supplementary Tables 8–11). In the comparison with speech sounds, but not with animal sounds, stronger activation for both types of facial expression sounds was observed in the primary visual cortex (Fig. 2A,B; Supplementary Tables 8–11).

Facial expression sounds—functional preference analysis

To statistically test for areas showing functional preference for the facial expression sounds, relative to all other sounds used in the experiment, in the blind participants, we performed a conjunction (i.e., logical AND) analysis of all contrasts between the sounds of facial expressions and the other sound categories (six contrasts included, see Materials and Methods and Fig. 2C). This stringent test of our hypothesis yielded a significant effect in the right lateral fusiform gyrus, in the typical location of the FFA (cluster center of mass MNI coordinates: 41 −48 −15; Fig. 2C). Furthermore, significant conjunction effects were also observed in the dorsolateral parts of the right visual cortex (the V5/MT area), the right STS, including the right pSTS, the left pSTS and right frontal regions (Fig. 2C).

Sighted Group

Animate sound categories versus object sounds

In the sighted group, both types of the facial expression sounds induced stronger response, relative to object sounds, in the auditory cortex and the pSTS, similarly to what was observed in the blind participants (Fig. 3A,B; Supplementary Tables 12, 13). Preferential activation for the facial expression sounds was also detected in frontal regions although in the sighted participants this effect was constrained primarily to the inferior frontal gyrus (Fig. 3A,B; Supplementary Tables 12, 13). Intriguingly, and in contrast to the results for the blind group, no preferential response in the fusiform gyrus was observed in the sighted participants, neither for the emotional nor for the nonemotional facial expression sounds (Fig. 3A,B; Supplementary Tables 12, 13). The emotional facial expression sounds induced some preferential activations, relative to object sounds, in other visual regions—particularly in the dorsolateral visual areas, the posterior part of the inferior occipital gyrus and the cuneus (Fig. 3A; Supplementary Table 12). However, similar effects were also observed in comparisons

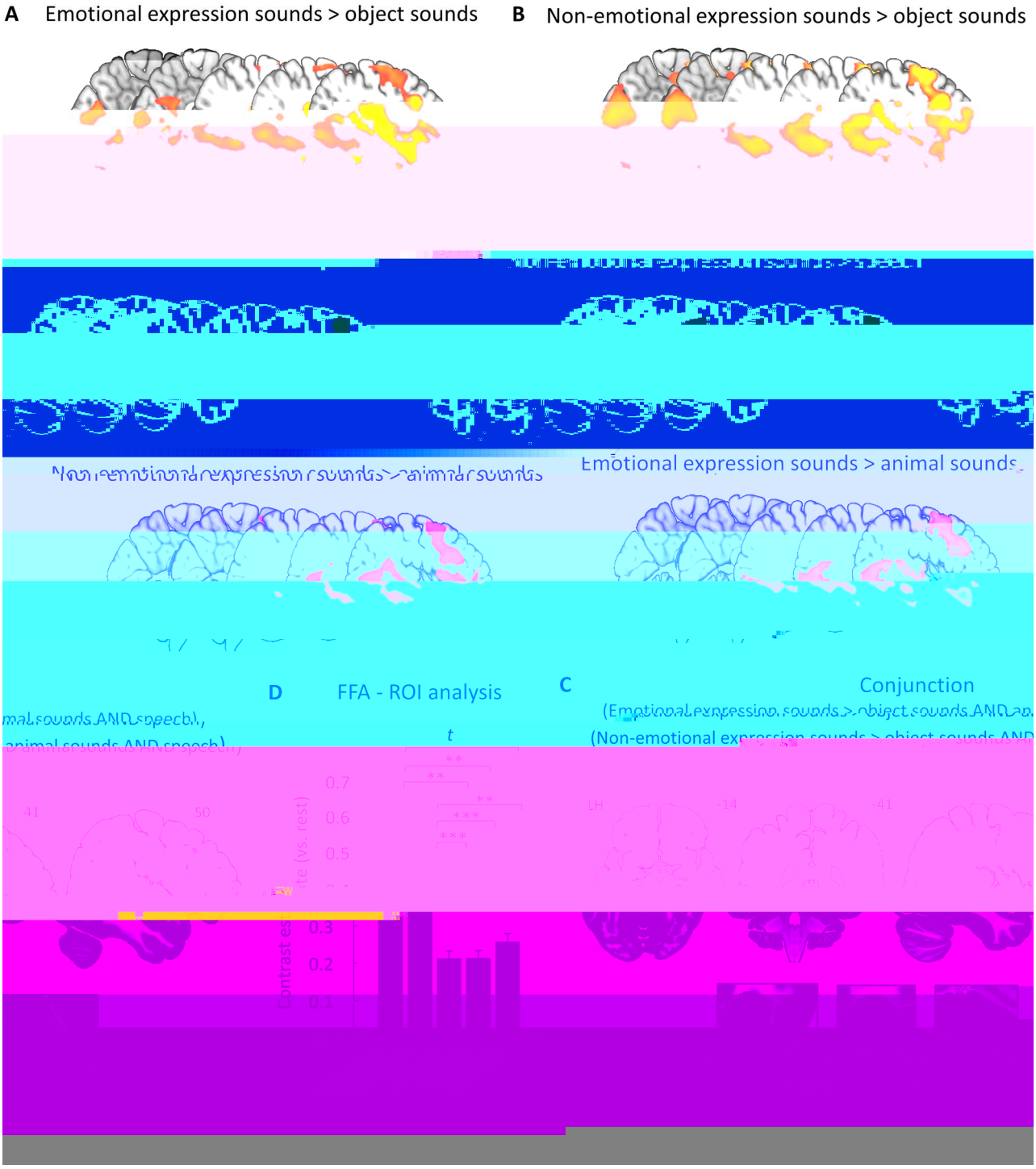

Figure 2. Functional preference for facial expression sounds in the fusiform gyrus of congenitally blind subjects. (*A*, *B*) Activation for the facial expression sounds relative to the activation for object sounds, speech sounds, and animal sounds. Results are presented separately for the expression sounds conveying emotional information (laughing and crying) and for the expression sounds that do not convey such information (yawning and sneezing). (*C*) Conjunction of the six contrasts presented in panels *A* and *B*. The burgundy ROIs represent the typical location of the FFA and the typical location of the right pSTS responsive to facial expressions as indicated by the Neurosynth meta-analyses. Close-ups of the two ROIs, in the planes corresponding to the whole-brain illustrations, are provided at the bottom for clarity. (*D*

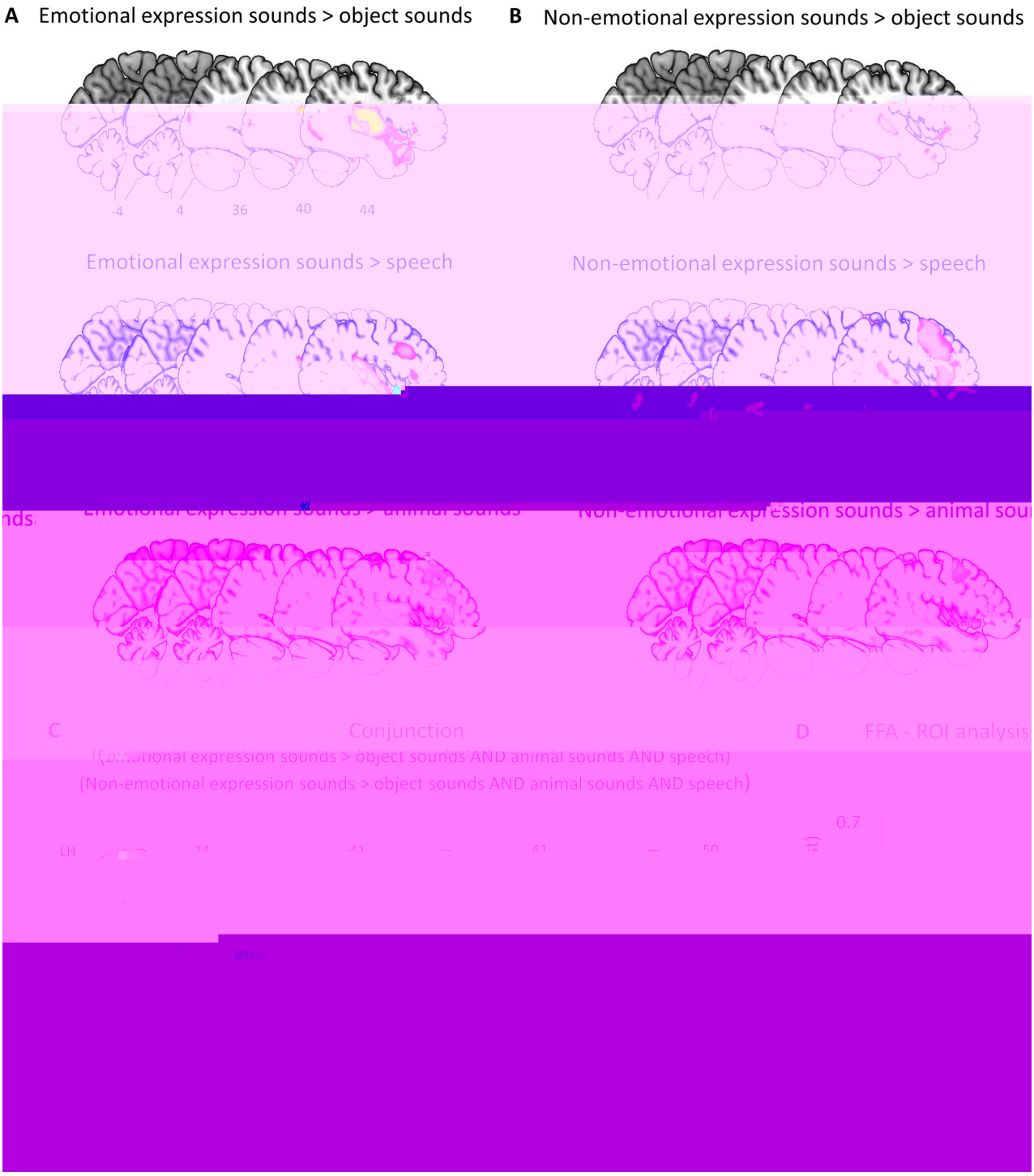

Figure 3. Lack of preferential univariate activation for facial expression sounds in the fusiform gyrus of sighted subjects. (*A*, *B*) Activation for the facial expression sounds relative to the activation for object sounds, speech, and animal sounds. Results are presented separately for the expression sounds conveying emotional information (laughing and crying) and for the expression sounds that do not convey such information (yawning and sneezing). (*C*) Conjunction of the six contrasts presented in panels *A* and *B*. The burgundy ROIs represent the typical anatomical location of the FFA and the part of the right pSTS that is responsive to facial expressions, as determined by Neurosynth meta-analyses. Close-ups of the two ROIs, in the planes corresponding to the whole-brain illustrations, are provided at the bottom for clarity. (*D*) Mean contrast estimates for each sound category in the FFA ROI (E expressions—emotional expressions; N-E expressions—nonemotional expressions). Numbers in panels *A* and *C* denote MNI coordinates of presented axial, coronal, or sagittal planes.

sighted participants showed significant effects in the auditory cortex, the right pSTS and frontal regions; however, no significant differences were detected in the visual cortex (Fig. 3A,B; Supplementary Tables 16, 17, 18, 19).

Facial expression sounds—functional preference analysis

The conjunction analysis of all six contrasts between the facial expression sounds and the other sound categories in the sighted participants yielded significant effects in the left and the right auditory cortex, the left and the

Figure 4. Direct test of between-group differences in sensitivity to facial expression sounds. Brain regions preferentially activated by facial expression sounds (averaged activation for emotional and non-emotional expressions), relative to other experimental conditions (averaged activation for speech, object sounds, and animals sounds), in the blind subjects but not (or not to the same extent) in the sighted subjects. The burgundy ROIs represent the typical anatomical location of the FFA and the part of the right pSTS that is responsive to facial expressions, as determined by Neurosynth meta-analyses. Numbers denote MNI coordinates of presented axial, coronal or sagittal planes.

right STS, including the right pSTS, and the right inferior frontal gyrus; however, it did not show any significant effects in the visual cortex (Fig. 3C).

Between-Group Comparisons and Region-of-Interest Analysis

Between-group differences in sensitivity to facial expression sounds

To directly test for differences in functional preference for the facial expression sounds across groups, we performed an additional condition (average from all facial expression sounds > average from all other sounds) by group interaction analysis. In line with the results observed within each group, this analysis indicated a stronger preference for facial expression sounds in the right lateral fusiform gyrus and in the right inferior occipital gyrus (typical anatomical locations of the FFA and the OFA, respectively) in the blind participants (Fig. 4; Supplementary Table 20). Furthermore, the same effect was observed in the early visual cortex and in the dorsolateral visual areas, bilaterally, as well as in the most posterior parts of the left and the right pSTS (Fig. 4; Supplementary Table 20). An inverse contrast, which tested for brain areas showing stronger functional preference for facial expression sounds in the sighted group, did not yield any significant results.

Region-of-interest analysis in the fusiform face area and the occipital face area

To further confirm spatial correspondence between the effect observed for facial expression sounds in the blind participants and the typical anatomical locations of the FFA and the OFA in sighted population, we performed an ROI analysis, in which those areas were defined based on the meta-analysis of face perception studies (see Materials and Methods). In the FFA ROI, a 2 (group) × 5 (sound category) ANOVA showed significant main effects of sound category ($F(4, 160) = 5.02$, $P = 0.001$, partial Eta squared = 0.11) and group ($F(1, 40) = 6.87$, $P = 0.012$, partial Eta squared = 0.15) as well as significant interaction between these factors ($F(4, 160) = 5.65$, $P < 0.001$, partial Eta squared = 0.12). In line with the whole-brain analysis results, pairwise comparisons showed stronger activation for both types of facial expressions sounds than for any other sound category included in the study in the blind group (Fig. 2D). In contrast, no significant differences between conditions were detected in the sighted group (Fig. 3D). Furthermore, direct comparisons between groups showed stronger activation in the blind group for emotional facial expressions (mean difference = 0.25, SEM = 0.08, $P_{FDR} = 0.01$) and nonemotional facial expressions (mean difference = 0.27, SEM = 0.08, $P_{FDR} = 0.005$). Trends towards a similar difference for other sound categories—animal sounds (mean difference = 0.16, SEM = 0.08, $P_{FDR} = 0.075$), speech sounds (mean difference = 0.11, SEM = 0.07, $P_{FDR} = 0.1$) and object sounds (mean difference = 0.13, SEM = 0.07, $P_{FDR} = 0.1$)—were also detected. In the OFA ROI, we also detected significant main effects of sound category ($F(4, 160) = 3.97$, $P = 0.004$, partial eta squared = 0.09) and group ($F(1, 40) = 6.87$, $P = 0.023$, partial eta squared = 0.12) as well as a significant interaction between these factors ($F(4, 160) = 5.08$, $P < 0.001$, partial Eta squared = 0.11). Pairwise comparisons showed that a pattern of results for the OFA ROI is overall similar to the pattern observed in the FFA ROI—that is, we found functional preference for facial expression sounds in the blind group and no significant differences across conditions in the sighted group (Supplementary Fig. 4).

Region-of-interest analysis in the fusiform face area—regression of frontal and temporal activations

In blind participants, the whole-brain conjunction analysis revealed functional preference for facial expression sounds not only in the typical location of the FFA, but also in several other brain regions, particularly in the right pSTS and the right frontal cortex (Fig. 2C). To investigate whether the activation in the typical location of the FFA contains information beyond that captured in the latter two regions, we repeated the above-described ROI analysis in the FFA with either the right frontal activations or the right pSTS activations for all sound categories included as covariates (see Materials and Methods). In both analyses, the 2 × 5 ANOVA for the FFA ROI showed significant interaction between the group and the sound category (both $F > 4.5$, both $P < 0.01$). Pairwise comparisons revealed result patterns similar to those pattern that was obtained in the original analysis—in particularly, the functional preference for both categories of facial expression sounds, compared to other

sound categories, was still observed in the blind group (Supplementary Fig. 5). Thus, the activations observed in the FFA seem to be relatively independent of activations in the right pSTS and the frontal cortex.

Region-of-interest analysis in the right posterior superior temporal sulcus

The whole-brain analysis suggests that the functional preference for the facial expression sounds used in our study can be observed, in both groups, in the part of the right pSTS that is known to be sensitive to information about dynamic features of face shape (Allison et al. 2000; Andrews and Ewbank 2004; Pitcher et al. 2014). To confirm this important control result, we performed an ROI analysis, in which the right pSTS was defined based on the meta-analysis of studies investigating brain responses to facial expressions (see Materials and Methods). The 2 (group) × 5 (sound category) ANOVA showed a significant main effect of sound category ($F(4, 160) = 35.91$, $P < 0.001$, partial Eta squared = 0.47) whereas the main effect of group and the interaction between the two factors were not significant (all $P > 0.15$). As expected, in both groups, post hoc tests showed stronger activation for both types of facial expression sounds relative to all other sound categories used in the experiment (Supplementary Fig. 6; all $P < 0.01$). Thus, although the whole-brain interaction analysis hinted at stronger functional preference for facial expression sounds in the most posterior part of the right pSTS in the blind group (Fig. 4), an activation pattern averaged from this whole area was comparable across groups.

Region-of-interest analysis in auditory cortices

The sounds from different domains and semantic categories inevitably differ on many auditory properties (see Materials and Methods). Could differences in acoustic features or acoustic complexity drive the fusiform gyrus response pattern in the blind participants? We find this possibility highly unlikely as the difference in activation of this region was found only in comparison between facial expression sounds and other sound categories. We did not observe activation differences in comparisons between other sounds categories (for example, between object sounds and animal sounds), despite the fact that these categories were also markedly different in terms of their auditory properties. Furthermore, in the whole-brain conjunction analysis we did not observe functional preference for facial expression sounds in the auditory cortices, particularly in the primary auditory cortex (A1). To illustrate this point further, we performed the ROI analyses in the A1 and in the STS, which hosts higher-level auditory areas (Supplementary Figs 7 and 8). In both these ROIs, the activation patterns were similar in the blind and the sighted group and markedly different from the pattern observed in the fusiform gyrus in the blind group. This finding confirms that differences in auditory properties are unlikely to explain the results reported here.

Winner-take-all OTC maps

To further explore the pattern of functional preference for specific sound categories in the OTC, winner-take-all maps were calculated for this region and for each group (see Materials and Methods). In both groups, the analysis resulted in an expected large-scale pattern with activation for inanimate sounds being stronger in the lateral occipital complex and the parahippocampal cortex, and the activation for animate sounds being stronger in the fusiform gyrus (Supplementary Fig. 9). In blind participants, the analysis indicated preferential activation of the fusiform gyrus for facial expression sounds as already shown by the statistical threshold analyses. Although in sighted participants the analysis seemed to indicate consistent preference of this region to speech sounds and, to a certain extent, to nonemotional expression sound, none of these effects reached significance in threshold analyses.

Functional Magnetic Resonance Imaging Experiment: Multivoxel Pattern Analysis

Multivoxel Pattern Classification of Sound Categories

The univariate analysis of the fMRI data revealed strong functional preference for facial expression sounds in the typical anatomical location of the FFA in the blind participants. In contrast, this area did not show any above-threshold univariate differentiation of presented sounds in the sighted group. Nonetheless, it is still possible that the FFA in the sighted participants represents the difference between the sound categories used in the experiment by means of dispersed, subthreshold activation patterns. To investigate this possibility, we performed a multivoxel pattern classification analysis in this area (see Materials and Methods).

We started by performing pairwise classifications, based on the FFA activation, for all pairs of sound categories used in the experiment (Fig. 5). In the blind group, the classifier successfully distinguished between all category pairs except for the only pair that did not include human sounds—that is, the animal—object sound pair. In line with our expectations, the classification was also, to some extent, successful in the sighted group. Specifically, in the sighted group, the classifier successfully distinguished between object sounds and each human sound category. Furthermore, the successful classification in this group was achieved for pairs including speech sounds and either of the facial expression sound categories.

We repeated the same analysis in three control ROIs—the broadly-defined auditory cortex (the primary auditory cortex and the STS ROIs combined), the PPA (Epstein and Kanwisher 1998), and the right pSTS. We documented highly successful classification of all category pairs, in both groups, in the auditory cortex (Supplementary Fig. 10), and in the right pSTS (Supplementary Fig. 11). In contrast, classification of animate category pairs was generally not successful in the PPA (Supplementary Fig. 12). As can be expected, this

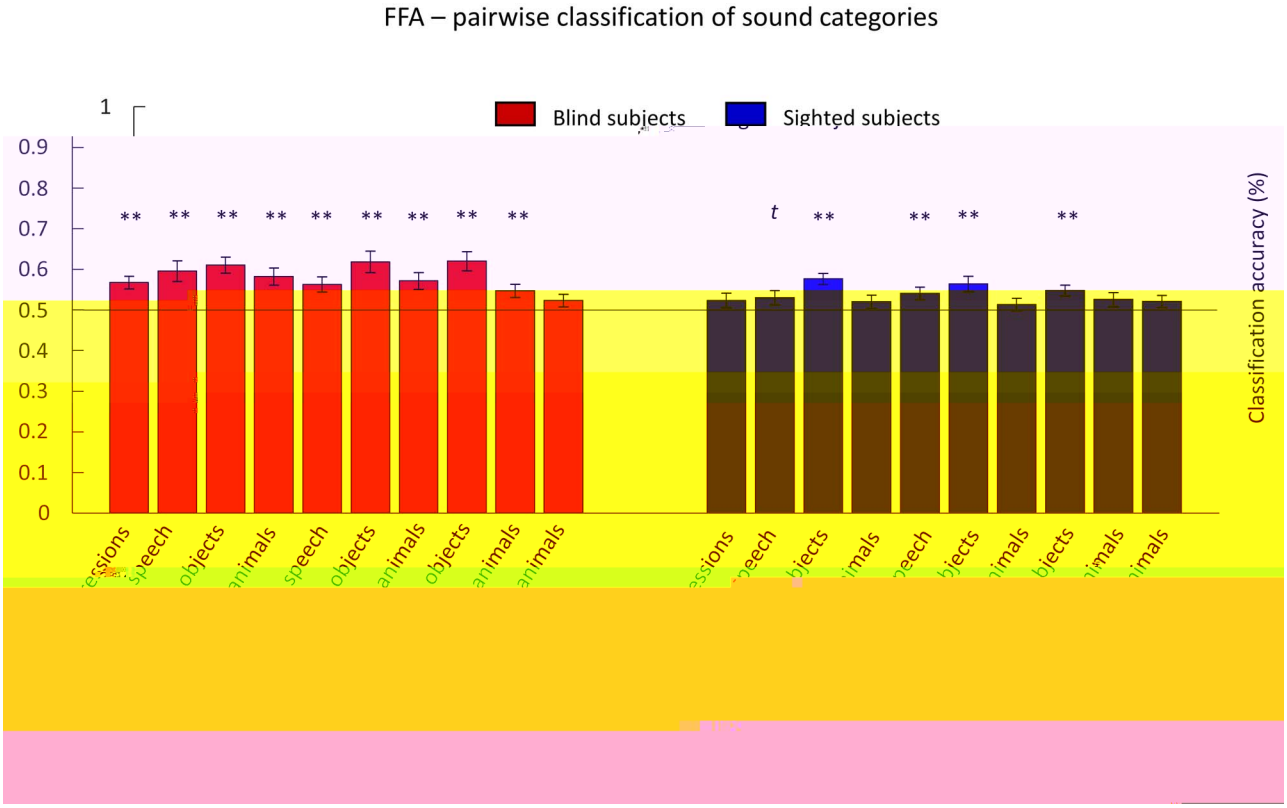

Figure 5. Human sounds induce distinctive multivoxel activation patterns in the fusiform gyrus, in both blind and sighted subjects. (*A*, *B*) Mean accuracy of a support vector machine classifier distinguishing between every pair of sound categories included in the experiment, in the blind and the sighted group. E expression—emotional expression sounds; N-E expressions—nonemotional expression sounds; $^{*}P < 0.05$, $^{**}P < 0.01$, t $P = 0.051$, corrected for multiple comparisons across all tests performed using FDR. Error bars represent standard error of the mean. Chance classification level is marked with a black line.

area seemed to distinguish mostly between the object sounds and the animate sound categories—especially in the blind group.

Multivoxel Pattern Classification of Specific Facial Expression Sounds

Does the typical anatomical location of the FFA code the differences between specific facial expressions presented through the auditory modality? To address this question, we trained a classifier to distinguish, based on the activation of this area, between four facial expressions presented in our study (crying, laughing, sneezing, and yawning) irrespective of the differences in voice characteristics and in the gender of the two actors (man and woman; see Materials and Methods). The sounds were successfully classified into specific facial expressions ine

Figure 6. Activation patterns in the fusiform gyrus of blind subjects code differences between specific facial expressions presented through the auditory modality. Mean accuracy of a support vector machine classifier distinguishing, based on the activation in the typical location of the FFA, between 1) four specific facial expressions (crying, laughing, sneezing and yawning), irrespectively of the actor; 2) two specific speech sounds, irrespectively of the actor; and 3) the actor (a man and a woman), irrespectively of the specific facial expression or the speech sound. The presented accuracy is adjusted for chance level (25% for facial expression classification; 50% for speech sound and gender classification). $^{**}P < 0.01$, corrected for multiple comparisons across all comparisons performed using FDR. Error bars represent standard error of the mean.

with the neural representational dissimilarity matrices (RDMs), representing actual differences in activation patterns observed for the sound categories in both groups (Fig. 7A,B; see Materials and Methods for details).

In the blind group, simple correlation analyses revealed significant correlations between neural RDM and the three models—the category model, the human/nonhuman model, and the animate/inanimate model. No significant correlation was observed for the pitch model (Fig. 7C). In the sighted group, the same analysis resulted in significant correlation only for the category model (Fig. 7C).

We then performed partial correlation analyses, in which each of the four models was correlated with neural RDMs whereas the three remaining models were treated as covariates (see Materials and Methods). In both groups, significant effects were observed only for the category model (Fig. 7C). Overall, the model representing distinctions known to be represented in the FFA, in the visual modality, explains more variance in this area's responses to sound categories than the three control models, in both the congenitally blind and the sighted participants.

Discussion

In this study, we found that sounds associated with human facial expressions induce preferential response in the lateral fusiform gyrus—the typical anatomical location of the FFA—in congenitally blind individuals. This effect is independent of the emotional content (or lack thereof) of the facial expression sounds and was observed in comparisons to object sounds, animal sounds, and human speech. In the sighted participants, the threshold univariate analyses showed no clear functional preference in the lateral fusiform gyrus for any of the sound categories. However, the multivoxel pattern classification showed that distinctions between human sounds and object sounds are robustly represented in the lateral fusiform gyrus in both groups. Furthermore, the RSA showed that the categorical model of response patterns that would be typically expected from the FFA correlates with the observed response patterns induced by sounds in both groups. This effect was observed even after accounting for the differences in pitch and more general human/nonhuman and animate/inanimate distinctions.

The observed differences in univariate activations in the blind group are in line with our initial predictions. They bring a fresh perspective to a debate about factors driving the development of functional specialization in the fusiform gyrus. In sighted individuals, this region is known to develop robust preference for animate objects, with the development of preference for human faces in the FFA being a prime example of this process. However, the origin of this specialization—particularly, the relative importance of visual experience and nonvisual factors in driving it—is relatively poorly understood (e.g., Caramazza and Shelton 1998; Mahon and Caramazza 2011; Saygin et al. 2012; Arcaro et al. 2017, 2019; Powell et al. 2018; Livingstone et al. 2019). Studying the activation profile of the fusiform gyrus in blind individuals seems critical to better understand the role of these two sets of factors. Yet, this research line has resulted in surprisingly mixed results, with approximately half of studies

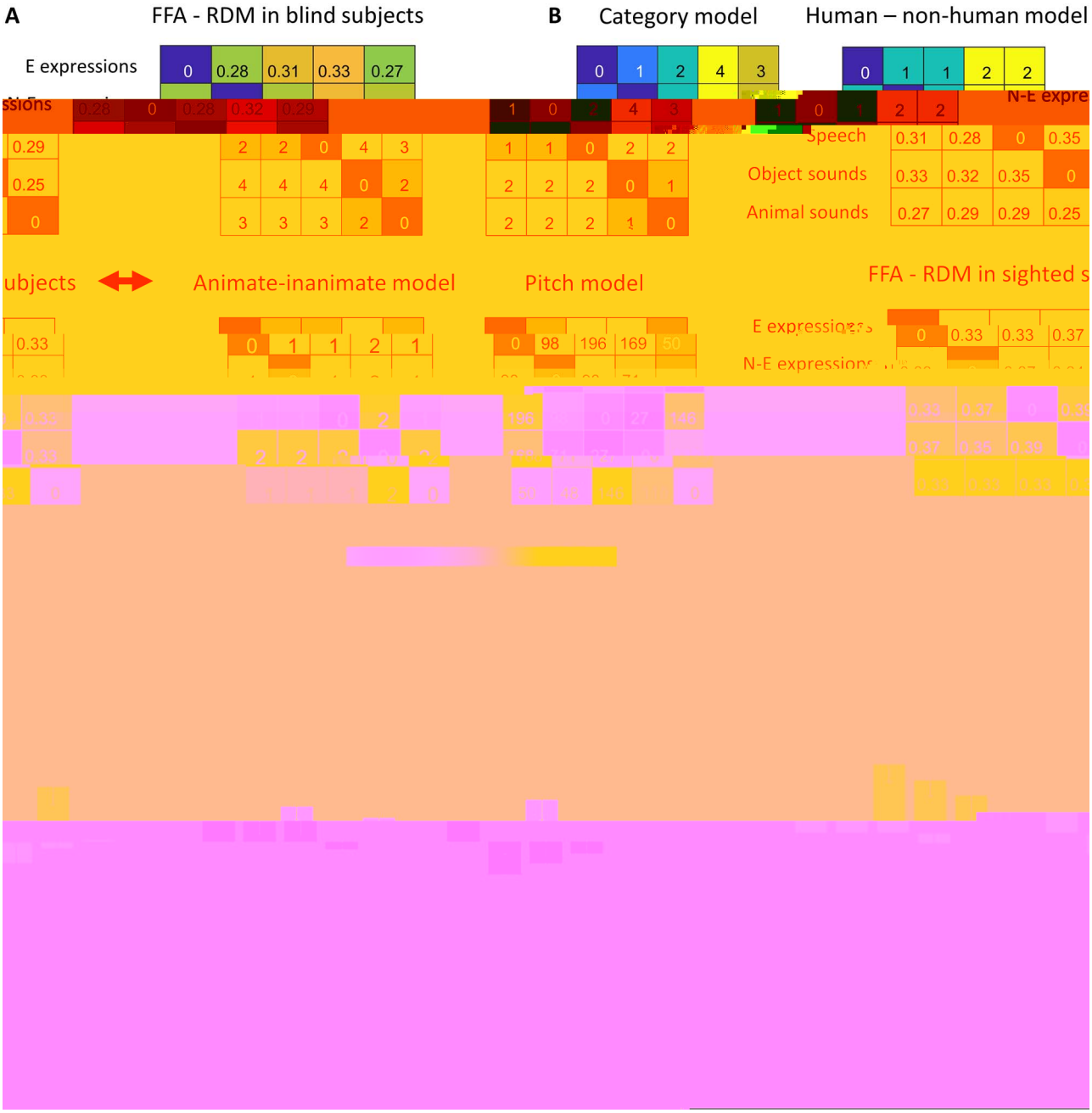

Figure 7. Responses patterns induced in the fusiform gyrus by sound categories correlate with the model of the FFA responses created based on visual studies. (A) Neural representational dissimilarity matrices (RDMs) representing differences in the response patterns observed in the typical anatomical location of the FFA, in each group. (B) Four dissimilarity models that were correlated with neural RDMs. (C) Simple and partial correlations between the neural RDMs and the dissimilarity models. In the partial correlation analysis, each model was correlated with neural RDMs whereas the three remaining models were treated as covariates. *$P < 0.05$, **$P < 0.01$, corrected for multiple comparisons across all tests performed using FDR. Error bars represent standard error of the mean. E expressions—emotional expression sounds, N-E expressions—nonemotional expression sounds.

documenting preference for face- and animal-related information even in the fusiform gyri in congenitally blind participants (Pietrini et al. 2004; Kitada et al. 2009; Handjaras et al. 2016, 2017; Fairhall et al. 2017; van den Hurk et al. 2017; Mattioni et al. 2020; Murty et al. 2020), and the remaining half of studies reporting null effects for the animate domain (Goyal et al. 2006; He et al. 2013; Kitada et al. 2013; Fairhall et al. 2014; Plaza et al. 2015; Dormal et al. 2018). Here, we propose an explanation of these seemingly incoherent results. In an extension of the conjecture formulated by Bi and colleagues (Bi et al. 2016), we hypothesized that specific animate representations hosted by the fusiform gyrus are differently associated with action system representations, and that the strength and transparency of this association determines whether or not the fusiform representations can be activated by nonvisual information. In support of this hypothesis, we found that facial expression sounds, which contain a wealth of information relevant to the action and the motor system, induce robust, preferential

responses in the typical anatomical location of the FFA in congenitally blind participants. In contrast, we found no clear functional preference for speech or animal sounds over inanimate object sounds in this area (or in the neighboring regions). Notably, functional preference for pictures of neutral faces and animals over pictures of objects is observed in the FFA in the visual modality (e.g., Kanwisher et al. 1999). Our results help clarify two issues about the role of visual experience in determining the organization of domain preferences in visually-driven cortex. First, they add to a growing number of studies showing that the lateral fusiform gyrus remains part of the cortical network for face processing, even in congenitally blind people (van den Hurk et al. 2017; Murty et al. 2020). In other words, this region responds preferentially to information from the same domain in congenitally blind and in sighted individuals. This preserved domain preference could be driven by innate connectivity patterns (Mahon and Caramazza 2011; Saygin et al. 2012; Powell et al. 2018) but it does not necessarily imply that this region retains the same functional role in the processing of face-related information in blind individuals. Second, and perhaps more importantly, our study shows that, in blind individuals, this functional organization can be revealed primarily for animate representations with transparent association with the action system.

On a more general level, the findings in the blind group help clarify a perplexing phenomenon—that is, the interaction of sensory modality by object domain in OTC. A recent conjecture proposed that the weaker auditory and tactile responsiveness of vOTC animate areas, relative to inanimate areas, can be explained by the fact that the mapping between shape and action representations for the animate domain is less transparent than for inanimate objects (Bi et al. 2016). In this study, we demonstrate the predictive value of this hypothesis by showing that effects characteristic of the inanimate domain can be also observed for specific subsets of animate entities with clear relationship between shape and action representations, at least in the absence of visual experience. This suggests that the OTC organization might be determined not only by local sensitivity to certain features, but also by metrics computed downstream in the brain, particularly in the action system (Mahon et al. 2007; Mahon and Caramazza 2011).

The fact that we did not observe the expected, univariate pattern of results in sighted participants might raise a question about generalizability of the above-described principle to the sighted population. A clear difference in the univariate results with human sounds obtained for the two groups seems to be inconsistent with the results from the inanimate domain, which are similar for these two groups (Mahon et al. 2009; Wolbers et al. 2011; He et al. 2013; Peelen et al. 2013; Wang et al. 2015). Nevertheless, the RSA results suggest that, also in sighted individuals, the FFA represents nonvisually generated information as meaningful, subthreshold activation patterns, and that this representation is precise enough to differentiate between, for example, various animate objects (as indicated by the fact that the category model explained more variance in the RSA than the animate–inanimate model). Furthermore, the RSA results suggest that the representation of nonvisually generated information in the fusiform gyrus is similar across groups, and similar to the representation of visual information revealed by visual face perception studies (Kanwisher et al. 1999). Based on these considerations, we propose to interpret the between-group difference in univariate results in terms of the adaptation of the FFA circuitry to the strength of incoming signals. In sighted individuals, the FFA receives strong feedforward inputs from early visual cortices, which might make this region less sensitive to relatively weak inputs from other brain systems. In congenitally blind individuals, lack of feedforward signals might lead to an adjustment of the FFA's sensitivity—as a result, this area might represent the differences in nonvisually generated information as above-threshold differences in univariate activation patterns. Such an interpretation is in line with studies showing that even short-term visual deprivation can increase the excitability of visual areas (Boroojerdi et al. 2000; Lunghi et al. 2015). This interpretation is also supported by our observation of generally weaker responses to sounds in the fusiform gyrus in sighted participants, compared to effects observed in blind participants (compare Figs 2D and 3D). On this view, the basic principles of the OTC organization are similar in blind and sighted individuals but lack of visual experience might be necessary for certain mechanisms to be clearly detectable, at least in the case of the animate domain, which generally seems to be more strongly linked to the visual system than is the inanimate domain.

Can our results be interpreted as evidence of a more dramatic functional repurposing of the fusiform gyrus in congenitally blind individuals? One possibility is that, in the absence of visual experience, this region takes over the computational role of other parts of the face perception network—for example, one can suppose that in blind individuals the computations carried out by the fusiform gyrus start to resemble the computations conducted in the right pSTS or the right IFG (see below for discussion of the results obtained in these areas). Although we cannot rule out such plasticity within the face perception network entirely, the fusiform gyrus results for the blind participants remain significant even after accounting for the activation of the right pSTS or the right IFG, a finding that speaks against the strong plasticity claim. Another possibility is that, in blind individuals, the fusiform gyrus assumes functions that are not linked to the face domain at all. Here, one candidate function is processing of the acoustic properties of sounds—thus, one can suppose that acoustic differences drive the observed difference between facial expression categories and the other stimuli. However, we find this possibility unlikely for several reasons. First, previous studies have already suggested that the vOTC of blind participants is not sensitive to

the acoustic properties of sounds (Watkins et al. 2013). Second, and in line with previous reports, 1) our control univariate and multivoxel pattern classification analyses showed markedly different patterns of results for the typical location of the FFA and for the auditory cortices in the blind participants; and 2) our RSA showed that the typical location of the FFA in the blind participants captures the similarity between sound categories, but not pitch similarity. Third, comparable preference for facial expression sounds was not observed in other OTC regions, such as the parahippocampal cortex or the lateral occipital cortex. It is unclear why only the lateral fusiform gyrus and not the neighboring areas would be sensitive to the acoustic properties of sounds. Another candidate function is processing of task demands. However, the task used in this study was to indicate whether a given sound was produced by a human or not. If the task "decide if this sound is produced by a human" or the motor/mental processes associated with such a decision activated the fusiform gyrus in the blind participants, then one would expect to find comparable activation for expression sounds and speech sounds in this region. This was clearly not the case in our study. Second, the task was very easy—thus, the difficulty associated with detecting the stimuli belonging to each category is unlikely to affect the results. Third, as described above, the functional preference for facial expression sounds relative to all control conditions was observed only in the typical anatomical location of the FFA, but not in other vOTC regions. There is no reason to expect such spatial specificity if the observed effects were driven by task demands or other general (i.e., not domain-specific) mechanisms. Finally, the human–nonhuman model, which was also the model representing the difference that the participants were asked to detect, did not explain the activation patterns in the fusiform gyrus as well as done by the category model. In summary, neither sensitivity to acoustic properties nor effects of task demands could fully explain the findings reported in this study.

Apart from the effects in the FFA and the pSTS—the canonical face perception areas—we observed functional preference for facial expression sounds in the dorsolateral visual cortex and the right frontal and prefrontal cortex, particularly in the blind group. Previous studies have already indicated that these areas become activated when sighted participants perceive facial expressions in the visual modality. For example, in their recent review, Pitcher and Ungerleider (2020) summarize the evidence that the dorsolateral visual cortex (V5/MT complex) provides critical information about facial expressions to the pSTS. In earlier experimental work, Pitcher et al. (2011) have shown that short videos of human facial expressions preferentially activate the right inferior frontal gyrus and the right motor cortex, when compared with short videos of objects. Based on this evidence, these areas could perhaps be considered as belonging to an "extended face perception network"—that is, they are likely to be face-relevant in a broad sense of being involved in the perception, evaluation, or understanding of facial expressions. The fact that we observed similar functional preference for the facial expression sounds, that is, in a nonvisual sensory modality, supports such an interpretation.

The experimental evidence shows that the main role of the FFA in sighted individuals is in processing invariant face features (e.g., Haxby et al. 2000; Calder and Young 2005). However, a growing number of studies suggest that, to some extent, the FFA is also sensitive to facial expression information (e.g., Ganel et al. 2005; Fox et al. 2009; Harry et al. 2013; Bernstein and Yovel 2015). Although the exact mechanisms and pathways underlying this sensitivity remain to be elucidated, one possibility is that information about facial expressions and other changeable aspects of face shape are processed by a specialized visual pathway (including the pSTS as a key region) and only later communicated to the FFA (Pitcher et al. 2014; Pitcher and Ungerleider 2020). One can hypothesize that the existence of such pathways, which transfer to the FFA the facial expression information that is already cognitively evaluated, allows this area to represent facial expression information even when received through nonvisual means.

Previous studies suggest that early visual experience is important for the development of typical face perception abilities, including configural face processing (e.g., Le Grand et al. 2001, 2003; Röder et al. 2013). Our results do not argue against these findings. We suggest that the fusiform gyrus remains a part of the face-perception network even in the absence of visual experience, illustrating the importance of innate connectivity patterns in shaping functional specialization of the brain. However, it is not our claim that the innate connectivity patterns alone are sufficient for the development of a fully-fledged face perception network, which would be capable of the typical configural face processing once vision were restored (if such were possible). Innate connectivity patterns are likely to be one of several factors, and local preference for certain low-level stimulus features is another likely candidate (Simion et al. 2001; Shimojo et al. 2003; Viola Macchi et al. 2004; Turati et al. 2006), which in combination lead to the development of the fully-fledged FFA in sighted individuals.

In conclusion, we have shown that auditory responsiveness of the lateral fusiform gyrus (the typical anatomical location of the FFA) in congenitally blind individuals is modulated by the type of evoked face shape representation and its relationship with face action representations. Thus, OTC representation of both inanimate and animate stimuli can be activated by auditory signals as long as the mapping between shape and associated action/motor computations is stable and transparent. However, our study suggests that the absence of potentially competing visual inputs seems necessary for this effect to clearly manifest in the case of animate representation.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Funding

Polish Ministry of Science (DN/MOB/023/V/2017 to Ł.B.); the National Science Center Poland (2020/37/B/HS6/01269 to Ł.B.); the Kosciuszko Foundation (a 2018–2019 fellowship to Ł.B.); the National Natural Science Foundation of China (82021004 to Y.B., 31925020 to Y.B., 32100837 to H.Y.); the Changjiang Scholar Professorship Award (T2016031 to Y.B.); and the 111 Project (BP0719032 to Y.B.).

Notes

Conflict of interest: The authors declare that no competing interests exist.

References

Allison T, Puce A, McCarthy G. 2000. Social perception from visual cues: role of the STS region. *Trends Cogn Sci*. 4:267–278.

Amedi A, Malach R, Hendler T, Peled S, Zohary E. 2001. Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci*. 4:324–330.

Amedi A, Floel A, Knecht S, Zohary E, Cohen LG. 2004. Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nat Neurosci*. 7:1266–1270.

Andrews TJ, Ewbank MP. 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*. 23:905–913.

Arcaro MJ, Schade PF, Vincent JL, Ponce CR, Livingstone MS. 2017. Seeing faces is necessary for face-domain formation. *Nat Neurosci*. 20:1404–1412.

Arcaro MJ, Schade PF, Livingstone MS. 2019. Universal mechanisms and the development of the face network: what you see is what you get. *Annu Rev Vis Sci*. 5:341–372.

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

Bernstein M, Yovel G. 2015. Two neural pathways of face processing: a critical evaluation of current models. *Neurosci Biobehav Rev*. 55: 536–546.

Bi Y, Wang X, Caramazza A. 2016. Object domain and modality in the ventral visual pathway. *Trends Cogn Sci*. 20:282–290.

Bischoff-Grethe A, Ozyurt IB, Busa E, Quinn BT, Fennema-Notestine C, Clark CP, Morris S, Bondi MW, Jernigan TL, Dale AM, et al. 2007. A technique for the deidentification of structural brain MR images. *Hum Brain Mapp*. 28:892–903.

Boroojerdi B, Bushara KO, Corwell B, Immisch I, Battaglia F, Muell-

James TW, Culham J, Humphrey GK, Milner AD, Goodale MA. 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain*. 126:2463–2475.

Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc B Biol Sci*. 361:2109–2128.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.

Kanwisher N, Stanley D, Harris A. 1999. The fusiform face area is selective for faces not animals. *Neuroreport*. 10:183–187.

Kitada R, Johnsrude IS, Kochiyama T, Lederman SJ. 2009. Functional specialization and convergence in the occipito-temporal cortex supporting haptic and visual identification of human faces and body parts: an fMRI study. *J Cogn Neurosci*. 21:2027–2045.

Kitada R, Okamoto Y, Sasaki AT, Kochiyama T, Miyahara M, Lederman SJ, Sadato N. 2013. Early visual experience and the recognition of basic facial expressions: involvement of the middle temporal and inferior frontal gyri during haptic identification by the early blind. *Front Hum Neurosci*. 7:7.

Konkle T, Caramazza A. 2013. Tripartite organization of the ventral stream by animacy and object size. *J Neurosci*. 33:10235–10242.

Kourtzi Z, Kanwisher N. 2000. Cortical regions involved in perceiving object shape. *J Neurosci*. 20:3310–3318.

Le Grand R, Mondloch CJ, Maurer D, Brent HP. 2001. Early visual experience and face processing. *Nature*. 410:890–890.

Le Grand R, Mondloch CJ, Maurer D, Brent HP. 2003. Expert face processing requires visual input to the right hemisphere during infancy. *Nat Neurosci*. 6:1108–1112.

Livingstone MS, Arcaro MJ, Schade PF. 2019. Cortex is cortex: ubiquitous principles drive face-domain development. *Trends Cogn Sci*. 23:3–4.

Lunghi C, Emir UE, Morrone MC, Bridge H. 2015. Short-term monocular deprivation alters GABA in the adult human visual cortex. *Curr Biol*. 25:1496–1501.

Magri C, Konkle T, Caramazza A. 2021. The contribution of object size, manipulability, and stability on neural responses to inanimate objects. *NeuroImage*. 237:118098.

Mahon BZ, Caramazza A. 2011. What drives the organization of object knowledge in the brain? *Trends Cogn Sci*. 15:97–103.

Mahon BZ, Milleville SC, Negri GA, Rumiati RI, Caramazza A, Martin A. 2007. Action-related properties shape object representations in the ventral stream. *Neuron*. 55:507–520.

Mahon BZ, 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.

Mattioni S, Rezk M, Battal C, Bottini R, Mendoza KEC, Oosterhof NN, Collignon O. 2020. Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *Elife*. 9:e50732.

Murty NAR, Teng S, Beeler D, Mynick A, Oliva A, Kanwisher NG. 2020. Visual experience is not necessary for the development of face selectivity in the lateral fusiform gyrus. *Proc Natl Acad Sci U S A*. 117:23011–23020.

O'Craven KM, Kanwisher N. 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci*. 12:1013–1023.

Oosterhof NN, Connolly AC, Haxby JV. 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front Neuroinform*. 10:27.

Peelen MV, Bracci S, Lu X, He C, Caramazza A, Bi Y. 2013. Tool selectivity in left occipitotemporal cortex develops without vision. *J Cogn Neurosci*. 25:1225–1234.

Peelen MV, He C, Han Z, Caramazza A, Bi Y. 2014. Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. *J Neurosci*. 34:163–170.

Peirce JW. 2007. PsychoPy-psychophysics software in python. *J Neurosci Methods*. 162:8–13.

Pietrini P, Furey ML, Ricciardi E, Gobbini MI, Wu WHC, Cohen L, Guazzelli M, Haxby JV. 2004. Beyond sensory images: object-based representation in the human ventral pathway. *Proc Natl Acad Sci U S A*. 101:5658–5663.

Pitcher D, Ungerleider LG. 2020. Evidence for a third visual pathway specialized for social perception. *Trends Cogn Sci*. 25:100–110.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*. 56:2356–2363.

Pitcher D, Duchaine B, Walsh V. 2014. Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Curr Biol*. 24:2066–2070.

Plaza P, Renier L, De Volder A, Rauschecker J. 2015. Seeing faces with your ears activates the left fusiform face area, especially when you're blind. *J Vis*. 15:197–197.

Powell LJ, Kosakowski HL, Saxe R. 2018. Social origins of cortical face areas. *Trends Cogn Sci*. 22:752–763.

Röder B, Ley P, Shenoy BH, Kekunnaya R, Bottari D. 2013. Sensitive periods for the functional specialization of the neural system for human face processing. *Proc Natl Acad Sci U S A*. 110:16760–16765.

Saygin ZM, Osher DE, Koldewyn K, Reynolds G, Gabrieli JD, Saxe RR. 2012. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat Neurosci*. 15:321–327.

Shimojo S, Simion C, Shimojo E, Scheier C. 2003. Gaze bias both reflects and influences preference. *Nat Neurosci*. 6:1317–1322.

Simion F, Macchi Cassia V, Turati C, Valenza E. 2001. The origins of face perception: specific versus non-specific mechanisms. *Inf Child Dev*. 10:59–65.

Smith SM, Nichols TE. 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 44:83–98.

Turati C, Macchi Cassia V, Simion F, Leo I. 2006. Newborns' face recognition: role of inner and outer facial features. *Child Dev*. 77: 297–311.

Viola Macchi C, Turati C, Simion F. 2004. Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychol Sci*. 15:379–383.

Wada Y, Yamamoto T. 2001. Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *J Neurol Neurosurg Psychiatry*. 71:254–257.

Wang X, Peelen MV, Han Z, He C, Caramazza A, Bi Y. 2015. How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *J Neurosci*. 35:12545–12559.

Watkins KE, Shakespeare TJ, O'Donoghue MC, Alexander I, Ragge N, Cowey A, Bridge H. 2013. Early auditory processing in area V5/MT+ of the congenitally blind brain. *J Neurosci*. 33: 18242–18246.

Whitfield-Gabrieli S, Nieto-Castanon A. 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect*. 2:125–141.

Wolbers T, Klatzky RL, Loomis JM, Wutte MG, Giudice NA. 2011. Modality-independent coding of spatial layout in the human brain. *Curr Biol*. 21:984–989.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*. 8:665–670.