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

Łukasz Bola[1#*], Huichao Yang[2,3#], Alfonso Caramazza[1,4], Yanchao Bi[2,3*]

[1] Department of Psychology, Harvard University

[2] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University

[3] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University

[4] Center for Mind/Brain Sciences (CIMeC), University of Trento

[#] These authors share first authorship

* Corresponding authors:

lukaszbola@fas.harvard.edu

ybi@bnu.edu.cn

Abstract

In high-level visual shape areas in the human brain, preference for inanimate objects is observed regardless of stimulation modality (visual/auditory/tactile) and subjects' visual experience (sighted/blind individuals), whereas preference for animate entities seems robust only in the visual modality. Here, we test a hypothesis explaining this effect: visual shape representations can be reliably activated through different sensory modalities only when they systematically map onto action system computations. We studied fMRI activations in congenitally blind and sighted subjects listening to animal, object, and human sounds. We found that face shape areas in blind individuals preferentially respond to human facial expression sounds, with transparent face shape/motor representation mapping, but not to speech or animal sounds. Using face areas' activation, we could distinguish facial expressions from other sounds in both groups. We conclude that auditory stimulation can activate visual shape representation of those stimuli – inanimate or animate - for which shape-action mapping is transparent.

Introduction

How does the brain transform modality specific signals into a representation that integrates stimulus properties from different senses? Here, we study an effect that might speak to this difficult question – that is, a stimulus domain by sensory modality interaction in high-level visual shape regions in the occipitotemporal cortex (OTC). Various studies have shown that different areas of this cortical territory show functional preference for inanimate objects regardless of stimulation modality (visual/auditory/tactile) and subjects' visual experience (sighted/blind individuals) (Mahon et al., 2009; Wolbers et al., 2011; He et al., 2013; Peelen et al., 2013; Dormal et al., 2018); in contrast, preference for animate entities seems considerably more tied to the visual modality (Goyal et al., 2006; He et al., 2013; Fairhall et al., 2014; Dormal et al., 2018). Recently, it has been proposed that this phenomenon might reflect a difference in transparency of mapping between high-level visual representations and their associated action computations across object domains (Bi et al., 2016). This conjecture originates from a theoretical framework which assumes that the object domain organization in the OTC is the result of evolutionary pressures to accommodate efficient mapping between locally-constrained visual representations and their associated downstream computations (Mahon & Caramazza, 2011). Within this framework, a relationship between visual representation and the downstream brain systems – particularly, the action/response system - is a major organizing principle for the OTC. Critically, this relationship is different for the inanimate and the animate domain. In the case of inanimate objects, mapping between visual shape or texture and functional relevance of an object is usually stable and transparent (e.g., elongation affords an action of gripping), which makes these visual qualities of inanimate objects relevant units for the action system. This privileged relationship between visual and action/motor representations might make the visual representation of inanimate objects less

tied to the visual modality and more accessible through different sensory systems. In the animate domain, the link between visual shape or texture and objects' functional relevance is usually much less articulated (e.g., similarly looking animals/humans might or might not pose a threat). The functional relevance of an animate object can be usually evaluated only after the categorization of the object as a whole (e.g., as friend/foe/food). This may have resulted in visual representation of animate entities that, at the stage of the OTC, is not transparently related to action or motor representations – a characteristic that might have made it less accessible through other sensory systems.

In the present study, we applied this theoretical insight to a recent contention concerning the responsiveness of the ventral OTC (vOTC) face processing area, the fusiform face area (FFA; Kanwisher et al., 1997; Kanwisher and Yovel, 2006), to non-visual face-related stimuli. A number of studies investigated whether the FFA shows functional preference for tactile, auditory or imagined information about the human face, relative to other types of stimulation in these modalities. When sighted subjects are explicitly 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 subjects not engaged in explicit imagery tasks and in congenitally and early blind subjects. Thus, no preferential FFA activation was reported when sighted and early blind subjects were listening to human voices pronouncing vowels (Dormal et al., 2018), and when sighted and congenitally blind subjects were touching 3-D models of faces (Goyal et al., 2006; a positive effect was found for late blind subjects). Furthermore, a null result for the FFA was reported when sighted subjects 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 significant increase in activation of the FFA when sighted and congenitally blind subjects were tactually identifying and comparing 3-D models of faces (Kitada et al., 2009; Murty et al., 2020). In the auditory modality, preferential FFA response was found in congenitally blind subjects listening to sounds of human facial actions (van den Hurk et al., 2017; Murty et al., 2020) or to verbal phrases conveying different emotional states (Fairhall et al., 2017). While differences between the tactile studies might perhaps be explained by differences in statistical power (n = 3 per subject group in Goyal et al., 2006, where an effect in the FFA of congenitally blind and sighted individuals was not reported), this explanation is not easily applicable to other studies discussed above. Here, in an extension of the shape-action mapping conjecture, we propose that non-visual responsiveness of the FFA is modulated by a transparency of coupling between visual representation and action system computations – specifically, representation of facial actions. This view predicts that representations of face shapes that transparently and stereotypically map onto specific types of facial motor actions (for example, face shape during laughing and the motor action of laughing) are more accessible through non-visual modality than other, static face representations. While this difference might be less clear in tactile studies, as tactile exploration can generate shape descriptions (e.g., elongations or curvatures) that naturally map onto similar visual descriptions, independently of visual-motor/action mapping, it should be clearly detectable in studies using sounds or words. In line with this prediction, auditory studies that found increased activation of the FFA used sounds of facial actions (van den Hurk et al., 2017; Murty et al., 2020) or phrases being indicative of the emotional state of a speaker (Fairhall et al., 2017) – that is, stimuli for which mapping between face shape and a specific type of facial motor actions is salient and systematic. In contrast, studies which reported null results for the FFA used stimuli for which the relationship between the visual shape

of a face and motoric actions might be less salient than for sounds of facial expressions or actions – either due to reduced information about the face shape (Dormal et al, 2018; note that speech sounds produced in a neutral tone convey information only about the mouth shape) or due to being indicative of a specific person rather than a type of a face representation (Fairhall et al., 2014).

To experimentally test our prediction, we designed a functional magnetic resonance imaging (fMRI) study in which congenitally blind and sighted subjects listened to object sounds and four categories of sounds produced by animate entities (see Methods). The first two animate categories included human sounds associated with either emotional or non-emotional facial expressions (sounds of laughing and crying vs. sounds of yawning and sneezing). The other two categories included speech sounds produced in a neutral tone and animal sounds. As discussed above, the visual shape of facial expressions, emotional or non-emotional, is stereotypically linked with specific types of facial motor actions. We thus predict that, in both groups, the FFA will show functional preference, relative to object sounds, for both categories of facial expression sounds included in the study – a result that would allow us to distinguish the FFA sensitivity to specific face shape representations from its potential sensitivity to emotional content in blind individuals (e.g., Fairhall et al., 2017). In contrast, we do not predict robust functional preference, neither in the FFA nor in other vOTC animate areas, for speech and animal sounds, for which the shape – motor action mapping is less salient than for facial expressions.

Results

Behavioral results

To keep the subjects attentive in the MRI scanner, we asked them to categorize presented sounds along the human/non-human dimension (see Methods). Both groups of subjects were highly

accurate in performing this task (mean accuracy ± SD, blind subjects: 95% ± 6%; sighted subjects: 92% ± 10%). No difference between groups or sound categories were detected; the interaction between these two factors was also not significant (Group x Sound Category ANOVA, all $p > 0.18$).

fMRI results: univariate analysis

We started the analysis of the fMRI data from performing omnibus F-tests, testing for main effects of sound category, group, and for interaction between these factors (see Methods). 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 (Table S1). When we increased the statistical threshold, in order to 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 (Fig. 1 – figure supplement 1). 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 posterior superior temporal sulcus (pSTS) (Fig. 1 – figure supplement 1; Table S2). 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 vs. object sounds

We first contrasted activation for sounds of every animate category with activation for object sounds in the blind group. As we expected, sounds of both emotional and non-emotional 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 occipital face area – OFA; Gauthier et al., 2000) (Fig. 1A-B; see also Table S3-4). Furthermore, sounds of emotional expressions only induced stronger activation in the amygdala (Fig. 1A; 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. 1A-B; Table S3-4). Contrasting the other animate categories - 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 (Fig. 1 – figure supplement 2; Table S5-6). However, in line with the shape – motor action mapping hypothesis, no reliable activation in the vOTC animate areas was observed for these animate sound categories (Fig. 1 – figure supplement 2; Table S5-6).

Facial expression sounds vs. speech and animal sounds

To directly test whether animate sound categories with different shape – motor action mapping transparency activate the vOTC to a different extent, we compared activation enhanced 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 location of the FFA and the OFA (Fig. 1A-B; Table S7-10). Stronger activation for facial expression sounds was also observed in dorsolateral visual areas, the right pSTS, the auditory cortex and frontal regions (Fig. 1A-B; Table S7-10). 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. 1A-B; Table S7-10).

Facial expression sounds vs. every other sound category (conjunction analysis)

To statistically test for areas showing functional selectivity for the facial expression sounds, relative to all other sounds used in the experiment, in the blind subjects, 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 Fig. 1C). 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. 1C). 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. 1C).

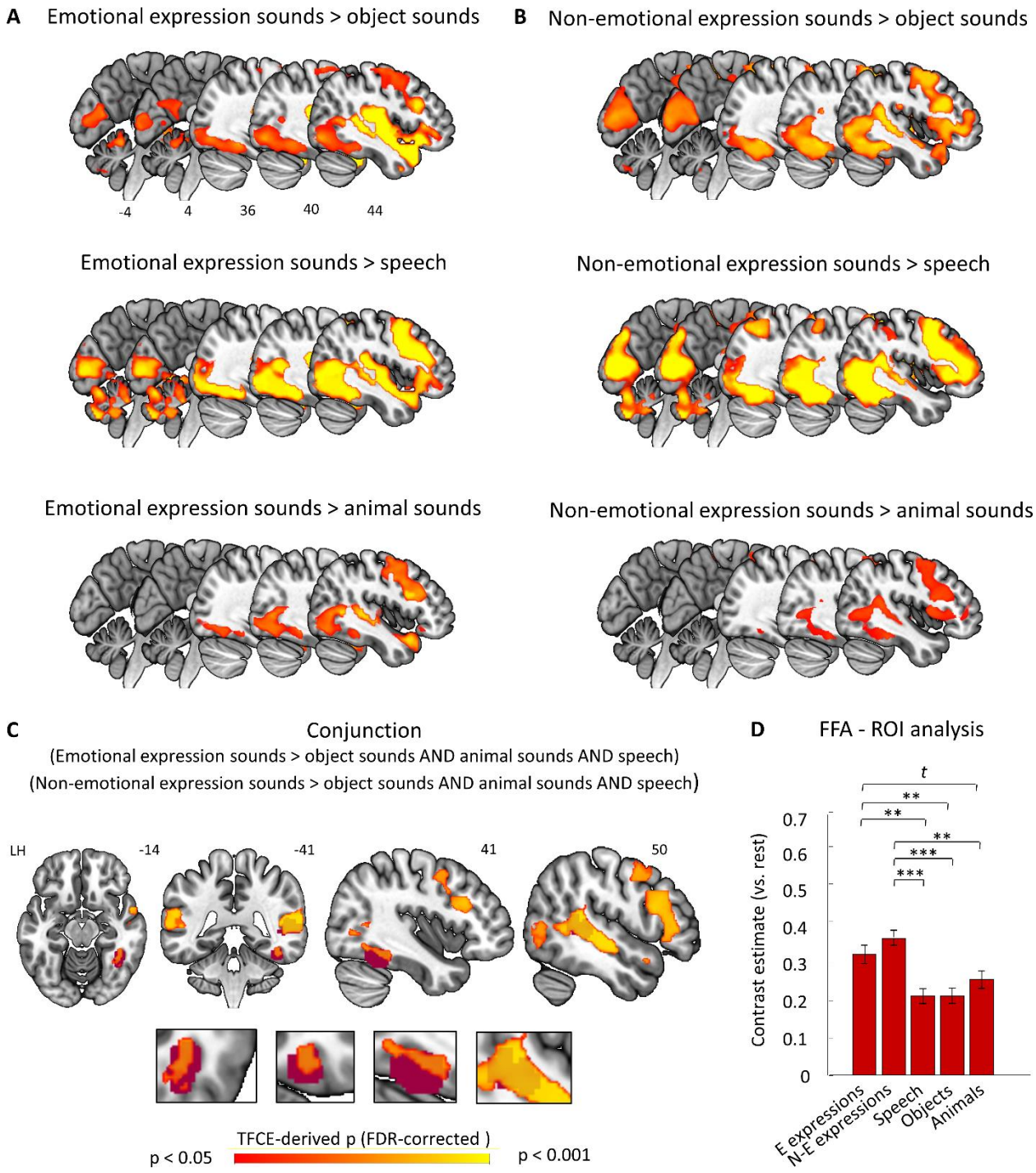


Figure 1. Functional selectivity for facial expression sounds in the fusiform gyrus of congenitally blind subjects. (A-B) Activation for the facial expression sounds relative to the activation for object sounds (upper panels), speech sounds (middle panels) and animal sounds (bottom panels). Results are presented separately for the expression sounds conveying emotional information (i.e., sounds of laughing and crying; panels A) and for the expression sounds that do not convey such information (i.e., sounds of yawning and sneezing; panels B). (C) Conjunction (i.e., logical AND analysis) of the six contrasts presented in panels A and B. Burgundy ROIs represent the typical location of the fusiform face area (FFA), as indicated by the Neurosynth meta-analysis of studies including the word "face" (896

studies included; map thresholded at Z = 15), and the typical location of the parts of the right posterior superior temporal sulcus (right pSTS) that are responsive to facial expressions, as indicated by an analogous metanalysis for the words "facial expressions" (250 studies included; map thresholded at Z = 4.8). Close-ups of the two ROIs, in the planes corresponding to the whole-brain illustrations, are provided at the bottom for clarity (from the left: the FFA ROI presented in the axial, coronal and sagittal plane, the right pSTS ROI presented in the sagittal plane). (D) Mean contrast estimates for each sound category in the FFA ROI (E Expressions – emotional expressions; N-E expressions – non-emotional expressions). Statistical thresholds: (A-C) For all analyses presented, Threshold-Free Cluster Enhancement (TFCE) values were calculated for every voxel within a broad grey matter mask, created based on a standard brain tissue probability map (all voxels in the brain image with a 20% and higher probability of belonging to the grey matter were included). The statistical significance of these values was then assessed using permutation testing (10000 iterations) and corrected for multiple comparisons using False Discovery Rate (FDR). In the conjunction analysis, color hues represent mean FDR-corrected p-values of the six contrasts included. These values were calculated only for voxels that passed the conjunction test at $p < 0.05$, FDR-corrected (i.e. only for voxels that were significant at $p < 0.05$, FDR-corrected, in all six contrasts included). (D) ** $p < 0.01$, *** $p < 0.001$, $t = 0.09$, FDR-corrected for multiple comparisons. Error bars represent standard error of the mean, adjusted to adequately represent within-subject variability across conditions using a method proposed by Cousineau (2005). Numbers in panels A and C denote MNI coordinates of presented axial, coronal or sagittal planes. Individual contrast estimates for each sound category in the FFA, analyzed in the ROI analysis, are available in Figure 1 and 2 – source data 1

Sighted group

Animate sound categories vs. 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 subjects (Fig. 2A-B; Table S11-12). Preferential activation for the facial expression sounds was also detected in frontal regions, although in the sighted subjects this effect was constrained primarily to the inferior frontal gyrus (Fig. 2A-B; Table S11-12). Intriguingly, and in contrast to the results for the blind group, no preferential response in the lateral fusiform gyrus was observed in the sighted subjects, neither for the emotional nor for the non-emotional facial expression sounds (Fig. 2A-B; Table S11-12). 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. 2A; Table S11).

However, similar effects were also observed in comparisons between speech sounds and object sounds, and, in the case of cuneus, between animal sounds and object sounds (see Fig. 2 – figure supplement 1; Table S13-14).

Facial expression sounds vs. speech and animal sounds

Direct comparisons between the facial expression sounds and the other animate sound categories in the sighted subjects showed significant effects in the auditory cortex, the right pSTS and frontal regions; however, no significant differences were detected in the visual cortex (Fig. 2A-B; Table S15-18).

Facial expression sounds vs. every other sound category (conjunction analysis)

The conjunction analysis of all six contrasts between the facial expression sounds and the other sound categories in the sighted subjects yielded significant effects in the left and the right auditory cortex, the left and the 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. 2C).

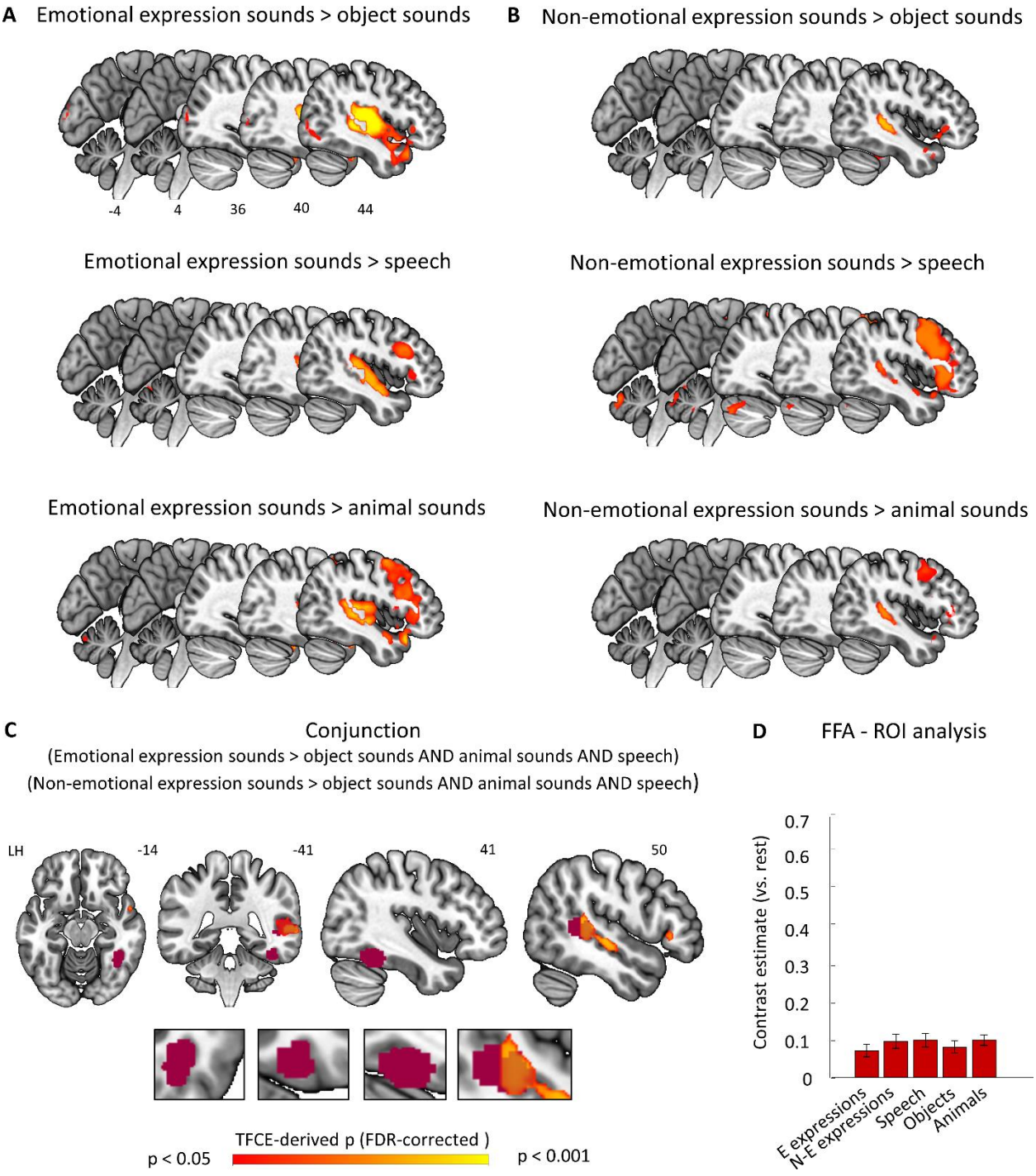


Figure 2. Lack of preferential univariate activation for facial expression sounds in the fusiform gyrus of sighted subjects. All analysis parameters and ROI definition procedures were identical to those applied to the data from the blind group (see Figure 1 for details). (A-B) Activation for the facial expression sounds relative to the activation for object sounds (upper panels), speech (middle panels) and animal sounds (bottom panels). Results are presented separately for the expression sounds conveying emotional information (i.e., sounds of laughing and crying; panels A) and for the expression sounds that do not convey such information (i.e., sounds of yawning and sneezing; panels B). (C) Conjunction (i.e., logical AND analysis) of the six contrasts presented in panels A and B. Burgundy ROIs

represent the typical location of the fusiform face area (FFA) and the typical location of the parts of the right posterior superior temporal sulcus (right pSTS) that are responsive to facial expressions, as determined by Neurosynth meta-analyses for the words “face” and “facial expressions”, respectively. Close-ups of the two ROIs, in the planes corresponding to the whole-brain illustrations, are provided at the bottom for clarity (from the left: the FFA ROI presented in the axial, coronal and sagittal plane, the right pSTS ROI presented in the sagittal plane). (D) Mean contrast estimates for each sound category in the FFA ROI (E Expressions – emotional expressions; N-E expressions – non-emotional expressions). Error bars represent standard error of the mean, adjusted to adequately represent within-subject variability across conditions using a method proposed by Cousineau (2005). Numbers in panels A and C denote MNI coordinates of presented axial, coronal or sagittal planes. Individual contrast estimates for each sound category in the FFA, analyzed in the ROI analysis, are available in Figure 1 and 2 – source data 1

Between-group comparisons

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, the interaction 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 subjects (Fig. 3; Table S19). 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. 3; Table S19). An inverse contrast, testing for brain areas showing stronger functional preference for facial expression sounds in the sighted group, did not yield any significant results.

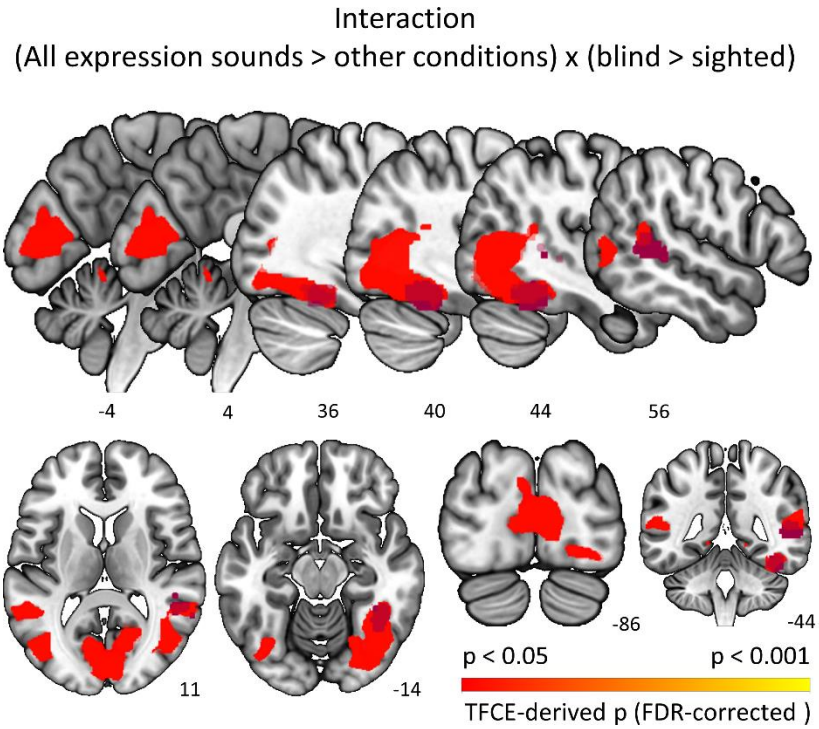


Figure 3. Direct test of between-group differences in functional preference for facial expression sounds. Brain regions preferentially activated by expression sounds (averaged activation for emotional and non-emotional expressions), relative to other experimental conditions (averaged activation for speech, object sounds and animals sounds), in the blind subjects but not (or not to the same extent) in the sighted subjects. All analysis parameters and ROI definition procedures were identical to those that were applied to within-group comparisons (see Figure 1 for details). Burgundy ROIs represent the typical location of the fusiform face area and the typical location of the parts of the right posterior superior temporal sulcus that are responsive to facial expressions, as determined by Neurosynth meta-analyses for the words "face" and "facial expressions", respectively. Numbers denote MNI coordinates of presented axial, coronal or sagittal planes.

ROI analysis in the fusiform face area

Finally, to confirm spatial correspondence between the effect observed for facial expression sounds in the blind subjects and the typical anatomical locus of the FFA, we performed a region-of-interest (ROI) analysis, in which the FFA ROI was defined based on the meta-analysis of face perception studies (i.e., Neurosynth meta-analysis of studies including the word "face"; Yarkoni et al., 2011; see Methods). In a 2 (Group) x 5 (Sound Category) ANOVA, we detected 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 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. 1D). In contrast, no significant differences between conditions were detected in the sighted group (Fig. 2 D). 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 non-emotional 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.

Responses in the primary auditory cortex

Can the fusiform gyrus response pattern in the blind subjects be driven by differences in acoustic features or acoustic complexity across sound categories? To investigate this possibility, we performed the ROI analysis in the primary auditory cortex in both groups (Fig. 3 – figure supplement 1). We observed activation pattern that was the same for the blind and the sighted subjects, and markedly different from the pattern that was observed in the fusiform gyrus in the blind group (see Fig. 1D). This suggests that the effects detected in this region in the blind subjects were not driven by the differences in acoustic characteristics of sounds, as such characteristics are likely to be captured by activation patterns of the primary auditory cortex.

<u>fMRI results: multivariate analysis</u>

The univariate analysis of the fMRI data revealed strong functional selectivity to facial expression sounds in the typical anatomical location of the FFA in the blind subjects. In contrast, this area did not show any univariate differentiation of presented sounds in the sighted group. Nonetheless, it is still possible that the FFA in the sighted subjects represents the difference between the facial expression sounds and the other sound categories by the means of disperse, subthreshold activation patterns. To investigate this possibility, we performed a multi-voxel pattern analysis (MVPA) in this area (see Methods).

Analogously to the univariate analysis, we first asked which categories of animate sounds can be distinguished from object sounds, based on the lateral fusiform gyrus activation patterns in each group. We found that an MVPA classifier, trained on the activation of this area (see Methods), was able to successfully differentiate between expression sounds and object sounds, in both blind and sighted subjects (Fig. 4A-B). Successful classification of sounds was also achieved for the speech - object sound pair (blind subjects: mean classification accuracy = 62 %, SEM = 2.4 %, $p_{FDR} < 0.001$; sighted subjects: mean classification accuracy = 55 %, SEM = 1.3 %, $p_{FDR} = 0.001$) but not for the animal - object sound pair (blind subjects: mean classification accuracy = 52 %, SEM = 1.6 %, $p_{FDR} = 0.093$; sighted subjects: mean classification accuracy = 52 %, SEM = 1.4 %, $p_{FDR} = 0.093$).

We then tested whether the activation of the lateral fusiform gyrus can be used to distinguish between animate sound categories with different shape - motor action mapping transparency (facial expression sounds – high transparency; speech an animal sounds – weak transparency). Indeed, we observed a successful classification of sounds into facial expression sounds and speech sounds, in both groups and in the case of both types of facial expressions (Fig. 4A-B). Additionally,

the MVPA classifier was able to distinguish between the facial expression sounds and animal sounds, in the blind, but not in the sighted subjects (Fig. 4A-B).

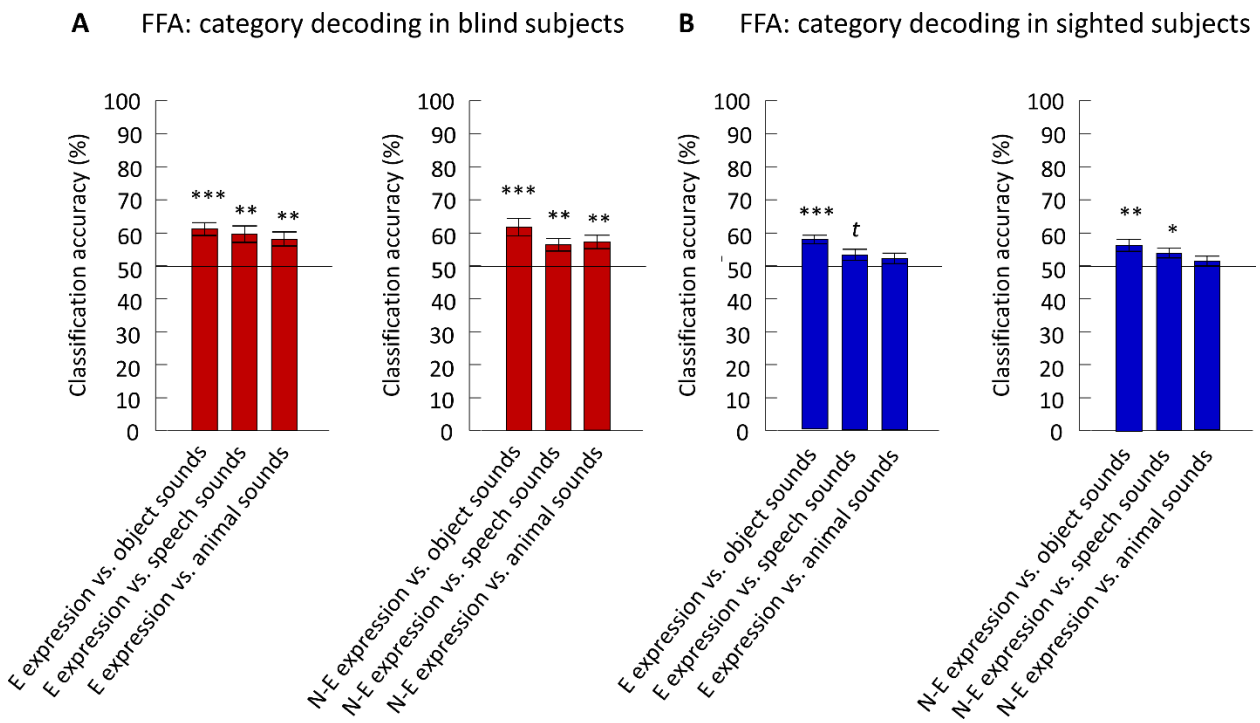


Figure 4. Facial expression sounds induce distinctive activation pattern in the fusiform gyrus, in both blind and sighted subjects. (A-B) Mean accuracy of a support vector machine classifier distinguishing between facial expression sounds and every other sound category included in the experiment, based on the activation patterns in the typical location of the fusiform face area (FFA). Results are presented separately for the blind (A) and the sighted (B) subjects and for the emotional expression sounds (E expression sounds; left panels) and the non-emotional expression sounds (N-E expression sounds, right panels). The FFA was defined based on the Neurosynth meta-analysis for the word "face" (896 studies included; the resulting map thresholded at Z = 10; for the purpose of the multivariate analysis, voxels in both the right and the left hemisphere were included, see Methods). Testing against chance was performed with one-tailed one-sample t-test. Statistical thresholds: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, $t = 0.057$, corrected for multiple comparisons across all comparisons performed using the false discovery rate. Error bars represent standard error of the mean. Chance classification level is marked with black line. Individual classification accuracy values are available in Figure 4 – source data 1.

Does the typical location of the FFA code the differences between specific facial expressions presented through the auditory modality? To answer this question, we trained the MVPA classifier to distinguish, based on the activation of this area, between four facial expressions presented in

our study (crying, laughing, sneezing and yawning), irrespectively of the differences in voice characteristics and in the gender of the two actors (man and woman; see Methods). Indeed, the sounds were successfully classified into specific facial expressions in the blind group (Fig. 5A-B). In contrast, no significant effects were observed in this group for the classification of speech sounds, irrespectively of the gender of the two speakers, or for the classification of the gender of the two speakers, irrespectively of the specific facial expression or speech sound (Fig. 5A). In the sighted group, none of these analyses yielded significant results (Fig. 5A).

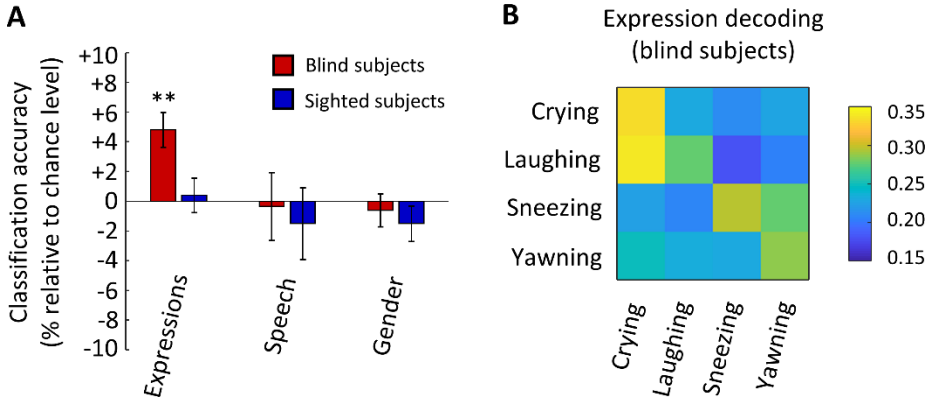


Figure 5. Activation patterns in the fusiform gyrus of blind subjects code differences between specific facial expressions presented through the auditory modality. (A) Mean accuracy of a support vector machine classifier distinguishing, based on the activation in the typical location of the fusiform face area (FFA), between: four specific facial expressions (crying, laughing, sneezing and yawning), irrespectively of the actor; two specific speech sounds, irrespectively of the actor; and the actor (man and woman), irrespectively of the specific facial expression or the speech sound. The presented accuracy is adjusted for the chance level (25 % for facial expression classification; 50 % for speech sound and gender classification). The FFA was defined based on the Neurosynth meta-analysis for the word "face" (896 studies included; the resulting map thresholded at Z = 10; for the purpose of the multivariate analysis, voxels in both the right and the left hemisphere were included, see Methods). Testing against chance was performed with one-tailed one-sample t-test. (B) A confusion matrix representing the classifier predictions for each facial expression sound in the blind subjects. Rows of the confusion matrix indicate the facial expression sound presented during the experiment and columns indicate classifier's predictions for this sound. The classifier's predictions are represented by color hues, with warmer colors for higher proportion of predictions and colder colors for lower proportion of predictions. Statistical thresholds: (A) ** $p < 0.01$, corrected for multiple comparisons across all comparisons performed using the false discovery rate. Error bars represent standard error of the mean. Individual classification accuracy values are available in Figure 5 – source data 1.

Discussion

In this study, we found that the sounds associated with specific 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 and animal sounds and human speech. In contrast to facial expression sounds, speech sounds or animal sounds did not yield preferential activation, relative to object sounds, in the vOTC animate areas in the blind group. In the sighted subjects, no clear functional preference in the vOTC was observed for any of the animate sound categories. However, the multivariate analysis showed that the expected distinctions between facial expression sounds and other sound categories – particularly, between facial expression sounds and the object/speech sounds – are represented in the lateral fusiform gyrus in both groups, with the representation in blind subjects being precise enough to differentiate even between specific expressions.

In the visual modality, the FFA responses are known to be the strongest for human faces. However, responses to many other animate categories, such as animals' heads, whole animals, whole humans or human bodies without heads were reported to be still stronger than responses to inanimate objects (Kanwisher et al., 1999). In our study, we observed a dichotomous pattern of results, with both types of facial expression sounds activating the FFA in blind subjects more strongly than the other categories, and the activation for speech sounds and animal sounds being not significantly different from the activation for object sounds. This clear dissociation between effects induced in the FFA of blind individuals by facial expression sounds and by other animate sound categories bring a new perspective to a recent debate concerning this area's sensitivity to auditory and tactile stimulation. Several studies have already shown that this brain area can be preferentially activated by auditory and tactile stimuli related to the human face, compared to other types of information

conveyed through these sensory modalities (blind subjects: van den Hurk et al., 2017; Murty et al., 2020; sighted subjects: Kitada et al., 2009). However, other studies investigating this issue reported null results, and a theoretical explanation of this difference has been lacking (blind and sighted subjects: Goyal et al., 2006; Dormal et al., 2018). Here, in an extension of the conjecture formulated by Bi and colleagues (Bi et al., 2016), we proposed that these conflicting findings reflect the fact that various face shape representations are differently associated with specific types of motor actions. Indeed, this difference is particularly clear in recent auditory studies: null effect in the fusiform gyrus of blind subjects was reported in a study using vowels spoken in a neutral tone – a stimuli for which a mapping between face shape and associated motor actions, while decodable for the mouth region, is not particularly salient (Dormal et al., 2018); in contrast, a significant increase in the FFA activation was reported for sounds of facial actions (such as laughing, chewing, blowing a kiss or whistling; van den Hurk et al., 2017) and for verbal phrases conveying information about a speaker's emotional state (Fairhall et al., 2017) – that is, stimuli for which a link between face shape and specific types of motor actions (of laughing, chewing, screaming, and so on) is salient and stereotypical. In line with these previous results, our study shows a clear preference of the FFA in the blind for facial expression sounds. This confirms our initial hypothesis and suggests that the transparency of mapping between visual and motor representations guides the auditory responsiveness of the FFA. Furthermore, our findings contribute to a growing body of evidence that visual experience is not necessary for the development of functional preference for face-related information in the human fusiform gyrus (van den Hurk et al., 2017; Murty et al., 2020). In contrast, we did not observe a modulation of the FFA response by emotional content in blind individuals (e.g., Fairhall et al., 2017; see Methods

and Table S20 for ratings of the sounds' emotional content, provided by an independent group of subjects).

On a more general level, our study contributes to the understanding of the nature of a perplexing phenomenon – that is, the sensory modality by object domain interaction in the 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 visual and action representations for the animate domain is less transparent than for inanimate objects (Bi et al., 2016). In this study, we demonstrated the predictive value of this hypothesis by showing that effects characteristic of the inanimate domain can be also observed for animate entities with stable and stereotypical relationship between visual and action/motor computations. Thus, our findings suggest that the OTC organization is determined not only by local sensitivity to certain visual features, but also by metrics computed downstream in the brain, particularly in the action system.

An unexpected finding in our study is the clear difference in vOTC univariate response to facial expression sounds across the congenitally blind and the sighted group. This seems to be inconsistent with results from the inanimate domain, which are remarkably similar for these two groups (He et al., 2013; Peelen et al., 2013; Wang et al., 2015). Nevertheless, the multivariate analysis showed that the expected distinctions between sound categories – particularly, between facial expression sounds and object sounds as well as between facial expression sounds and speech sounds – are represented in the vOTC face areas in both groups. Based on this result, and on our observation of generally weaker response to sounds in the FFA of the sighted (see ROI analysis), we think that the adaptation of the FFA circuitry to the strength of incoming signals is a plausible explanation of the observed between-group difference. In sighted individuals, the FFA receives strong feedforward inputs from early visual cortices, which might make this region relatively

insensitive to small differences between inputs from other brain systems. In congenitally blind individuals, lack of feedforward signals might lead to the adjustment of the FFA sensitivity – as a result, this area might be able to capture the subtle differences induced by auditory stimuli. Such an adjustment of sensitivity, following lack of sight, might be weaker for the inanimate representation, computed in other parts of the vOTC, as this representation is generally more available to information from other brain systems (see discussion above). In agreement with our findings, the two previous studies that showed univariate activation in the fusiform gyrus of blind individuals, induced by auditory stimulation, failed to find similar, robust effects in this region in sighted individuals (Fairhall et al., 2017; van den Hurk et al., 2017). Overall, while the difference between the blind and the sighted subjects was not predicted in our study, it does not affect the theoretical implications of our results, described above – unless some form of functional repurposing in the fusiform gyrus in the blind group is assumed (e.g., Fairhall et al., 2017), for which our study provides no convincing evidence.

In conclusion, we show that auditory responsiveness of the lateral fusiform gyrus (the typical anatomical location of the FFA) in congenitally blind individuals can be explained in terms of the type of evoked face shape representation and its relationship with motor representation. 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.

Methods

Subjects

Twenty congenitally blind subjects (13 males, 7 females, mean age ± SD = 46.25 ± 13.34 y, average length of education ± SD = 9.4 ± 3.9 y) and 22 sighted subjects (15 males, 7 females,

mean age ± SD = 45.91 ± 10.41 y, average length of education ± SD = 9.41 ± 2.09 y) participated in the study. Prior to the data collection, we decided to recruit approximately 20 subjects per group. Based on previous studies (e.g., He et al., 2013; Peelen et al., 2013; van den Hurk et al., 2017), we expected such a sample size to be more than sufficient to study auditory-induced effects in the OTC. The groups were matched for age, gender, handedness, and years of education (all $p > 0.25$). All subjects were native Mandarin speakers and had no history of neurological disorders. All sighted subjects had normal or corrected-to-normal vision. Detailed characteristics of each blind subject are provided in Table 1. All experimental protocols were approved by institutional review board of Department of Psychology Peking University, China (2015/05/04), as well as by the institutional review board of Harvard University (IRB15-2149), in accordance with the Declaration of Helsinki. Informed consent was obtained from each participant prior to study.

Table 1. Characteristics of blind subjects.

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 anophthalmos	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 anophthalmos	None
CB16	59	F	9	R	Congenital eye's tumor disease and dysplasia	None
CB17	54	F	9	R	Congenital microphthalmia	None
CB18	59	M	9	R	Congenital anophthalmos and cataract	None
CB19	23	M	11	R	Unknown	None
CB20	30	M	3	R	Congenital glaucoma and cataract	None

fMRI experiments

In the fMRI experiment, subjects were presented with sounds belonging to five categories: emotional facial expressions (laughing and crying), non-emotional facial expressions (sneezing and yawning), speech sounds (a single Chinese character repeated three times in a neutral tone;

two different characters were used), 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 included three unique sound exemplars (e.g., three different exemplars of a male actor laughing or three different exemplars of a horse neighing). All sounds were 2-s long and matched for sound intensity (root mean square value: -20 dB). Sounds for the fMRI experiment were chosen from a larger dataset based on behavioral ratings of their emotional content (or lack thereof) and recognizability of depicted actions and sound categories, obtained from separate groups of sighted Chinese subjects (see Table S20 for the ratings for the sounds used in the fMRI study; demographic data of the raters: 11 males, 11 females, mean age ± SD = 24.05 ± 2.84, mostly students of Beijing Normal University). All human sounds were recorded specifically for this study. Object and animal sounds were downloaded from the internet.

The timing of the sound presentation in the fMRI scanner was kept the same across subject 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 exemplars' presentation within each block was randomized in order to modulate purely auditory properties across blocks belonging to a given subcategory. The subjects' task was to respond to each sound and categorize it along the human/non-human dimensions (i.e., to answer whether a sound was produced by a human or non-human entity). For each sound exemplar, a response was recorded for 2.5 s following sound presentation's onset.

Subjects participated in eight experimental runs. Each run included two blocks for every subcategory (e.g., male actor laughing), resulting in eight blocks for each category (e.g., emotional facial expressions) per run. Thus, 64 blocks for each category and 16 blocks for each subcategory were presented, in total. The blocks were interspersed with rest periods with no stimuli presented (from 0- to 18-s long) so that each run lasted 6.8 min. The order of each category presentation and the length of rest periods were determined using Optseq (https://surfer.nmr.mgh.harvard.edu/optseq/). A separate Optseq sequence was created for each run and the order of runs was randomized across subjects. Within each category of stimuli, the order of blocks belonging to specific subcategories was randomized with the rule that a given subcategory cannot be repeated two times in a row.

The stimulus presentation was controlled with a program written in Psychopy 3.0.12b (Peirce, 2007). The sounds were presented through MRI-compatible headphones. Subjects gave their answers by pressing buttons on a response pad with the index and the middle finger (e.g., the index finger – human sounds; the middle finger – non-human sounds), and a response mapping to a given finger was counterbalanced across subjects. Before the actual fMRI experiment, each subject completed a short training session outside the scanner, in order to ensure that the task was clear and that he/she can classify stimuli correctly.

MRI 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 multi-slice 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; multi-band 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 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 subjects, sighted subjects) x sound category (emotional facial expressions, non-emotional facial expressions, speech, objects, animals) ANOVA. Analysis were performed in SPSS 25 (IBM). The reaction time was not analyzed since (1) the primary goal of the behavioral task was to verify whether subjects were attentive in the scanner, for which the accuracy is a better measure; (2) The time needed to categorize a given sound as belonging to a specific category is likely to depend not only on a subject 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.

MRI data analysis

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

mri_deface script (https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface; Bischoff-Grethe et al., 2007) The preprocessing was performed using 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 subject 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 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 subject's head motion larger than 0.5 mm were marked); (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; (8) Spatial smoothing with Gaussian kernel at two levels: 8 mm FWHM for the univariate analysis and 2 mm FWHM for the multivariate analysis (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 subject. 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, non-emotional expression sounds, speech sounds, object sounds, animal sounds). For the multivariate analysis, the data smoothed with 2-mm FWHM kernel were modelled at the level of specific subcategories (twenty experimental

conditions: five categories x 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 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 beta estimates and contrast images were computed for each experimental condition (i.e., each category in the first model, each subcategory in the second model), relative to rest periods.

For the univariate group analysis, the contrast images for five sound categories vs. rest periods, for both groups, were entered into a random-effect 2 (Group) x 5 (Sound Category) ANOVA model, created using the 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 – since 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 vs. 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 pairwise comparisons. For each test at the group level, Threshold-Free Cluster Enhancement (TFCE; Smith and Nichols, 2009) values were calculated, with a standard parameters (E = 0.5, H = 2; see Smith and Nichols, 2009), for every voxel within a broad grey matter mask, created based on canonical brain tissue probability maps implemented in SPM (all voxels in the brain image with

a 20% and higher probability of belonging to the grey matter were included). The statistical significance of these values was then assessed using permutation testing (10000 iterations with labels of compared conditions randomized) and corrected for multiple comparisons using the false discovery rate (FDR; Benjamini and Hochberg, 1995). The calculation of these measures 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).

For the univariate ROI analysis, the FFA was defined using a statistical association map produced by an automatic Neurosynth meta-analysis (https://neurosynth.org/) of studies including the word "face". The logic behind Neurosynth association maps is to 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). The performed meta-analysis included 896 studies and the association map was thresholded at Z-value = 15, resulting in a cluster in the right fusiform gyrus consisting of 227 voxels (functional data resolution; cluster center of mass MNI coordinates: 41 -49 -20). MarsBaR SPM toolbox (http://marsbar.sourceforge.net/) was used to extract an average contrast estimate for all voxels in the FFA ROI, for each subject and experimental condition. Analogously to the univariate whole-brain analysis, these values were then entered into a random-effect 2 (Group) x 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 three analysis steps – that is, for pairwise comparisons within each group (10 comparisons across sound categories per group; see Fig. 1D and Fig. 2D) as well as for pairwise comparisons between groups (5 comparisons).

For the multivariate analysis, the FFA was defined using the same Neurosynth association map (i.e., target word: "face"; 896 studies included) thresholded at Z-value = 10. The ROI was defined bilaterally (936 functional voxels included in total; 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 and bilateral ROI definition was that the multivariate analysis relies on the disperse and subthreshold activation and deactivation patterns, which might be well represented also by cross-talk between hemispheres (for example, a certain subcategory might be represented by activation of the right FFA and deactivation of the left analog of this area). The multivariate analysis was performed on single-subject beta estimates for each sound subcategory vs. rest periods using CoSMoMVPA (v. 1.1.0; Oosterhof et al., 2016). A support vector machine, as implemented in LIBSVM (v. 3.23; Chang and Lin, 2011), was used to perform data classification. 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. Obtained classification accuracy values were tested against chance level using one-tailed one-sample t-tests, with the FDR correction applied when appropriate (see Figures 4-5).

Acknowledgements

This work was supported by a Polish Ministry of Science grant (DN/MOB/023/V/2017) and a Kosciuszko Foundation fellowship to Ł.B., the National Natural Science Foundation of China (31925020, 31671128 to Y.B.), Changjiang Scholar Professorship Award (T2016031 to Y.B.), the 111 Project (BP0719032 to Y.B.), and the Fundamental Research Funds for the Central Universities.

Competing interests

The authors declare that no competing interests exists.

References

Bedny M (2017) Evidence from Blindness for a Cognitively Pluripotent Cortex. Trends Cogn Sci 21:637–648.

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.

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, Brown GG, Fischl B (2007) A technique for the deidentification of structural brain MR images. Hum Brain Mapp 28:892–903.

Chang CC, Lin CJ (2011) LIBSVM: A Library for support vector machines. ACM Trans Intell Syst Technol 2:27.

Cousineau D (2005) Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. Tutor Quant Methods Psychol, 1: 42-45.

Dormal G, Pelland M, Rezk M, Yakobov E, Lepore F, Collignon O (2018) Functional Preference for Object Sounds and Voices in the Brain of Early Blind and Sighted Individuals. J Cogn Neurosci 30:86–106.

Fairhall SL, Anzellotti S, Ubaldi S, Caramazza A (2014) Person- and Place-Selective Neural Substrates for Entity-Specific Semantic Access. Cereb Cortex 24:1687–1696.

Fairhall SL, Porter KB, Bellucci C, Mazzetti M, Cipolli C, Gobbini MI (2017) Plastic

reorganization of neural systems for perception of others in the congenitally blind. Neuroimage 158:126–135.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995) Analysis of fMRI time-series revisited. Neuroimage 2:45–53.

Gardumi A, Ivanov D, Hausfeld L, Valente G, Formisano E, Uludağ K (2016) The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. Neuroimage 132:32–42.

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform "face area" is part of a network that processes faces at the individual level. J Cogn Neurosci, 12:495-504.

Goyal MS, Hansen PJ, Blakemore CB (2006) Tactile perception recruits functionally related visual areas in the late-blind. Neuroreport 17:1381–1384.

He C, Peelen M V, Han Z, Lin N, Caramazza A, Bi Y (2013) Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. Neuroimage 79:1–9.

Ishai A, Ungerleider LG, Haxby J V. (2000) Distributed neural systems for the generation of visual images. Neuron 28:979–990.

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.

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.

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.

Livingstone MS, Arcaro MJ, Schade PF (2019) Cortex Is Cortex: Ubiquitous Principles Drive Face-Domain Development. Trends Cogn Sci 23:3–4.

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.

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. bioRxiv:2020.02.25.964890.

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 J V. (2016) CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front Neuroinform 10.

Peelen M V., 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.

Peirce JW (2007) PsychoPy-Psychophysics software in Python. J Neurosci Methods 162:8–13.

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–

98.

van den Hurk J, Van Baelen M, Op de Beeck HP (2017) Development of visual category selectivity in ventral visual cortex does not require visual experience. Proc Natl Acad Sci U S A 114:E4501–E4510.

Wang X, Peelen M V, 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.

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.

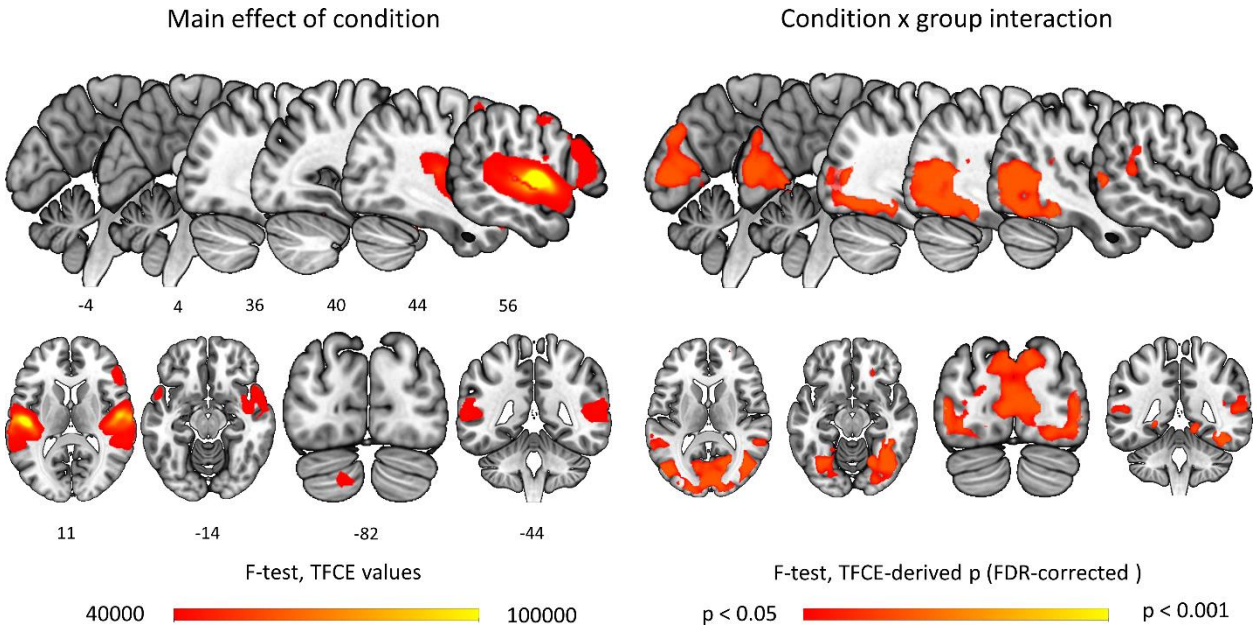


Figure 1 – figure supplement 1. Results of omnibus F-tests. Testing for the main effect of group did not yield significant results. Statistical thresholds: Initially, all analysis parameters were identical to those applied to the data reported in the main text (i.e., statistical threshold was set at false-discovery-rate corrected $p < 0.05$, derived based on calculation of TFCE values and permutation testing; see Figure 1 for details). In the case of the testing for the main effect of condition, this yielded significant effects in wide network of regions, including frontoparietal regions, the temporal lobe, the OTC, and the early visual cortex (see Table S1). To achieve a higher degree of spatial specificity, the statistical threshold for this analysis was increased to arbitrarily chosen TFCE value (i.e., 40000).

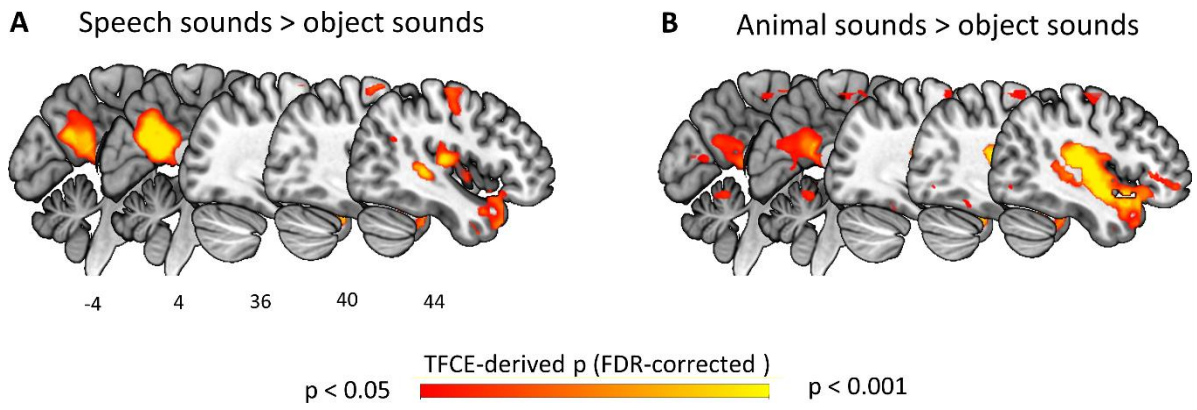
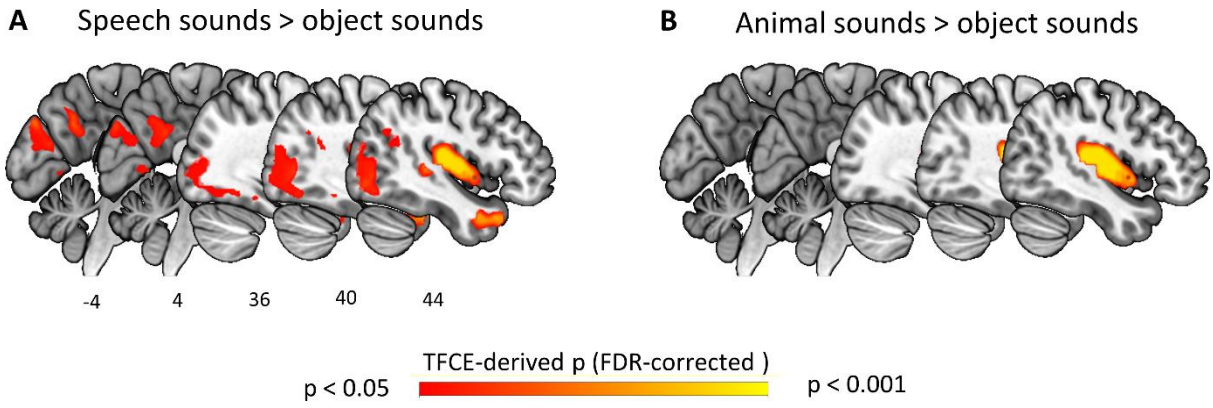


Figure 1 – figure supplement 2. Preferential activation induced by speech sounds (A) or animal sounds (B), relative to object sounds, in blind subjects. All analysis parameters were identical to those applied to the data reported in the main text (see Figure 1 for details).



757

758 **Figure 2 – figure supplement 1. Preferential activation induced by speech sounds (A) or animal sounds (B), relative**
759 **to object sounds, in sighted subjects.** All analysis parameters were identical to those applied to the data reported
760 in the main text (see Figure 1 for details)

761

762

763

764

765

766

767

768

769

770

771

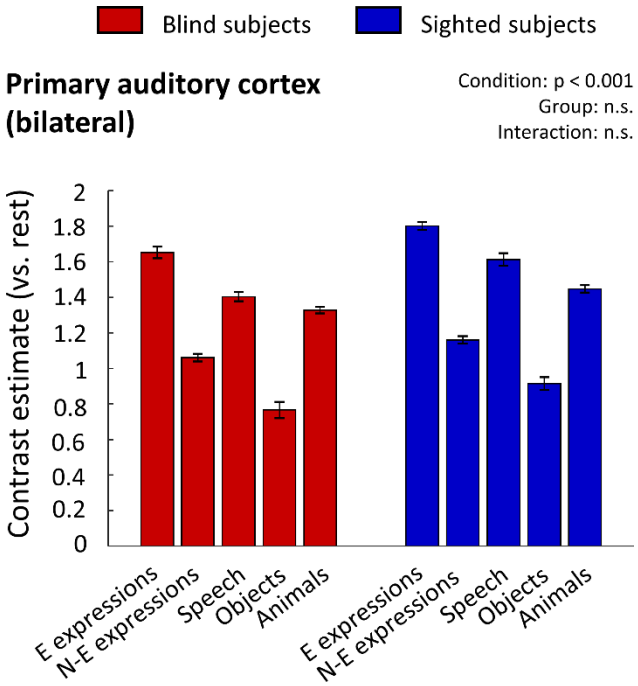


Figure 3 – figure supplement 1. Region-of-interest (ROI) analysis of the univariate activation patterns in the primary auditory cortex Mean contrast estimates, calculated relative to rest periods, are presented for each condition and group for the primary auditory cortex. The ROI was defined bilaterally, based on the Neurosynth metanalysis of studies including the word "primary auditory" (114 studies included; map thresholded at Z = 10; ROI size = 672 functional voxels). Condition x Group ANOVA showed a significant effect of condition ($p < 0.001$; all other ANOVA effects were not significant). Responses to all conditions are significantly different from each other (all $p < 0.001$). Error bars represent standard error of the mean, adjusted to adequately represent within-subject variability across conditions, in each group, using a method proposed by Cousineau (2005).