



Plastic reorganization of neural systems for perception of others in the congenitally blind



S.L. Fairhall^{a,*}, K.B. Porter^b, C. Bellucci^c, M. Mazzetti^c, C. Cipolli^c, M.I. Gobbini^{c,d,*}

^a Center for Mind/Brain Sciences, University of Trento, Italy

^b Department of Psychology, Harvard, Cambridge, MA, USA

^c Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale (DIMES), Medical School, University of Bologna, Bologna, Italy

^d Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

ARTICLE INFO

Keywords:

Emotion
Fusiform face area
Superior temporal sulcus
Language
fMRI
Decoding

ABSTRACT

Recent evidence suggests that the function of the core system for face perception might extend beyond visual face-perception to a broader role in person perception. To critically test the broader role of core face-system in person perception, we examined the role of the core system during the perception of others in 7 congenitally blind individuals and 15 sighted subjects by measuring their neural responses using fMRI while they listened to voices and performed identity and emotion recognition tasks. We hypothesised that in people who have had no visual experience of faces, core face-system areas may assume a role in the perception of others via voices. Results showed that emotions conveyed by voices can be decoded in homologues of the core face system only in the blind. Moreover, there was a specific enhancement of response to verbal as compared to non-verbal stimuli in bilateral fusiform face areas and the right posterior superior temporal sulcus showing that the core system also assumes some language-related functions in the blind. These results indicate that, in individuals with no history of visual experience, areas of the core system for face perception may assume a role in aspects of voice perception that are relevant to social cognition and perception of others' emotions.

1. Introduction

There is extensive work on the neural systems for face and voice recognition (Gobbini and Haxby, 2007; Haxby et al., 2000; Tsao et al., 2008; Yovel and Belin, 2013). Voices, as compared to non-vocal sounds, activate Temporal Voice Areas (Belin et al., 2004; Pernet et al., 2015), which include the middle and anterior superior temporal sulcus (maSTS), the superior temporal gyrus (STG), as well as the insula and the prefrontal cortex (Ethofer et al., 2012; Leaver and Rauschecker, 2010; Moerel et al., 2012; Remedios et al., 2009; Romanski et al., 2005). The distributed system for face perception in humans (Fairhall and Ishai, 2007; Gobbini and Haxby, 2007; Haxby et al., 2000, 2002; Haxby and Gobbini, 2011; Ishai et al., 2005) includes visual areas involved in the perception of invariant visual attributes diagnostic of identity and of changeable aspects diagnostic of facial expression and direction of attention (the 'core system'), and some additional areas that play a role in extracting information, such as person knowledge, associated with faces ('the extended system').

The two networks for voice and face processing are largely distinct with anatomical overlap confined to the posterior STS/STG regions (Watson et al., 2014a, 2014b). The interdependence between these networks is uncertain. The function of the core system areas for face perception, however, appears to extend beyond face-processing to a more general and abstract role in social cognition and perception of others (Gobbini et al., 2011). For example, viewing point-light displays depicting biological motion or geometric shapes depicting social interactions, neither of which depicts biological form, activates both the face selective region of the lateral fusiform gyrus (the 'fusiform face area', FFA) and posterior superior temporal sulcal (pSTS) areas of the core system (Bonda et al., 1996; Castelli et al., 2000; Gobbini et al., 2007; Grossman and Blake, 2002; Schultz et al., 2003).

Examining cortical activation in individuals who have not had visual experience provides insight into the principles that guide cortical organization in terms of both preservation of regional function in the absence of visual input and the nature of plastic changes. Studies of the congenitally blind have demonstrated the striking preservation of organization

* Corresponding author. Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale (DIMES), Medical School, University of Bologna, Bologna, Italy; Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA.

** Corresponding author. Center for Mind/Brain Sciences, University of Trento, Italy.

E-mail addresses: scott.fairhall@unitn.it (S.L. Fairhall), mariaida.gobbini@unibo.it, maria.i.gobbini@dartmouth.edu (M.I. Gobbini).

<http://dx.doi.org/10.1016/j.neuroimage.2017.06.057>

Received 17 June 2016; Received in revised form 25 May 2017; Accepted 22 June 2017

Available online 30 June 2017

1053-8119/© 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

that had been thought to be shaped by visual experience (He et al., 2013; Holig et al., 2014b; Mahon et al., 2009; Peelen et al., 2013; Pietrini et al., 2004; Ricciardi et al., 2014; Wang et al., 2015), as well as the extensive capacity of the visual system to take over new functions (Amedi et al., 2003, 2004; Buchel et al., 1998a; Buchel et al., 1998b; Cohen et al., 1999; Collignon et al., 2007, 2011; for a review, see Noppeney, 2007; Raz et al., 2005; Sadato et al., 1996). Bedny et al. (2011) showed that reorganization of function in the visual cortices of the congenitally blind includes language processing in the occipital lobe extending into the left posterior fusiform gyrus. Holig et al. (2014b) found an area in the anterior ventral temporal cortex that shows a priming effect for voice identity in the congenitally blind, near an area that shows face-selectivity and view-invariant representation of face identity in the sighted (Guntupalli et al., 2017), suggesting a role of this area in the supramodal representation of person identity.

Here, we ask whether the visual cortices in the ‘core system’ retain their preference for information relevant to social interaction in the congenitally blind. In theoretical terms, ascertaining that the organization of the core system for perception of others is preserved in the absence of visual experience with faces would provide evidence for the general hypothesis that the principles underlying the development of the functional organization of this system are better captured by more abstract concepts involving social perception than by visual perception of a particular stimulus class, i.e. faces.

In the present study blind and sighted participants processed the identity and emotional content of voices speaking word phrases or producing non-linguistic sounds. By manipulating the social content of the stimuli (full phrases versus vocalisations, variation in emotional content), rather than the presence or absence of conspecific information (human versus non-human sounds), we were able to present identity and emotional content across all stimuli while simultaneously manipulating the richness of person information contained in the stimuli.

We employed a combination of univariate and multivariate information-based measures to determine where emotion and identity are represented in the brains of the blind as compared to the brains of the sighted.

2. Materials and Methods

2.1. Participants

Seven blind (age: \bar{x} = 35.4, sd = 10.9; five female) and 15 sighted (age: \bar{x} = 32.2, sd = 13.5, ten female) participants partook in this study. Six blind subjects were so from birth, one lost sight at 16 months of age. Three had residual light detection in one or both eyes but none could perceive form or motion. All participants were right handed. Informed consent was given by all participants and all procedures approved by the University of Trento Human Research Ethics Committee.

2.2. Stimuli

Four professional actors (two female) depicted four emotional states (happiness, disgust, fear and neutral). Two different Stimulus-Types were presented. *Verbal* stimuli (word phrases) consisted of four phrases for each of the four emotions uttered by each of the four speakers (e.g. “Che schifo! (That’s disgusting!)”; “Che bello! (That’s wonderful!)”; “(Aah! Aiuto! (Aaah! Help!))”; “Sto leggendo (I’m reading)”). Phrases matched the emotional state both in terms of semantic content and intonation. *Vocalisations* were non-linguistic stimuli where emotion was conveyed solely through intonation. Each speaker portrayed each emotion once. Thus, we used 64 different stimuli with verbal phrases (4 phrases for 4 emotions by 4 speakers) and 16 different stimuli with nonverbal utterances (4 emotions by 4 speakers). Each phrase and vocalisation was recorded on an open-source audio-recording software (Audacity: <http://audacity.sourceforge.net>).

A preliminary study was conducted on 32 university students (who

were unaware of the aim of the study) to select the stimuli with accuracy of recognition higher than 70% for each emotion conveyed by phrases and vocalizations. For the 64 selected verbal phrases, the average accuracies were 82.1% for Happiness; 87.5% for Disgust; 73.2% for Fear; and 80.4% for Neutral. The 16 vocalisations were correctly rated as expressing Happiness 92.8%; Disgust 100%; Fear 71.4% and Neutral 100%.

Rather than artificially modulating the phrases and the vocalisations, auditory stimuli were left in their natural state. The same stimuli were used in both tasks (one-back repetition detection for identity and emotion) and both groups (blind and sighted). Therefore, all analyses emphasized contrasts that are orthogonal to the stimuli, namely the effects of Group and Task (see below). Consequently, stimuli varied in mean length both as a function of actor [1.48, 1.50, 1.05, 1.30 s; $F(3,60) = 5.4$, $p < 0.01$] and portrayed emotion [disgust: 1.36; happiness: 1.60; fear: 1.19 neutral: 1.18 s; $F(3,60) = 4.4$, $p < 0.01$]. Vocalisations (mean = 3.1 s) were longer than verbalisations (mean = 1.3 s) to facilitate recognition of the emotion conveyed by intonation, which was more difficult for vocalizations due to lack of semantic content.

2.3. Task

Immediately prior to the fMRI experiment, the participants were familiarized with the task. This practice session consisted of 16 trials with the same structure used during the fMRI experiment. Responses were recorded using an identical device to that in the scanner.

In separate fMRI runs, participants performed one-back repetition detection tasks based on either emotional content or identity. Because these tasks used the same stimuli, any effect of task is unrelated to stimulus variations (Hoffman and Haxby, 2000). At the beginning of each run the experimenter specified if participants had to recognize whether the stimulus presented conveyed the same emotion as the previous stimulus (independently of speaker identity or stimulus type), or was spoken by the same actor as the previous stimulus (independently of the emotion portrayed or stimulus type). Subjects pressed a button with the index finger for a “yes” response and with the middle finger for a “no” response. Trials consisted of the presentation of the auditory stimulus and a response period until 4.4 s after stimulus onset. Participants were told prior to the experiment that there were 4 different actors and 4 different emotions.

2.4. Procedure

Stimulus conditions followed a $4 \times 4 \times 2$ design (4 emotions, 4 speakers, 2 stimulus-types) and were presented in two different tasks. Four runs were acquired for each subject. Each run had 128 trials, consisting of 64 verbal phrases (all combinations of phrase, emotion, and speaker) and 64 vocalisations (each of 16 vocalisations – all combinations of emotions and speakers – was repeated four times). Each run also contained 16 4.4 s null events and was divided into four equal parts with 32 trials and 4 null events in each part. Stimuli were counterbalanced across runs so that the probability of a stimulus being preceded by each category of emotion or actor was equal. The order of runs with emotion and identity repetition tasks was counterbalanced across subjects. Within each run 10-second null events separated four blocks of stimuli. Each run began and ended with a 20 s baseline period.

2.5. MRI scanning

Stimuli were played through a SereneSound Audio System (Resonance Technology, Inc., California, USA.) at a level fixed across participants so as to be clearly audible above the scanner noise. Data collection was conducted at the Center for Mind/Brain Sciences (CIMeC), University of Trento on a Bruker BioSpin MedSpec 4T using a USA Instruments 8-channel phased-array head coil. One thousand two hundred and sixty volumes of 37 AC–PC aligned slices were acquired over 4 runs using an

echo-planar 2D imaging sequence (image matrix = 70×64 , repetition time = 2200 ms, echo time = 33 ms, flip angle = 76° , slice thickness = 3 mm, gap = 0.45 mm, with 3×3 mm in plane resolution). An additional high-resolution ($1 \times 1 \times 1$ mm³) T1-weighted MPRAGE sequence was acquired (sagittal slice orientation, centric phase encoding, image matrix = 256×224 [Read \times Phase], field of view = 256×224 mm [Read \times Phase], 176 slices with 1-mm thickness, GRAPPA acquisition with acceleration factor = 2, duration = 5.36 min, repetition time = 2700, echo time = 4.18, TI = 1020 ms, 7° flip angle).

2.6. fMRI analysis

2.6.1. Preprocessing

Analysis was performed in SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). After discarding the first 4 vol of each run, all images were corrected for head movement. Slice-acquisition delays were corrected using the middle slice as reference. All images were normalized to the standard SPM8 EPI template (Montreal Neurological Institute [MNI] stereotactic space), resampled to a 3-mm isotropic voxel size, and spatially smoothed using an isotropic Gaussian kernel of 8-mm full-width half-maximum (FWHM). The time series at each voxel for each participant were high pass filtered at 128 s and prewhitened by means of an autoregressive model AR(1).

2.6.2. Univariate analysis

Subject-specific beta weights were derived with a general linear model (GLM). 64 regressors were formed by convolving the onset of each combination of Actor/Emotion/Stimulus-Type separated for the identity-task and emotion-task runs with the SPM8 hemodynamic response function. The 6 head-motion parameters were included as additional regressors of no interest. Subject-specific beta values for each level of task, stimulus-type and group were entered into a group-level random-effects GLM to allow statistical inference at the population level.

2.6.3. Face-selective localiser

In line with our initial hypothesis, to compare the results obtained in this study to classic face-selective regions, we selected ROIs identified in a previous independent study (Fairhall et al., 2014) through a face localiser in an independent group of 18 sighted subjects in the bilateral FFA and the bilateral pSTS. In this experiment, subjects viewed famous and unknown faces or places while performing a categorisation task (for further details see Fairhall et al., 2014). Using the ROIs from an independent group of subjects rather than a localiser derived from our sighted participants allowed us to identify the face-sensitive FFA and pSTS ROIs and to avoid biases associated with using subject-specific data for one group (sighted) and generic data for the other (blind). The data from this independent group did not play any further role in the analyses of the responses to voices in the current experiment.

2.6.4. Multivariate analysis

MVPA (Haxby et al., 2001, 2014; Norman et al., 2006; Tong and Pratte, 2012) was performed on single trials (i.e. the pre-processed volume acquired over the 4.4–6.6 s after stimulus onset). Decoding was implemented separately eight times: at each level of Stimulus-Type and Task decoding was performed twice, once with the goal of decoding emotion and once with the goal of decoding identity.

During classifier training MVP analyses built classifiers using linear discriminant analysis (LDA; Carlson et al., 2003; McLachlan, 2004) to classify the patterns of response to four conditions (either four emotions or four actors). MVP analysis determines whether the patterns of response to each condition (here four emotions or four speaker identities) in a region can be reliably distinguished by building a classifier on training data and testing the classifier's validity on test data (Norman et al., 2006). LDA tests whether the test response vectors are closer, using Euclidean distances, to training data response vectors from the same condition as compared to training data response vectors from any of the

other three conditions. In the current MVP classification analyses, each trial in the testing data was classified as the condition to which it was closest in the training data, as indexed by the means of all training trials for each condition. There were two runs for each task, and each run contained 64 trials with verbal statements and 64 trials with nonverbal vocalisations. Classifiers were trained on the 64 trials in one run with of a given stimulus type, then tested on the 64 trials of the same stimulus type in the other run with the same task. Thus, there were 16 trials for each of four conditions in the training data, and 16 trials per condition in the left-out test data. A searchlight analysis was performed (Kriegeskorte et al., 2006) within a grey matter mask using the searchlight function of the Princeton MVPA Toolbox (<http://www.csmb.princeton.edu/mvpa>).

The searchlight volume was spherical with a three-voxel-radius sphere (~ 123 3 mm voxels). Classification accuracy for each subject was entered into a $2 \times 2 \times 2$ random effects GLM for each searchlight (factors: group, stimulus type, task; chance for one out of four classification of 4 emotions or 4 identities = 25%). Because of the large number of samples in our classification analysis, mean single subject classification accuracies were normally distributed ranging between 0.15 and 0.35, allowing the use of parametric tests. As the main effect of the factor Stimulus-Type and the interactions of this factor with Group and Task were not significant, we present only the main effects and interactions involving the Group and Task factors.

Subsequent to the searchlight analysis, ROI analyses were performed using the MarsBar toolbox for SPM (<http://marsbar.sourceforge.net>). ROIs for the analysis of the interaction between group and task for emotion decoding were defined based on results from the searchlight analysis. Six spherical ROIs (3 voxel radius) were centred on the voxels with peak group differences in decoding accuracy (main effect of group, see Table 4) and contained all significant voxels ($p < 0.001$, for main effect of group).

ROI selection via independent orthogonal contrasts can inflate the chance of type-1 errors if sample sizes are unequal (Kriegeskorte et al., 2009). For instance, if the contrast Task versus Baseline were used to identify ROIs for subsequent analysis of the main effect of group (a 'contrast A + B' situation), then unequal sample sizes would inflate the probability of false positives. In our case, we used orthogonal contrasts of a different nature. We use the contrast GrpA > GrpB to delineate ROIs, then the simple effects of, and interaction with, the factor 'Task' were interrogated. To ensure that this would not be subject to the same statistical bias, we ran a series of simulations on random Gaussian data generated under the null hypothesis. For 10,000 virtual voxels, data was synthesised for 2 groups (N; Grp1 = 7; Grp2 = 15) and two conditions. Mimicking our analysis, 100 of these voxels were selected as a virtual ROI. The proportion of false positives within this ROI was calculated over 160000 random data sets. This approach was very sensitive to bias. Modelling of the documented contrast A + B ROI-selection bias resulted in 100% penetrance of false positives. However, the ROI selection contrasts used in this study had no influence on type-1 errors (5.04% main effect, 4.96% interaction). We are therefore confident that our data and analytic approach do not reflect circularity and are statistically unbiased.

Unless otherwise stated, statistics maps were generated using an initial voxel-wise threshold of $p < 0.001$. Correction for multiple comparisons was subsequently performed at the cluster level using the family wise error (FWE) correction as implemented in SPM8. For MVPA analysis, the critical p was set to $p < 0.0125$ as four separate whole-brain comparisons were investigated in this study.

3. Results

Participants (factor: Group) listened to four actors portraying four emotional states. Emotions were conveyed either through verbalisations or non-linguistic vocalisations (factor: Stimulus-Type). In alternate fMRI blocks, participants performed a 1-back matching task on either the identity of the actor or the conveyed emotion (factor: Task).

3.1. Behavioural performance

Mean reaction time and accuracy were analysed in two separate ANOVAs with the factors, Group, Stimulus-Type and Task (see Table 1). Mean reaction times (and standard error) for the emotion (2032 ± 59 msec) and actor (2006 ± 73 msec) repetition-detection tasks were faster for verbalisations (1843 ± 56 msec) than vocalisations (2195 ± 80 msec; $F_{(1,20)} = 38.5$, $p < 0.001$) but, importantly, there was no main effect of Group or Group by Stimulus-Type interaction (p -values > 0.2). Accuracy was high for both the emotion ($88 \pm 2\%$) and actor ($72 \pm 2\%$) repetition-detection tasks. It should be noted that participants were performing a 4AFC 1-back detection task where failure to classify a stimulus on any given trial affects accuracy on the next trial. Consequently, these accuracy scores may reflect an underestimate of the single trial performance. Performance on the emotion repetition-detection task was more accurate than on the identity repetition-detection task ($F_{(1,20)} = 41.2$, $p < 0.001$) but, critically, did not differ as a function of the factors Group or Stimulus-Type (either as main effects or interactions, all p -values > 0.15).

3.2. fMRI data

In order to identify the main effects of Group, Stimulus Type and Task, we performed a univariate analysis of the whole brain and regions of interest (ROIs). We selected ROIs (see Methods) in the bilateral FFA and the bilateral pSTS to investigate whether these core face system areas play a more general, supramodal role in perception of others.

In order to investigate the information represented in core face system areas during voice perception in the congenitally blind, we performed a searchlight multivariate pattern classification analyses for voice identity and emotion.

3.2.1. Univariate analysis

Voxel-wise analysis revealed a main effect of Group, with greater activation in blind subjects over a broad expanse of the visual cortices in the occipital lobe extending into parietal and posterior temporal regions (Fig. 1A). Additionally, a region of the right auditory cortex was more active for blind than sighted individuals. Analysis of the main effect of Stimulus-Type (verbalisation vs vocalisation) revealed that non-linguistic vocalisations elicited greater activation in the lateral fissure and Heschl's gyrus and linguistic verbalisations produced greater activation along the superior and middle temporal sulcus (Fig. 1B).

To identify brain regions where Stimulus-Types were processed differently in blind and sighted individuals, we examined the interaction between Group and Stimulus-Type (Fig. 1C i). The anatomical distribution is consistent with core regions of the face-processing network. As a direct test of our hypotheses, we conducted an analysis within a face-selective localiser obtained from 18 independent participants (Fig. 1C i; threshold $p < 0.001$, contrast: faces $>$ places; see Methods for further details). Using a small volume correction, we confirmed that the regions identified from the interaction between Group and Stimulus Type in the present experiment lie within the face-selective network. Both bilateral FFA and right pSTS show significant effects within this search volume (left FFA, $p < 0.002$; right FFA $p < 0.012$; right pSTS $p < 0.025$, cluster-level corrected; see Table 2). Speech-selective effects were also evident with a statistical mask of regions showing a greater response for

Table 1
Mean (standard error) accuracy and reaction time for behavioural task.

		Accuracy		Reaction Time (msec)	
		Verbal	Vocal	Verbal	Vocal
Sighted	Emotion	0.89 (0.02)	0.87 (0.02)	1902 (64)	2247 (90)
	Actor	0.71 (0.03)	0.70 (0.03)	1912 (71)	2245 (109)
Blind	Emotion	0.91 (0.03)	0.90 (0.03)	1747 (93)	2138 (131)
	Actor	0.78 (0.04)	0.76 (0.05)	1669 (104)	2036 (160)

verbalisations ($p < 0.001$, Fig. 1B, red), with blind participants showing greater activation than sighted subjects for verbalisations in the left and right lateral fusiform gyri (c.f. Table 2). Signal plots extracted from the three independent face localiser ROIs (IFFA, rFFA, rpSTS; Fig. 1C i) confirm the direction of the selective response verbalisations in the blind. Inspection of single subject data revealed that the right lateral fusiform effect (at $p < 0.005$) was present in six of the seven blind subjects and was not present in any of the 15 sighted subjects even at a more liberal threshold of $p < 0.05$ (see Fig. 1C). This suggests that cortex in the brains of congenitally blind subjects that is the homologue of the FFA in sighted subjects may play a role in extracting information from voices.

No main effect of Task or interactions with Task survived correction for multiple comparisons.

3.2.2. Multivariate pattern analyses

We performed separate multivariate pattern classification analyses of emotion and identity (one out of four classification for each analysis) at each level of Stimulus Type and Task in each subject using a searchlight analysis and linear discriminant analysis.

The resulting brain maps were entered into a $2 \times 2 \times 2$ random effects GLM (factors: group, stimulus type, task).

3.2.2.1. Overall decoding accuracy. To investigate overall decoding accuracy, within our second level GLM, we contrasted decoding accuracies across all level of Group, Stimulus Type and Task to chance.

3.2.2.2. Emotion. Results revealed that large portions of the cortex contain information about the emotion conveyed by stimuli (Fig. 2A). These included primary and secondary auditory cortex, bilateral frontal operculum, a bilateral region of the face area of the sensory motor cortex and a superior section of the medial prefrontal cortex as well as a small section of the right SMA (see Table 3).

3.2.2.3. Identity. Four-way classification of identity was associated with a less extensive network of brain regions (Fig. 2B). It was possible to decode identity in primary and secondary auditory cortices and also in the right hippocampus (see Table 3). Emotion decoding in this right hippocampal region did not differ from chance ($t_{(21)} = 0.57$, $p = 0.58$), in contrast to auditory regions where emotion decoding accuracies were greater than identity decoding accuracies.

The above patterns of Emotion and Identity decoding were present in both Blind and Sighted groups when considered in isolation. Formal and quantitative analysis of group differences are considered in the next section.

3.3. Main effect of group and interaction with task

3.3.1. Emotion

Decoding accuracies for emotional expression were higher in blind than in sighted participants in bilateral pSTS and the lateral fusiform gyrus as well as in the cuneus and a small section of right auditory cortex (Fig. 3, Table 4). No regions showed greater decoding accuracy for emotion in sighted as compared to blind participants.

Within the fusiform gyrus, [defined by the AAL template (Tzourio-Mazoyer et al., 2002), dilated by 3 mm and a MNI y coordinate > -75], the region that showed a group effect for emotion decoding abutted the left FFA and overlapped the right FFA (as identified in the independent face localiser (Fairhall et al., 2014)). Of the 27 right fusiform gyrus voxels that showed a group effect of decoding, 7 (26%) fell within the right FFA (149 voxels). For comparison, Fig. 4 shows the superimposed locations of the group effect of emotion, the verbal/vocal by group interaction and the independent face localiser ($p < 0.001$, extent > 50), along with published coordinates for the visual word from area (VWFA).

To further qualify the representation of emotion in these brain regions, the main effect of Task and the interaction between Group and

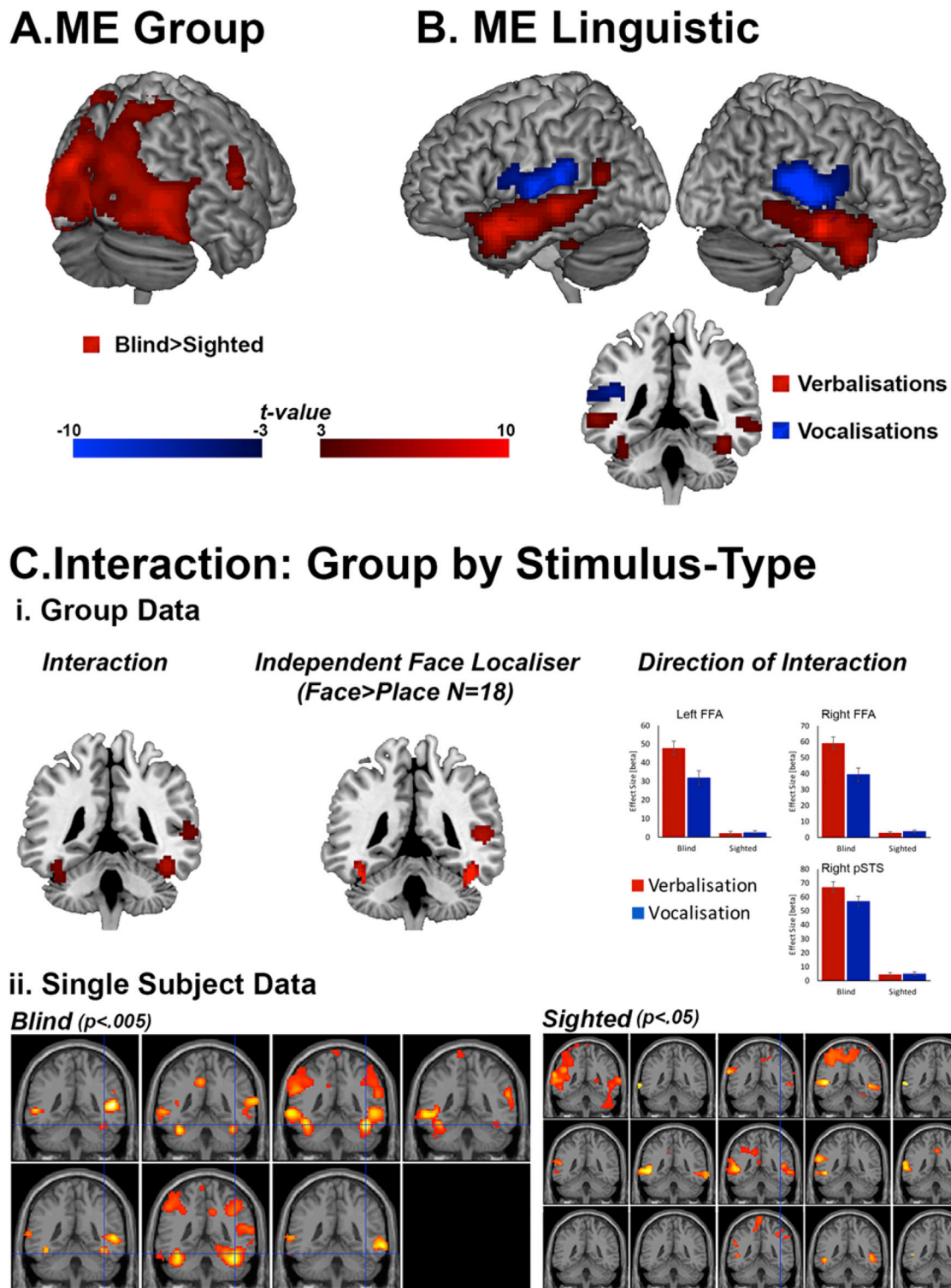


Fig. 1. Univariate Analysis. A) Main effect of group ($p < 0.001$). Listening to voices produced greater activation in blind participants over an expanse of occipital temporal cortex. B) Main effect of Stimulus-Type ($p < 0.001$). Vocalisation produced stronger activation in the Sylvian fissure while verbalisations produced stronger effects along the superior temporal and fusiform gyri. C) Interaction of Group and Stimulus-Type at the group level (i - left) note the overlap with classical face-selective regions characterized from an independent localiser (i - middle) (for both figures $y = -43$, $p < 0.001$). Signal plots of Group and Stimulus-Type, effects extracted from ROIs derived from the independent face localiser, show the direction of the interaction (l - right; error bars depict between subject SEM). ii) Interaction for single subjects- 7 blind and 15 sighted (blind: $p < 0.005$; sighted $p < 0.05$). See Table 2 for cluster extents and peak locations.

Task were examined in ROIs centred at the six peak locations for the main effect of group in the left and right pSTS, left and right fusiform gyri, the cuneus, and right STG (Table 4 and Fig. 3). By using this orthogonal contrast, the main effect of Group permitted us to identify ROIs for subsequent analysis of Task and Group x Task effects in a statistically

unbiased manner despite the unequal group size (see Methods). Because the same stimuli were used for both tasks and for both groups, the main effect of Task and the interaction of Task by Group are orthogonal to possible confounding effects of variations in the low-level properties of the auditory stimuli (length and actor. In the left pSTS, the Group x Task

Table 2

Significance, extent and location of univariate analysis of the main effects and interaction of Group and Stimulus-Type. All clusters larger than 50 voxels are shown for main effects. No significant effects were seen outside statistical masks used in the interactions.

	Region	Cluster $p_{(\text{corrected})}$	extent (vox.)	Peak t-Value	x	y	z	
Main effect of Group	parietal/visual cortex	<0.001	6257	8.8	-15	-91	22	
	rSTG	0.022	124	5.6	66	-19	16	
Main effect of Stimulus type Verbal > Vocal	lMTG	<0.001	792	10.3	57	-10	-20	
	rMTG	<0.001	866	9.0	-57	2	-20	
	lpSTS	0.106	75	5.7	-60	-58	13	
	rFFA	0.020	127	5.7	39	-46	-20	
	lFFA	0.144	66	4.2	-39	-43	-23	
	Vocal > Verbal	l Heschl's	<0.001	1208	16.7	-39	-28	10
		r Heschl's	<0.001	1335	15.0	54	-22	7
	Interaction Group x Stimulus Type (masked by Verbal > Vocal)	lFFA	0.004	88	6.4	45	-40	-20
		rFFA	0.011	60	4.2	-42	-40	-20
		rpSTS	0.175	4	3.6	57	-46	-2
(masked by face-localiser)		lFFA	0.002	81	6.4	45	-40	-20
		rFFA	0.012	33	4.2	-42	-40	-20
		rpSTS	0.025	19	4.1	60	-43	7

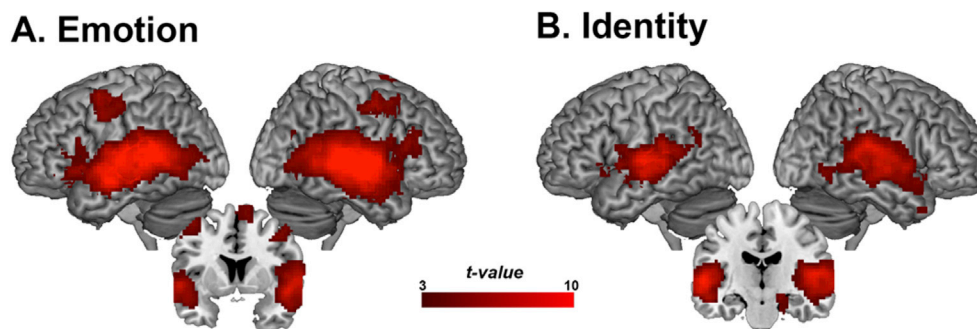


Fig. 2. Searchlight Analysis of overall decoding-accuracy of emotions (A) and identities (B). Shown is the main effect of decoding-accuracy across all levels of the Group by Task by Stimulus-Type GLM versus the chance accuracy baseline. Information about conveyed emotion is present across the extensive section of primary and secondary auditory cortices as well as face regions of the sensory motor cortex and medial prefrontal structures. It was possible to decode actor-identity in a less extensive subsection of auditory cortices as well as the right hippocampus. See Table 3 for cluster extents and peak locations.

Table 3

Significance, extent and location of cortical areas in which overall decoding of Emotion and Identity were significant (collapsed across factors Group, Task and Stimulus-Type).

	Region ^a	Cluster $p_{(\text{corrected})}$	extent (vox.)	Peak t-value	x	y	z
Emotion	IPT	<0.001	4320	13.5	60	-19	1
	rPT	<0.001	3940	13.2	-48	-13	-5
	rSMC	<0.001	289	5.9	48	-4	49
	lSMC	<0.001	355	5.6	-45	2	58
	medial PFC	<0.002	135	4.4	6	29	52
	SMA	<0.010	93	4.2	12	5	70
Identity	IPT	<0.001	1742	9.3	-54	-19	-2
	rPT	<0.001	2324	8.5	51	-16	10
	rHipp	<0.004	105	4.5	12	-7	-23

^a Abbreviations: 'l', left; 'r', right PT: planum temporale; SMC: sensorimotor cortex; SMA: supplementary motor area; PFC: prefrontal cortex; Hipp: hippocampus.

interaction was significant ($p = 0.029$): the emotion task (compared to the identity task) yielded higher decoding accuracies of emotional expression in blind ($p = 0.002$) but not sighted participants ($p = 0.37$). The representation of emotion in the right pSTS was also strengthened by Task (main effect $p < 0.001$) but, unlike the left pSTS, this did not differ between groups (Group \times Task interaction: $p = 0.40$) with both blind ($p = 0.023$) and sighted ($p < 0.001$) showing a modulation by Task. Task had a moderate effect on decoding accuracy in the left ($p = 0.048$) but not right fusiform gyrus ($p = 0.28$).

Additionally, the cuneus showed strong modulation by Task in blind ($p < 0.001$) but not sighted participants ($p = 0.64$; Group \times Task interaction: $p = 0.001$). Activation in the right auditory cortex was not

modulated by Task for either Group (p -values > 0.45).

3.3.2. Identity

The contrast blind > sighted revealed two brain regions that have higher accuracies for decoding identity in the blind, namely the left superior temporal gyrus (STG) and right Heschl's gyrus, regions that are associated with the auditory processing of language (Table 5). Neither of these regions was modulated by Task ($p < 0.5$).

No regions showed significantly greater decoding accuracy for identity in sighted as compared to blind participants.

4. Discussion

We investigated the reorganization of the visual face-processing system in the absence of visual experience with faces in the congenitally blind.

We found plastic changes related to perception of information in voices that is relevant for understanding others in regions that correspond to the core face-processing system in the sighted. Blind subjects exhibit enhanced selectivity for verbal stimuli compared to non-linguistic vocalizations in regions corresponding to the bilateral FFA and right pSTS. Multivariate decoding analyses showed accuracies for decoding emotion were higher for the blind in bilateral pSTS and bilateral areas of the lateral fusiform gyrus near the FFA. Accuracies for decoding emotion were also higher in the blind in right auditory cortex, and accuracies for decoding identity were higher in right and left auditory cortex, suggesting intramodal plasticity consistent with the well-documented within sensory alterations in the blind (Noppeney, 2007).

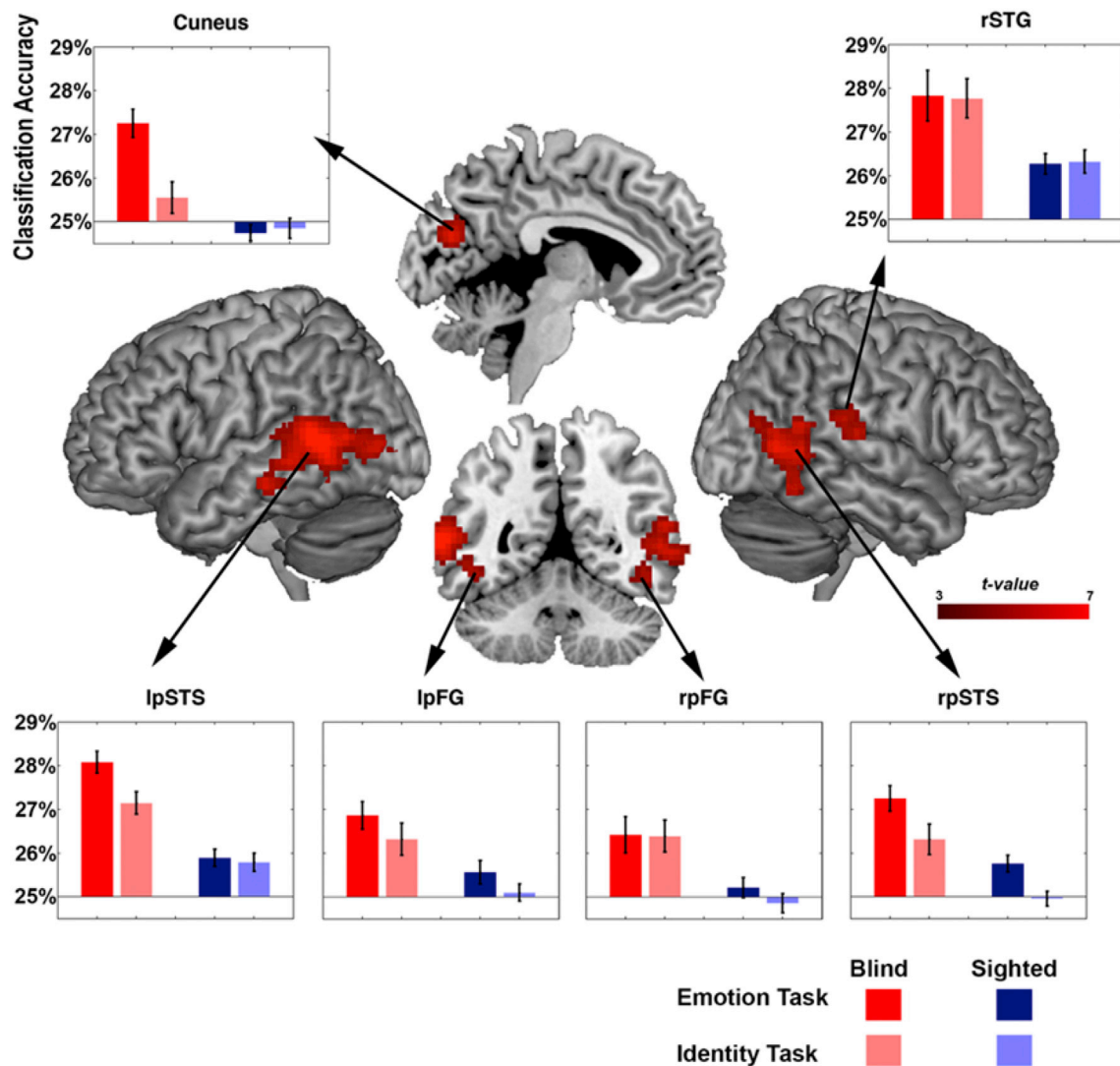


Fig. 3. Brain maps: Searchlight results for the contrast Blind > Sighted for emotion decoding accuracy. Plots: Decoding accuracy (above chance) for the ROI analysis of Task and Task by Group interaction (error bars: SEM). Note: Direct comparisons between the two groups are not valid in this situation due to the ROI selection procedure. Only the effects of Task and the Task by Group interaction are statistically unbiased. See Table 4 for cluster extents and peak locations.

Table 4
Significance, extent and location of the main effect of Group for emotion decoding (see Fig. 3).

Cluster	Region ^a	Cluster $p_{(\text{corrected})}$	extent (vox.)	Peak t-value	MNI		
					x	y	z
Cluster 1	lpSTS	<0.001	503	7.74	-66	-52	7
	lpFG			4.15	-45	-52	-11
Cluster 2	rpSTS	<0.001	287	5.58	54	-64	4
	rpFG			4.62	45	-55	-14
Cluster 3	Cuneus	0.002	116	5.49	-3	-76	22
Cluster 4	rSTG	0.002	118	4.48	66	-25	13

^a Abbreviations: 'l', left; 'r', right; 'a', anterior; 'p', posterior. STS: superior temporal gyrus; FG: fusiform gyrus; STG: superior temporal gyrus.

Overall stronger responses to auditory stimuli with higher information content in the blind as compared to sighted subjects are not surprising since the blind rely mainly on sounds to obtain information that the sighted can obtain through vision. However, the enhancement of person-related information content in parts of the neural system for representation of others reveals a reorganization of components that are dedicated to processing visual stimuli in sighted subjects.

4.1. Plastic reorganization within the face-processing system

In our study, of the extensive occipito-parietal regions showing non-specific plasticity in the blind brain (Fig. 1A), a small section responded selectively to verbal information (Fig. 1C), and overlapped with the core-face network. The enhanced univariate response to verbal stimuli in the FFA was highly reliable in the blind, being present in 6 of the 7 individuals studied.

Unlike other regions of the ventral stream, the FFA only exhibits a category selective response during the presentation of images, unless explicit visual imagery tasks are performed (Epstein et al., 2007; Fairhall and Caramazza, 2013; Fairhall et al., 2014; Noppeney et al., 2006). Moreover, previous studies demonstrating the preservation of categorical organization in object-sensitive visual cortices of the congenitally blind have not found selective activation of this region (He et al., 2013; e.g. Mahon et al., 2009; Peelen et al., 2013; Bi et al., 2016). In individuals who were congenitally blind due to dense cataracts but had their site restored, the face-selectivity of responses in the FFA and the face-specificity of the N170 are reduced (Grady et al., 2014; Röder et al., 2013), suggesting an early critical period during which visual experience with faces is necessary to develop face-selectivity in the core system. Here we observe that spoken sentences stimulate this brain region in the

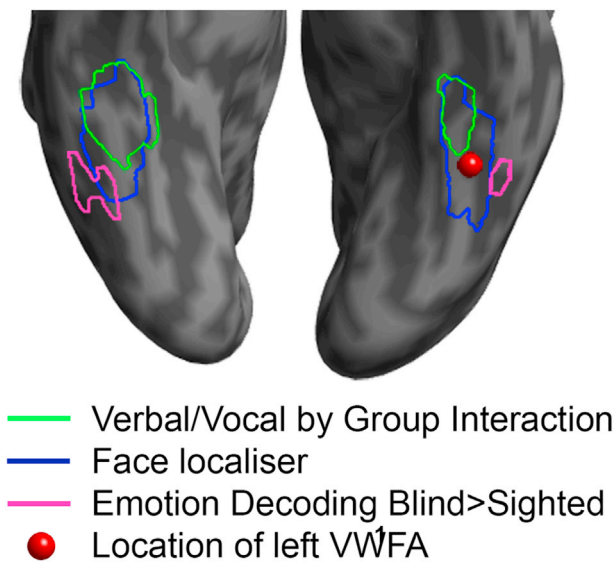


Fig. 4. Superimposed locations of ventral stream effects reported in this study. Outlines depict cluster boundaries ($p < 0.001$) for the verbal/vocal by group interaction (green), independent face localiser (blue), group effect of emotion (blind > sighted, pink). Additionally, a location for the VWFA is indicated from reference (taken from Rauschecker et al., 2012).

Table 5
Significance, extent and location of the main effect of Group for identity decoding accuracy.

	Cluster	extent (vox.)	Peak max T value	MNI			
				x	y	z	
Identity	lSTG	<0.001	404	5.52	-63	-13	-11
	rHeschl	0.006	91	4.99	51	-13	10

blind. Natural spoken language is rich in information about others. To a greater extent than vocalisations, natural spoken language conveys cues highly relevant for social interaction, in addition to semantic and linguistic information.

This fact may be interpreted in ways that are not mutually exclusive. The enhanced response in the blind may reflect the plastic reorganization of the system for perception of others or be more restricted to reorganization of these areas for representation of language. The lateral fusiform cortex in the blind may respond more specifically to linguistic (or semantic) features. Bedny et al. (2011) found that a left fusiform area responded selectively to coherent sentences in a working memory task, but this area was 3 cm more posterior than the area that responded selectively to emotional content in voices in the current study. In the sighted brain there is a dynamic relationship between the letter-string selective left visual word form area (VWFA; McCandliss et al., 2003) and the FFA. In literate, but not illiterate populations, reading competes with face recognition during development, shifting the locus of the left FFA and enhancing the laterality of the FFA to the right (Dehaene et al., 2010; see Dehaene et al., 2015 for a review; Dundas et al., 2013; Ossowski and Behrmann, 2015; Behrmann and Plaut, 2015). Here we provide the counterpoint to the development of face processing in the absence of literacy, showing that in the absence of experience with faces, spoken language functions may encroach upon face-selective regions. The ventral temporal areas that show an effect of verbal versus vocal content and greater decoding of emotion in the congenitally blind are close to the location of the VWFA. This finding is particularly striking as, while Braille and auditory sensory-substitution device reading (Reich et al., 2011; Striem-Amit et al., 2012) share some of the demands of visual reading, verbal language has none of these properties. The reorganization of the FFA for spoken language processing suggests that the lateral fusiform gyrus possesses properties that predispose it for language

related cognition, independently of the orthographic demands of reading. It is notable that selective activation of the FFA is not observed in the blind when syntactically rich sentences are compared to less linguistically rich control stimuli (Bedny et al., 2011) or when theory of mind stories are compared to non-social controls (Bedny et al., 2009). By contrast, we find activation in the FFA and pSTS to spoken sentences that convey emotional content. Our findings suggest that in the absence of visual experience core face regions may assume a role in aspects of voice perception that are relevant to social cognition and perception of others' emotions.

Verbalisations also produced a greater activation than vocalisations in a second region of the core face-processing network, the pSTS. This region is activated both by speech and faces in the sighted, and has been shown to respond to a wide variety of person-related cognition independent of input modality (Fairhall and Caramazza, 2013; Peelen et al., 2010; Saxe and Kanwisher, 2003). The activation of the pSTS, along with the FFA, supports our hypothesis that these regions have a more general function in the perception of socially relevant stimuli, which is not bound to visual experience.

The emotion conveyed by voices could be decoded from activity in the bilateral FFA and pSTS with greater accuracy in the blind than in the sighted. Different emotions conveyed by facial expressions can be decoded from patterns of activity in the pSTS (Said et al., 2010) and this extends to supramodal representations of conveyed emotion that converge across facial expression, body movements, and vocal intonation (Peelen et al., 2010). In the blind, the absence of visual input and visual experience with faces appears to enhance the role of the pSTS in representing vocally-conveyed social information. In the sighted, decoding was also above chance in left pSTS, consistent with Peelen et al. (2010), but insensitive to attention to emotion. In the right pSTS, conveyed emotion was only represented in the sighted if it was task-relevant (see Fig. 3). In the blind, the overall strength of the representation emotion was higher, and this was modulated by task relevance in both hemispheres. This underscores the importance of this region in the supramodal representation of emotion and indicates that the development of this specialisation is not contingent on experience with faces.

4.2. Preservation of function outside the face processing system

We also observed representations of distinctions among emotions and identities common to both blind and sighted participants. Because we used naturalistic stimuli, decoding of identity and emotion in primary auditory cortex as well as voice-sensitive auditory cortices, such as the planum temporale and STG, may be due to differences in the low-level acoustic features (see Belin et al., 2004 for a review; see also Formisano et al., 2008). Our results are consistent with an ROI MVPA study in sighted individuals, which found the most informative voxels for emotion expression span both primary auditory and voice-sensitive cortices (Ethofer et al., 2009).

In addition, we found representations of distinctions among emotions and identities in non-auditory cortices. While the influence of low-level acoustic properties on these regions cannot be fully discounted, their anatomical location makes this less likely. The presence of these effects in the blind is informative as it demonstrates non-dependence on visual experience. Information about emotional expression was present for both groups in bilateral patches of the precentral gyrus in the face area of the motor strip and in the bilateral frontal operculum. Sensory-motor face representations are active both during the production and perception of facial expressions (Adolphs, 2010; Hennenlotter et al., 2005) and disruption of the face region of the somatosensory strip with transcranial magnetic stimulation impairs the recognition of facial expression in sighted individuals (Pitcher et al., 2008). The frontal operculum is activated by both the perception and execution of facial expressions (Montgomery and Haxby, 2008; Montgomery et al., 2009). The present results show that the representation of information about emotional states in these regions is supramodal encompassing information

conveyed by voices as well as by facial expression. Moreover, these results suggest that the role of facial sensorimotor and frontal opercular cortices in representing the emotion of others develops even in those who have never seen a facial expression and cannot be explained by visual mimicry.

We also observed accurate decoding of actor identity in the right hippocampus. This was the only region to show significant decoding accuracy for identity but not for emotion. Two recent studies investigated voice identity priming in congenitally (Holig et al., 2014b) and late (Holig et al., 2014a) blind subjects. Stimuli were disyllabic pseudowords. At uncorrected statistical levels, identity-related priming was seen in a number of regions, including a small volume in the fusiform gyrus, anterior to the FFA, but no effect in the pSTS. We do not observe an effect for identity decoding in this area ($p < 0.05$, uncorrected), rather our results emphasise the role of the hippocampus. The hippocampus contains neurons that respond to specific people over a wide range of views and settings (Quiroga et al., 2005) and multivariate analyses has revealed representations of face identity that are tolerant to varying view points (Anzellotti et al., 2014). This result indicates that the representation of identity in this region may be multimodal – representing identity derived from both faces and voices – and is not dependent on vision or face perception.

No effects of identity were seen within the core-system for face perception in blind participants. Kilgour and colleagues (2005) reported activation of the fusiform gyrus in sighted subjects haptic face perception, but others have not (Pietrini et al., 2004), nor has this effect been reported in the blind (Kitada et al., 2013). As the identities we used were not familiar to the subject, the possibility that this effect depends on the familiarity cannot be ruled out. Von Kriegstein and colleagues (2005) have reported that the fusiform gyrus, amongst other areas, shows an increased response in the sighted when listening to spoken sentences produced by personally familiar people. In both studies (Kilgour et al., 2005; von Kriegstein et al., 2005) the activation of the fusiform cortex could reflect face imagery in sighted subjects.

In sum, our results support the general hypothesis that the principles that guide the development of the functional organization of the core face-processing network may not be restricted to visual perception of a particular class of stimuli but, rather, is based on a broader domain for the social perception of others and can be redirected to voice perception in the absence of visual experience.

Acknowledgement

Above all, we would thank the blind participant who travelled from across Italy to take part in this study. We also thank Jim Haxby for helpful comments and discussion, Alfonso Caramazza for his support, Nick Oosterhoff for his advice on MVPA as well as the UIC of Bologna and the actors and actresses of the Ten Teatro (<http://www.tenteatro.it/it/tenteatro-1>) for their help with creation of the stimuli. This research was supported by the Provincia Autonoma di Trento and the Fondazione Cassa di Risparmio di Trento e Rovereto. This work was supported by a European Research Council Starting Grant awarded to Scott Fairhall (StG2014-640594).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.neuroimage.2017.06.057>.

References

Adolphs, R., 2010. Social cognition: feeling voices to recognize emotions. *Curr. Biol.* 20, R1071–R1072.
 Amedi, A., et al., 2003. Early 'visual' cortex activation correlates with superior verbal memory performance in the blind. *Nat. Neurosci.* 6, 758–766.
 Amedi, A., et al., 2004. Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nat. Neurosci.* 7, 1266–1270.

Anzellotti, S., Fairhall, S.L., Caramazza, A., 2014. Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex* 24, 1988–1995.
 Bedny, M., Pascual-Leone, A., Saxe, R.R., 2009. Growing up blind does not change the neural bases of Theory of Mind. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11312–11317.
 Bedny, M., et al., 2011. Language processing in the occipital cortex of congenitally blind adults. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4429–4434.
 Behrmann, M., Plaut, D.C., 2015. A vision of graded hemispheric specialization. *Ann. N. Y. Acad. Sci.* 1359, 30–46.
 Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. *Trends Cogn. Sci.* 8, 129–135.
 Bi, Y., Wang, X., Caramazza, A., 2016. Object domain and modality in the ventral visual pathway. *Trends Cogn. Sci.* 20, 282–290.
 Bonda, E., et al., 1996. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J. Neurosci.* 16, 3737–3744.
 Buchel, C., et al., 1998a. Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* 121 (Pt 3), 409–419.
 Buchel, C., Price, C., Friston, K., 1998b. A multimodal language region in the ventral visual pathway. *Nature* 394, 274–277.
 Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15, 704–717.
 Castelli, F., et al., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325.
 Cohen, L.G., et al., 1999. Period of susceptibility for cross-modal plasticity in the blind. *Ann. Neurol.* 45, 451–460.
 Collignon, O., et al., 2007. Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cereb. Cortex* 17, 457–465.
 Collignon, O., et al., 2011. Functional specialization for auditory-spatial processing in the occipital cortex of congenitally blind humans. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4435–4440.
 Dehaene, S., et al., 2010. How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364.
 Dehaene, S., et al., 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* 16, 234–244.
 Dundas, E.M., Plaut, D.C., Behrmann, M., 2013. The joint development of hemispheric lateralization for words and faces. *J. Exp. Psychol. Gen.* 142, 348–358.
 Epstein, R.A., Parker, W.E., Feiler, A.M., 2007. Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J. Neurosci.* 27, 6141–6149.
 Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive cortices. *Curr. Biol.* 19, 1028–1033.
 Ethofer, T., et al., 2012. Emotional voice areas: anatomical location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cereb. Cortex* 22, 191–200.
 Fairhall, S.L., Ishai, A., 2007. Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex* 17, 2400–2406.
 Fairhall, S.L., Caramazza, A., 2013. Category-selective neural substrates for person- and place-related concepts. *Cortex* 49, 2748–2757.
 Fairhall, S.L., et al., 2014. Person- and place-selective neural substrates for entity-specific semantic access. *Cereb. Cortex* 24, 1687–1696.
 Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322, 970–973.
 Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41.
 Gobbini, M.I., et al., 2007. Two takes on the social brain: a comparison of theory of mind tasks. *J. Cognit. Neurosci.* 19, 1803–1814.
 Gobbini, M.I., et al., 2011. Distinct neural systems involved in agency and animacy detection. *J. Cognit. Neurosci.* 23, 1911–1920.
 Grady, C.L., Mondloch, C.J., Lewis, T.L., Maurer, D., 2014. Early visual deprivation from congenital cataracts disrupts activity and functional connectivity in the face network. *Neuropsychologia* 57, 122–139.
 Grossman, E.D., Blake, R., 2002. Brain areas active during visual perception of biological motion. *Neuron* 35, 1167–1175.
 Guntupalli, J.S., Wheeler, K.G., Gobbini, M.I., 2017. Disentangling the representation of identity from head view along the human face processing pathway. *Cereb. Cortex* 27, 46–53.
 Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233.
 Haxby, J.V., et al., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.
 Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2002. Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67.
 Haxby, J.V., Gobbini, M.I., 2011. Distributed neural systems for face perception. In: *The Oxford Handbook of Face Perception*, Vol. 1.
 Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* 37 (37), 435–456.
 He, C., et al., 2013. Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage* 79, 1–9.
 Hennenlotter, A., et al., 2005. A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage* 26, 581–591.
 Hoffman, E.A., Haxby, J.V., 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat. Neurosci.* 3, 80–84.
 Holig, C., et al., 2014a. Brain systems mediating voice identity processing in blind humans. *Hum. Brain Mapp.* 35, 4607–4619.

- Holig, C., et al., 2014b. Crossmodal plasticity in the fusiform gyrus of late blind individuals during voice recognition. *Neuroimage* 103, 374–382.
- Ishai, A., Schmidt, C.F., Boesiger, P., 2005. Face perception is mediated by a distributed cortical network. *Brain Res. Bull.* 67, 87–93.
- Kitada, R., Okamoto, Y., Sasaki, A.T., Kochiyama, T., Miyahara, M., Lederman, S.J., 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.
- Kilgour, A.R., Kitada, R., Servos, P., James, T.W., Lederman, S.J., 2005. Haptic face identification activates ventral occipital and temporal areas: an fMRI study. *Brain Cogn.* 59, 246–257.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3863–3868.
- Kriegeskorte, N., et al., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540.
- Leaver, A.M., Rauschecker, J.P., 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
- Mahon, B.Z., et al., 2009. Category-specific organization in the human brain does not require visual experience. *Neuron* 63, 397–405.
- McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299.
- McLachlan, G., 2004. *Discriminant Analysis and Statistical Pattern Recognition*, vol. 544. John Wiley & Sons.
- Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J. Neurosci.* 32, 14205–14216.
- Montgomery, K.J., Haxby, J.V., 2008. Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 20, 1866–1877.
- Montgomery, K.J., Seeherman, K.R., Haxby, J.V., 2009. The well-tempered social brain. *Psychol. Sci.* 20, 1211–1213.
- Noppeney, U., et al., 2006. Two distinct neural mechanisms for category-selective responses. *Cereb. Cortex* 16, 437–445.
- Noppeney, U., 2007. The effects of visual deprivation on functional and structural organization of the human brain. *Neurosci. Biobehav. Rev.* 31, 1169–1180.
- Norman, K.A., et al., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cognit. Sci.* 10, 424–430.
- Ossowski, A., Behrmann, M., 2015. Left hemisphere specialization for word reading potentially causes, rather than results from, a left lateralized bias for high spatial frequency visual information. *Cortex* 72, 27–39.
- Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. *J. Neurosci.* 30, 10127–10134.
- Peelen, M.V., et al., 2013. Tool selectivity in left occipitotemporal cortex develops without vision. *J. Cogn. Neurosci.* 25, 1225–1234.
- Pernet, C.R., et al., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119, 164–174.
- Pietrini, P., et al., 2004. Beyond sensory images: object-based representation in the human ventral pathway. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5658–5663.
- Pitcher, D., et al., 2008. Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J. Neurosci.* 28, 8929–8933.
- Quiroga, R.Q., et al., 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107.
- Rauschecker, A.M., Bowen, R.F., Parvizi, J., Wandell, B.A., 2012. Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1568–E1577.
- Raz, N., Amedi, A., Zohary, E., 2005. V1 activation in congenitally blind humans is associated with episodic retrieval. *Cereb. Cortex* 15, 1459–1468.
- Reich, L., et al., 2011. A ventral visual stream reading center independent of visual experience. *Curr. Biol.* 21, 363–368.
- Remedios, R., Logothetis, N.K., Kayser, C., 2009. Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18010–18015.
- Ricciardi, E., Handjaras, G., Pietrini, P., 2014. The blind brain: how (lack of) vision shapes the morphological and functional architecture of the human brain. *Exp. Biol. Med.* 239, 1414–1420.
- Röder, B., Ley, P., Shenoy, B.H., 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.
- Romanski, L.M., Averbeck, B.B., Diltz, M., 2005. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* 93, 734–747.
- Sadato, N., et al., 1996. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* 380, 526–528.
- Said, C.P., et al., 2010. Graded representations of emotional expressions in the left superior temporal sulcus. *Front. Syst. Neurosci.* 4, 6.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* 19, 1835–1842.
- Schultz, R.T., et al., 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 358, 415–427.
- Striem-Amit, E., et al., 2012. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–652.
- Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63 (63), 483–509.
- Tsao, D.Y., Moeller, S., Freiwald, W.A., 2008. Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19514–19519.
- Tzourio-Mazoyer, N., et al., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., Giraud, A.L., 2005. Interaction of face and voice areas during speaker recognition. *J. Cogn. Neurosci.* 17, 367–376.
- Wang, X., et al., 2015. How visual is the visual Cortex? Comparing connective and functional fingerprints between congenitally blind and sighted individuals. *J. Neurosci.* 35, 12545–12559.
- Watson, R., et al., 2014a. People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex* 50, 125–136.
- Watson, R., et al., 2014b. Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. *J. Neurosci.* 34, 6813–6821.
- Yovel, G., Belin, P., 2013. A unified coding strategy for processing faces and voices. *Trends Cogn. Sci.* 17, 263–271.