

AUDITORY PERCEPTION OF NATURAL SOUND CATEGORIES – AN FMRI STUDY

M. SHARDA AND N. C. SINGH*

National Brain Research Centre, NH-8, Nainwal Mode, Manesar, Gurgaon 122050, India

Abstract—Despite an extremely rich and complex auditory environment, human beings categorize sounds effortlessly. While it is now well-known that this ability is a result of complex interaction of bottom-up processing of low-level acoustic features and top-down influences like evolutionary relevance, it is yet unclear how these processes drive categorization. The objective of the current study was to use functional neuroimaging to investigate the contribution of these two processes for category selectivity in the cortex. We used a set of ecologically valid sounds that belonged to three different categories: animal vocalizations, environmental sounds and human non-speech sounds, all matched on acoustic structure attributes like harmonic-to-noise ratio to minimize differences in bottom-up processing as well as matched for familiarity to rule out other top-down influences. Participants performed a loudness judgment task in the scanner and data were acquired using a sparse-temporal sampling paradigm. Our functional imaging results show that there is category selectivity in the cortex only for species-specific vocalizations and this is revealed in six clusters in the right and left STG/STS. Category selectivity was not observed for any other category of sounds. Our findings suggest a potential role of evolutionary relevance for cortical processing of sounds. While this seems to be an appealing proposition, further studies are required to explore the role of top-down mechanisms arising from such features to drive category selectivity in the brain. © 2012 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: auditory, categorization, acoustic, fMRI, non-speech vocalizations, STG.

INTRODUCTION

Human beings have extremely rich and complex auditory environments composed of stimuli which vary in source, location, intensity and behavioural relevance. Despite this complexity, humans are skilled at auditory categorization and discrimination. The mechanisms underlying this ability have been a subject of active investigation in the last decade and have revealed much about the neural

organization of sounds in the regions of primary (PAC) and secondary auditory cortex (SAC).

Past research on primates as well as humans has demonstrated a hierarchical organization in the auditory cortex, with the core area in the PAC responding preferentially to simple sounds like pure tones and musical notes and the surrounding belt and parabelt regions of the SAC, responding to more complex sounds like ripple noise and speech (Belin et al., 2000; Wessinger et al., 2001).

While there is now general consensus that the basis for this categorization is a complex combination of bottom-up and top-down processes (Lewis et al., 2005, 2009), the detailed mechanisms are still not well understood. For instance, little is known about how complex properties of source identification for communication emerge in the auditory cortex. This is perhaps, because, fundamental properties like frequency encoding and spectro-temporal computations are already present much lower down in the auditory pathways, making the PAC and SAC much 'higher' areas in the hierarchy of information processing (King and Nelken, 2009). In addition, the presence of multiple relay centres in the subcortical structures as well as descending corticofugal tracts, lead to several interactions of top-down control and bottom-up processes making the representation of sound categories a complex process to investigate.

The idea of a category-selective module in the auditory brain has been supported by a number of studies. Lesion studies have shown selective deficits for certain categories including environmental sounds and animal vocalizations (Saygin et al., 2010), whereas imaging studies have shown cortical selectivity for conspecific vocalizations, environmental sounds (Belin et al., 2000; Doehrmann et al., 2008) and musical sounds. However, the neurophysiological basis of this segmentation is still not very well understood.

There is much speculation about whether category selectivity in the auditory cortex is driven by a lower level acoustic feature (Lewis et al., 2009) or an inherent behavioural characteristic lending these categories perceptual salience. Recently, the proposition of the harmonic-to-noise ratio (HNR) as a probable signal attribute was shown by Lewis et al. (2009). Using animal sounds and artificial ripple noise, they showed a parametric activation of the auditory cortex with increasing HNR. Subsequently, a study by Leaver and Rauschecker (2010) demonstrated a dissociation between stimulus features and stimulus categories for certain sounds. They reported an anteroventral organization in the superior temporal lobe, with the posterior regions responding more to stimulus

*Corresponding author. Tel: +91-124-233-8920/28x333; fax: +91-124-233 8927/28.

E-mail address: nandini@nbrc.ac.in (N. C. Singh).

Abbreviations: A, animal sounds; HNR, harmonics-to-noise ratio; NV, environmental sounds; PAC, primary auditory cortex; ROI, region of interest; SAC, secondary auditory cortex; V, human vocalizations.

features like spectral structure and temporal variability and more anterior regions to stimulus categories like human speech and musical instrument sounds. However, the extent to which lower level features like the spectral content as measured by the HNR as well as higher level attributes like familiarity contribute to categorization in the cortex, still remains to be explored.

In the present study we used a series of ecologically valid natural sounds of our environment, which were matched on features of familiarity and were all evolutionarily relevant at some level. In order to minimize the role of acoustic heterogeneity of the stimuli during perception of the sounds, we matched them on standard measures like amplitude and duration and additionally on HNR. Thus, our objective was to identify the mechanism of cortical selectivity for sounds from three different categories which were matched on familiarity as well as low-level acoustic structure. Keeping in mind the context-modulation of the

auditory cortex (Bartlett and Wang, 2005), the task ensured a purely acoustic-level engagement of the subjects, thereby ruling out other higher order effects like attention across the stimuli. Additionally, we employed a ‘sparse temporal sampling’ design to acquire the functional imaging data. This was done to ensure minimum interference from scanner noise as well as to tap in into both the cortical and subcortical networks involved in auditory perception which may not be revealed during continuous acquisition designs (Langers and Melcher, 2011).

EXPREIMENTAL PROCEDURES

Participants

Twenty right-handed participants (age range: 20–41 years, 10 females) were tested. All participants spoke at least two languages and had similar educational backgrounds with no previous history

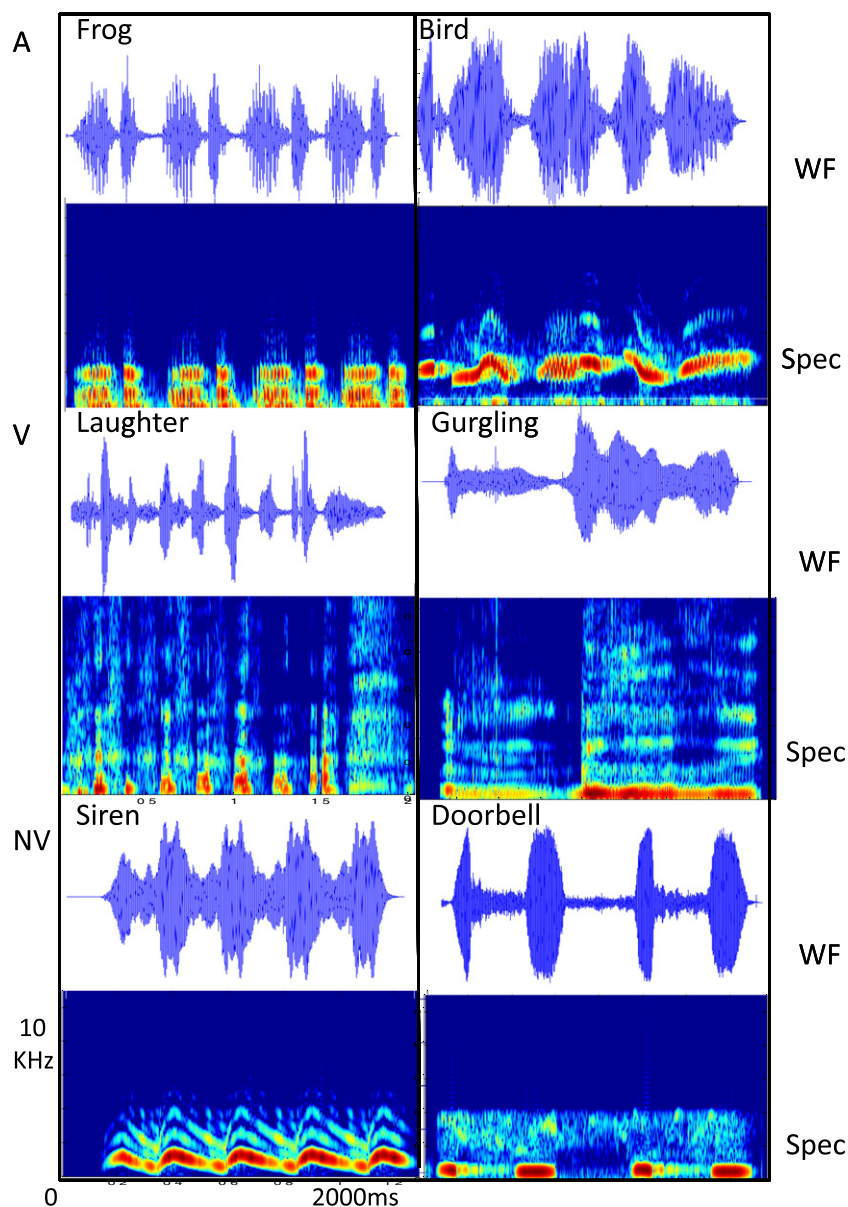


Fig. 1. Waveforms (WF) and spectrograms (Spec) of representative sounds from each category, A, V and NV are shown.

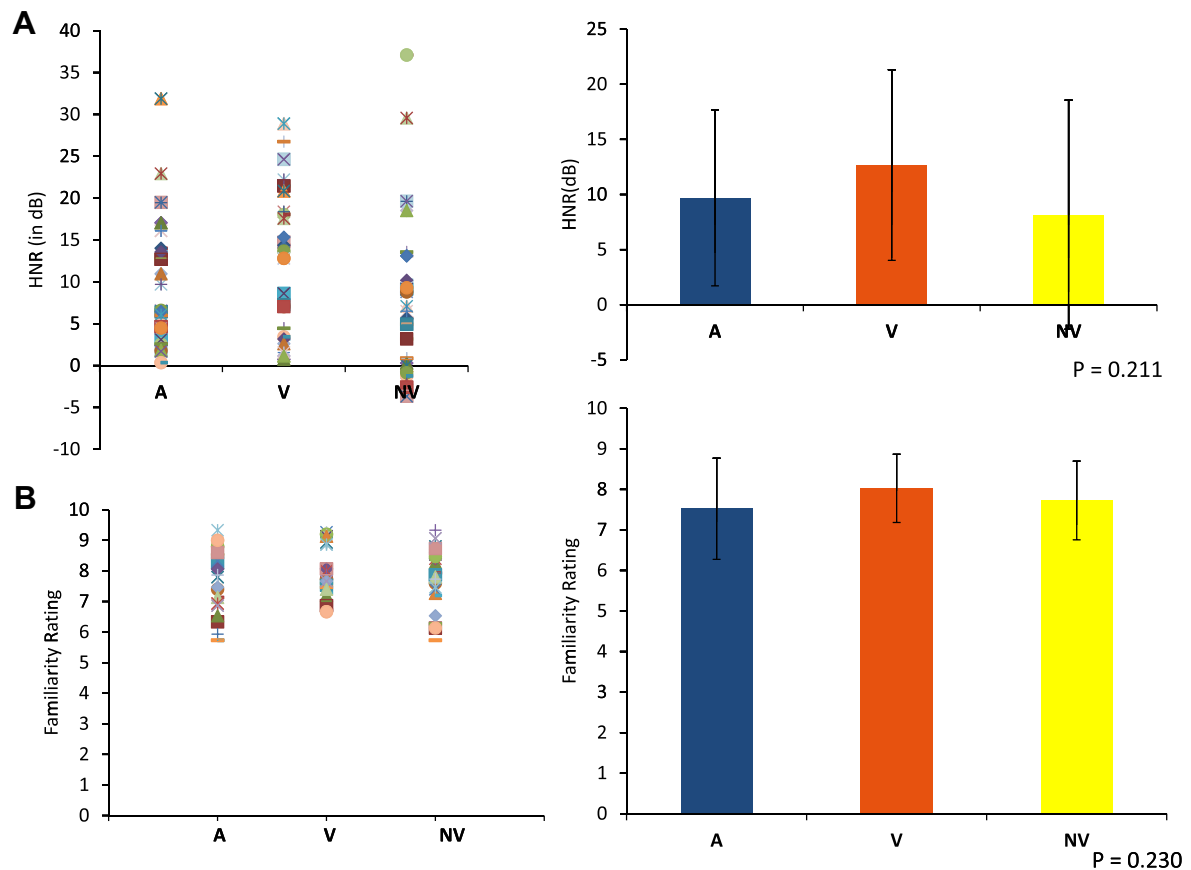


Fig. 2. (A) The distribution of HNR across the three categories A (Animal), V (Vocal) and NV (Environmental) showed no significant differences. (B) Familiarity ratings of sounds from the three categories did not differ significantly.

of neurological and psychiatric disorders, or hearing impairment. Informed consent was obtained from all participants as required by the Institute Ethics Committee.

Stimulus creation and presentation

Sound stimuli consisted of 75 sound pairs taken from various online databases. These included 25 sound pairs each from three different categories – A: animal sounds (e.g. frog croaking, bird singing), NV: environmental sounds (e.g. river flowing, telephone ringing) and V: human vocal non-speech sounds (e.g. coughing, laughing), (see Appendix 1 for list). All sound stimuli were of 2.0 ± 0.25 -s duration, matched for total RMS power using Goldwave (www.goldwave.com) software. Each sound pair consisted of a loud version (LS, 200% power) and soft version (SS, 50% power) of the same sound. Sound stimuli were converted to one channel (mono) and 24-bit digitized at a sampling rate of 22,050 Hz. All sound stimuli were subsequently rated by 15 listeners for familiarity on a numerical scale (as described below).

Analysis of sounds

HNR. The harmonic-to-noise ratio for each sound was calculated using PRAAT (<http://www.fon.hum.uva.nl/praat/>) software as described in Leaver and Rauschecker (2010). The HNR algorithm calculates the degree of periodicity in a signal by measuring the ratio of the strength of the harmonic and noisy components of a signal (Boersma, 1993). HNR is calculated in a single time window as follows:

$$\text{HNR} = 10 * \log_{10} \left[\frac{r'(t_{\max})}{1 - r'(t_{\max})} \right]$$

where $r'(t_{\max})$ is the energy of the harmonic components over time and $(1 - r'(t_{\max}))$ represents the strength of the noisy component of the signal. The values are represented in dB. Thus, higher HNR values denote a harmonic stimulus with a high spectral structure, while lower values indicate a noisy stimulus.

Familiarity rating

Fifteen participants were asked to rate the familiarity of all 75 sounds on a scale of 0–10. The instructions to the participant described the level of familiarity for a particular rating. A high level of familiarity for the sound was to be rated between 7 and 10, a moderate level of familiarity between 4 and 6 and a low level of familiarity to be rated between 0 and 3. Mean familiarity for all sounds was calculated for each category.

Behavioural training and fMRI task

Prior to the fMRI task, all participants underwent a training run with a short set of sound stimuli consisting of 15 sounds (different from those used in the fMRI task). The sounds were presented in pairs of LS (loud version sound) and SS (soft version of same sound), ordered pseudo-randomly across trials, and the participants were instructed to perform a loudness judgment task using a button-press response. For the fMRI task, there were 75 task-related trials, 25 for each condition (A, V, NV) and 10 null trials. Representative waveforms and spectrograms for sounds from each category are shown in Fig. 1. Each trial was 12 s long

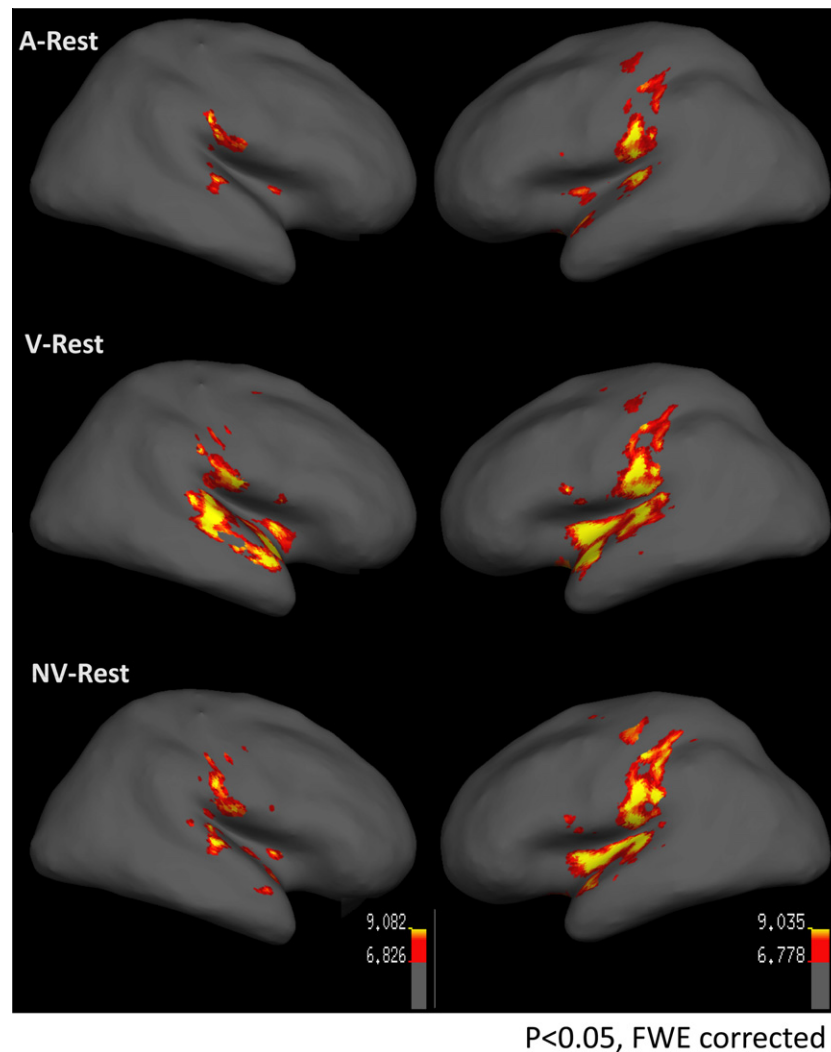


Fig. 3. Condition-wise activations for three categories of sounds, A, V and NV with respect to baseline rest condition. Activations are shown on an inflated standard brain in MNI space. For all three conditions, we observe the recruitment of an extensive brain network comprising the primary and secondary auditory cortices as well as some motor and inferior frontal regions. For a list of co-ordinates and regions activated see [Table 1a](#), [1b](#), and [1c](#).

and consisted of a blank period (3 s) followed by the first sound (2 s), a 1-s gap before the other sound was presented (2 s) and a 4-s period for the participants' response as to whether the first sound was loud or soft indicated by a key press. The order of conditions was pseudo-randomized across trials. The total scan duration was 17 min. The main purpose of the task was to allow the participants to focus strictly on the auditory stimuli and pay attention. The results were used only to assess if the participants understood the task correctly and were able to hear the sounds. The response times of the participants across different conditions did not differ significantly ($f_{59,2} = 1.432$, $p = 0.247$), thereby ruling out any category-based attention differences.

In order to present stimuli without interference from the scanner sound, the technique of *sparse temporal sampling* (STS) was adopted during scanning (Edmister et al., 1999; Hall et al., 1999). In this approach, single volumes are acquired sparsely with a long temporal distance. Using a long repetition time (TR = 12 s) and a regular volume acquisition time (e.g., 2 s), auditory stimuli could be presented with minimal interference from scanner noise. During the training and fMRI scanning session, sound stimuli were delivered binaurally using a Windows

PC and E-prime 1.0 software (Psychology Software Tools) and MR-compatible headphones (In Vivo Eloquence system) were used during scanning.

fMRI data acquisition

Images were acquired on a 3-T Philips Achieva scanner using a standard whole-head coil. A custom-built head holder was used to prevent head movement during the scan. Thirty axial slices (5.0-mm thick, 1.0-mm gap) parallel to the AC-PC line (where the anterior commissure (AC) and posterior commissure (PC) meet) and covering the whole brain were imaged using a T_2^* -weighted gradient-echo EPI sequence (TR = 12 s, TE = 35 ms, flip angle = 90°). A total of 85 volumes were acquired.

The field of view was 230 × 230 mm and matrix size, 64 × 64. The images were reconstructed as a 128 × 128 matrix with 1.79 × 1.79-mm resolution. A high-resolution 3D T_1 -weighted image with the following parameters, TR = 8.4 s; TE = 3.7 ms; flip angle = 8°; 25-cm field of view; 170 slices; 252 × 223 matrix, was acquired before the functional scans.

Table 1a. MNI coordinates as shown by AAL atlas for condition A-Rest at $p < 0.05$, FWE-corrected

A-Rest Brain regions (AAL)	MNI co-ordinates xyz (mm)	T-value
Left STG	–64 –106	7.18
	–66 –268	6.91
Right STG	64 –20 14	10.11
	50 –28 14	7.72
Left STS, STG	–46 –248	11.30
	–482 –4	10.21
Left Insula	–36 –4 –8	8.05
	–3680	6.91
Right insula	4620	8.52
Left thalamus	–6 –320	9.44
	–14 –180	7.98
Right lingual gyrus	10 –32 –2	7.75
Left postcentral gyrus	–52 –2022	11.70
	–50 –2246	8.53
	–44 –3244	7.83
	–32 –2648	7.48
Left rolandic operculum	–44 –42	8.85
Left supramarginal gyrus	–64 –2620	9.26
Right supramarginal gyrus	60 –2024	9.63
Left cerebellum	–28 –58 –30	8.34
	–20 –60 –28	7.44
	–20 –66 –56	7.37
	–30 –62 –52	6.98
	–24 –64 –52	6.87
	–6 –48 –52	8.07
Right cerebellum	14 –54 –24	12.28
	22 –48 –32	8.61
	20 –68 –50	8.25
	26 –60 –56	7.32
	30 –52 –46	7.86
Left olfactory lobe	–248 –14	8.45
Left pallidum	–24 –14 –2	7.89
Right putamen	2862	8.29
Left precentral gyrus	–56620	8.58
Vermis	4 –64 –28	8.77

fMRI data analysis

The first two volumes were discarded to allow for signal equilibration. Functional MRI data were then analysed using SPM5 analysis software (<http://www.fil.ion.ucl.ac.uk/spm>). Functional images were realigned to correct for motion, spatially transformed to standard stereotaxic space (based on the Montreal Neurological Institute [MNI] coordinate system), smoothed with an 8-mm full-width at half-maximum Gaussian kernel to decrease spatial noise prior to statistical analysis. Translational movement in millimetres (x, y, z) and rotational motion in degrees (pitch, roll and yaw) were calculated based on the SPM5 parameters for motion correction of the functional images in each participant. No participants had movement greater than 2.5-mm translation or 3 degrees of rotation; therefore, none were excluded from further analysis. The responses for four conditions (A, V, NV and Rest) were modelled using a canonical Hemodynamic Response Function and temporally filtered using a high-pass filter of 128 s. A fixed effects model was used to analyse single-subject data for three contrasts, A-Rest, V-Rest and NV-Rest, which were then entered into a random effects analysis to identify activations at the group level. One-sample t -tests were utilized to report effects of each of the above contrasts at the group level at an FWE-corrected threshold of $p < 0.05$ to address effects of multiple comparisons. A region of interest (ROI) analysis was then

Table 1b. MNI coordinates as shown by AAL atlas for condition V-Rest at $p < 0.05$, FWE-corrected

V-Rest Brain regions (AAL)	MNI coordinates xyz (mm)	T-value
Left STG, STS	–484 –8	14.27
Right STG	66 –24 12	14.24
	50 –28 12	12.69
Left insula	–36 –2 –8	15.59
Right insula	460 –6	15.99
Right amygdala	20 –2 –16	7.49
Left Hippocampus	–16 –22 –10	7.16
Right lingual gyrus	12 –30 –8	10.48
Right parahippocampal gyrus	24 –24 –20	9.23
Right thalamus	8 –18 –4	7.44
	14 –162	7.23
	8 –244	7.19
Left inferior Frontal operculum	–566 18	9.10
Right superior frontal gyrus	34 –660	7.40
Left rolandic operculum	–50 –2222	13.61
Right rolandic operculum	62 108	6.97
Right supplementary motor area	4646	8.26
	4052	8.02
Right supramarginal gyrus	52 –2636	6.93
Right anterior cingulum	4262	6.98
Right cerebellum	12 –54 –22	9.92
	24 –58 –30	9.50
	22 –46 –32	8.57
	22 –66 –52	8.52
	16 –72 –42	7.29
	10 –42 –48	7.13
Left cerebellum	–28 –60 –28	10.03
Left putamen	–208 –2	6.98
Right putamen	3040	7.63

performed for ROIs selected from the random effects analysis (Tables 2a and 2b) using the MarsBar toolbox (Brett et al., 2002) in SPM5.

RESULTS

Analysis of sounds

HNR. Sounds from all three categories – A, V, NV had HNR with means (standard deviation) ranging from 9.69 (7.8) dB for animal sounds, 12.67 (8.4) dB for human vocal sounds and 8.18 (10.1) dB for environmental sounds. However, there was no significant difference ($f_{2,72} = 1.589$, $p = 0.211$) across the three categories of sounds which ensured that any functional differences that might be seen are not a result of differences in HNR (Fig. 2A).

Familiarity rating

Sounds from all three categories were matched on familiarity ratings and their mean ratings ranged from 7.52 (1.24) for animal sounds, 8.02 (0.84) for human vocalizations and 7.72 (0.97) for environmental sounds. There were no significant differences ($f_{2,72} = 1.5$, $p = 0.230$)

across the three categories of sounds which ensured that any functional differences that might be seen are not a result of differences in familiarity (Fig. 2B).

Cortical networks for auditory perception

Averaged group responses derived from the random effects analysis for each contrast, A-Rest, V-Rest and NV-Rest, activated a bilateral superior temporal network which included the bilateral Superior Temporal Gyri and Superior Temporal Sulci (Fig. 3 and Tables 1a, 1b and 1c). These areas, along with non-primary auditory areas such as bilateral Insulae, Inferior Frontal Gyrus, Right Lingual Gyrus, Right Parahippocampal Gyrus, Right Amygdala, Left Inferior Parietal Lobule and bilateral Thalamus, were also observed in all three conditions (Fig. 4 and Tables 1a, 1b, and 1c) and appear to form a common network for sound perception. In order to compare levels of activations across the conditions, we performed an ROI analysis in regions activated by all conditions. The ROIs were defined as spheres of 5–10 mm radius around peak co-ordinates at an FWE-corrected threshold of $p < 0.05$ and the MarsBar toolbox for SPM was used for extracting BOLD (parameter estimate) values for each ROI. Table 2a lists the ROIs defined. Statistically significant differences were seen only in two clusters of STG and bilateral insula with greater activity for V as compared to A and NV sounds.

Category selectivity

In order to identify category-specific activity in the brain, we performed a conjunction analysis. Category selectivity for a particular condition was defined as the conjunction of the subtraction of one condition with the two others. For example, to study category selectivity for condition A, we identified commonly activated voxels between the contrasts A-V and A-NV. This was done using the inclusive contrast method in SPM5. Even at a lenient threshold of $p < 0.05$ (uncorrected), we did not see any category selectivity for condition A and condition NV. However, a preferential selectivity for human vocal sounds, exhibited by a conjunction of contrasts V-A and V-NV, was observed in bilateral Anterior Superior Temporal Sulci and Gyri, in Right Inferior Frontal Gyrus as well as in Right Middle Temporal Gyrus at an FWE-corrected threshold of $p < 0.05$ (Fig. 5). A subsequent ROI analysis of all six clusters from the conjunction showed a significantly higher BOLD activity in these regions in the V condition as compared to the A and NV conditions (Table 2b).

DISCUSSION

The objective of the current study was to identify the cortical mechanisms underlying category selectivity of species-specific sounds. We used sounds from three natural categories – animal sounds, environmental sounds and human nonspeech sounds which were all matched on acoustic structure attributes to minimize differences in bottom-up processing as well as on familiarity to minimize top-down influences. Our functional imaging results show that there is category selectivity in the cortex

Table 1c. MNI coordinates as shown by AAL atlas for condition NV-Rest at $p < 0.05$, FWE-corrected

NV-Rest Brain regions (AAL)	MNI coordinates xyz (mm)	T-value
Left STG	–64 –26 20	13.93
	–66 –28 8	6.97
Left STS, STG	–48 –22 10	12.46
Right STG	66 –22 14	11.69
Right STS, STG	48 –30 14	9.28
	42 –16 0	8.91
	48 0 –12	8.66
Right Heschl's gyrus	32 –32 8	7.21
Right hippocampus	30 –28 6	7.05
Right insula	46 0 2	11.18
Left parahippocampal gyrus	–18 –24 –18	7.72
Right parahippocampal gyrus	20 –40 –6	8.51
	22 –22 –20	7.80
Left superior frontal gyrus	–22 –8 54	7.86
Right rolandic operculum	62 10 14	7.43
Left supplementary motor area	24 50	9.32
Right supramarginal gyrus	52 –20 24	9.68
Left post central gyrus	–50 –20 24	14.85
Right thalamus	12 –30 2	10.77
	20 –18 16	10.42
	20 –28 14	7.38
Right cerebellum	14 –54 –24	13.05
	22 –66 –52	10.91
	28 –62 –56	10.01
	6 –40 –52	8.00
Left cerebellum	–22 –56 –26	7.85
	–22 –66 –56	7.21
Vermis	4 –60 –4	7.15
	2 –70 –14	7.08
	2 –62 –26	8.50
Right caudate	20 28 2	10.53
	20 –6 18	8.26
	18 4 12	7.12
	20 10 12	7.14
Right putamen	20 12 0	9.26
	28 6 –2	7.72
	28 –16 0	7.79
	30 –24 0	6.87

only for species-specific vocalizations and this is revealed in six clusters in the right and left STG/STS.

A common auditory brain network

All categories of sounds activated a common auditory brain network comprising not just the primary and secondary auditory cortices, but also regions like the insula, parahippocampal gyri and motor cortices. The recruitment of such an extensive brain network is consistent with earlier reports of multimodal interaction in real-world sound perception (Ghazanfar, 2009; Lewis et al., 2011). The insula subserves an important role in practically all sensory functions and has numerous direct connections with thalamic nuclei as well as with the auditory cortex (Bamiou et al., 2003). Lesions in the bilateral insula lead to complete auditory agnosia in patients (Engelien et al., 1995). There is also evidence from functional imaging studies in

Table 2a. Description of ROIs defined to compare activity across the three conditions, A, V and NV

<i>ROI selection</i>			
<i>Auditory brain network</i>			
ROI name	MNI coordinates (x,y,z)	ROI size	Region
LSTG1_5	−46 −24 −8	5-mm sphere	Left superior temporal sulcus
LSTG2_5	−484 −8	5-mm sphere	Left superior temporal sulcus
RaSTG_5	480 −12	5-mm sphere	Right superior temporal sulcus
RpSTG_5	64 −20 14	5-mm sphere	Right superior temporal gyrus
RSTG_5	50 −28 12	5-mm sphere	Right superior temporal gyrus
Lins_10	−36 −2 −8	10-mm sphere	Left insula
Rins_10	460 −6	10-mm sphere	Right insula
LIPL_10	−46 −56 48	10-mm sphere	Left inferior parietal lobule
Lthal_5	−14 −18 0	5-mm sphere	Left medial geniculate nucleus
Rthal2_5	20 −28 14	5-mm sphere	Right medial geniculate nucleus
RLG_5	12 −30 −8	5-mm sphere	Right lingual gyrus
LPHG_NV	−18 −24 −18	Cluster at $p < 0.05$, FWE corrected	Left parahippocampal gyrus
Ramyg_V	20 −2 −16	Cluster at $p < 0.05$, FWE corrected	Right amygdala

Table 2b. Description of the vocal-selective clusters from the conjunction analysis of V-A and V-NV contrasts

<i>ROI selection</i>			
<i>Vocal selective areas</i>			
ROI name	MNI coordinates (x,y,z)	Cluster size	Region
LSTG1_Vsel	−64 −10 0	110	Left superior temporal gyrus
RSTG1_Vsel	50 −22 −6	3	Right superior temporal gyrus
RSTG2_Vsel	58 −12 −6	16	Right superior temporal gyrus
RSTG3_Vsel	56 −8 0	7	Right superior temporal gyrus
RIFG_Vsel	52 34 0	8	Right inferior frontal gyrus
RMTG_Vsel	60 −26 −4	3	Right middle temporal gyrus

humans and electrophysiological studies in primates, postulating a role for the insula in sound detection (Remedios et al., 2009) and an integral role, even in passive listening tasks. The activation of visually-sensitive areas such as the parahippocampal gyrus and the lingual gyrus in all three conditions possibly reflect the ‘imageability’ of these sounds. As shown previously (Engel et al., 2009), these regions are involved in visual form and feature extraction during perception of imageable sounds. In addition to the auditory brain network, we also see activation of somatomotor areas and cerebellum. These structures have been previously implicated in matching or discrimination tasks requiring button-press actions (Brown and Martinez, 2007) as compared to passive listening tasks which typically elicit activations restricted to primary and secondary auditory areas.

The consistent activation of bilateral superior temporal gyri reflects the complex nature of all the sounds. The absence of a distinct locus in Heschl’s gyrus of the primary auditory cortex for all sounds may be attributed to the absence of simple sounds like pure tones in our paradigm.

In addition to the recruitment of an extensive cortical network, we found activations in various subcortical regions like the thalamus, caudate and putamen. A robust activation of the subcortical networks involved in auditory processing is commonly reported in functional imaging studies using a sparse sampling design. Thus, our exper-

iment seems to reveal the subcortical regions like the thalamus (medial geniculate nuclei) which are part of the classical auditory pathway as well as non primary auditory regions like the caudate and putamen as shown recently by Langers and Melcher (2011).

Category selectivity for species-specific vocalizations

The conjunction analysis revealed category-selective clusters for species-specific vocalizations in anterior and middle portions of right STG, MTG and STS as well as a large cluster in left STG, consistent with findings from earlier studies (Binder et al., 1994; Belin, 2006). The activation of these regions by human vocal sounds as compared to animal or environment sounds reflects the species-specificity of this category (Fecteau et al., 2004) and is also consistent with results from primate data, which demonstrate the role of regions homologous to anterior STG and STS to conspecific calls (Rauschecker, 1998; Tian and Rauschecker, 2004). This response specificity in STG associated with human vocal sounds has also been attributed to the processing of complex spectro-temporal patterns in human vocal sounds (Samson et al., 2011). Category selectivity was not observed for any other category of sounds in our study. This finding maybe explored further to investigate the role of evolutionary significance in driving top-down

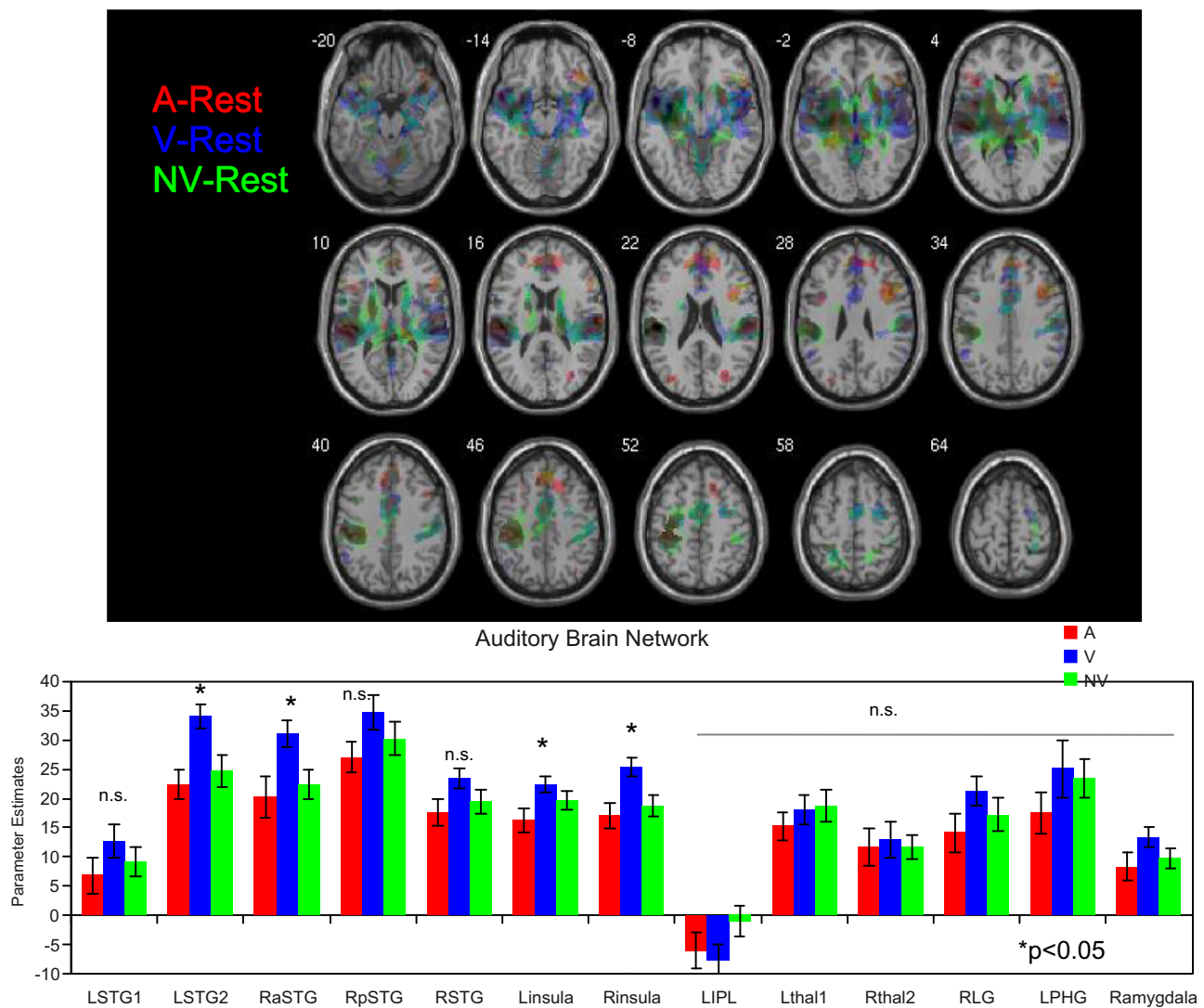


Fig. 4. Brain network for auditory processing. The top panel shows a condition-wise overlay of regions activated by A (red), V (blue) and NV (green) sounds ($p < 0.001$, uncorrected) superimposed on a canonical anatomical image. The lower panel shows a (ROI) region of interest analysis for brain areas activated across all conditions. These include both auditory cortex regions like the bilateral superior temporal gyri as well as non primary auditory areas like the insula. For a list and description of ROIs see Tables 2a and 2b. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mechanisms. The role of recognizing conspecific sounds for survival as well as social interaction has been demonstrated in both humans and non-human primates (Petkov et al., 2008). The idea of a specific 'Voice area' in the brain has been around for some years now (Belin et al., 2000, 2004) and stems from the importance of recognizing conspecific sounds for the purpose of extracting information like speaker identity, speaker intent and affect and other non-linguistic information which is crucial for social interaction. Previous studies focusing on similar mechanisms have suggested the presence of specific acoustic features like harmonicity of vocalizations, which might be driving this specialized processing. However, in the present study we demonstrate that despite being matched on features of HNR and familiarity, species-specific vocalizations are still processed as a distinct category as opposed to animals and environmental sounds in the

bilateral temporal cortex. Although there might be other complex acoustic features that might lead to differential processing of conspecific vocalizations, it is yet to be demonstrated. At the same time, the evolutionary significance of this category cannot be ignored. Future studies looking at category selectivity may use sounds from different categories matched on intermediate level features of spectro-temporal complexity to further elucidate this idea and explore the role of evolutionary significance in category selectivity in the brain.

Acknowledgements—This work was supported by external funding from the Department of Information Technology, Government of India and core funding from the National Brain Research Centre. The authors would like to thank Mr. J. Ahlawat and Mr. Abhilash for assistance in data acquisition and Ms. S. Cherodath and Ms. A. Chakraborty for comments on the manuscript.

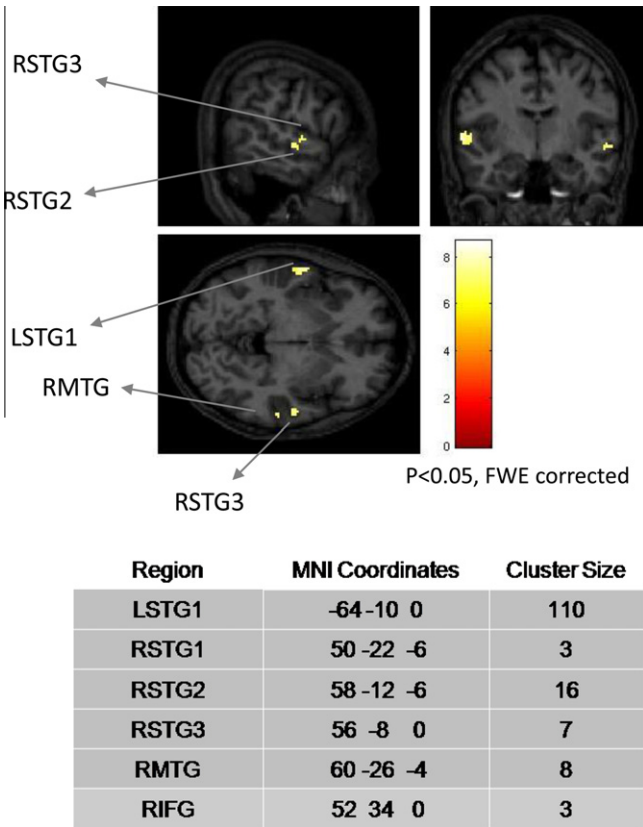


Fig. 5. Voice-selective areas. The conjunction analysis using an inclusive contrast for V-A and V-NV contrasts gave the ‘voice-selective clusters’ shown in the upper panel.

APPENDIX 1. LIST OF ALL SOUNDS USED FROM THREE CATEGORIES.

Animal	Vocal	Environmental
Lion roar	Child cough	Firecracker
Frog	Cough	Lake
Sheep	Baby gurgle	River flowing
Duck	Throat choking	Hammer
Rooster	Child sighing	Thunder
Horse	Group laugh	Footsteps
Buffalo	Child rolling Rs	Stream
Dog bark	Female laugh	Wind blowing
Parrot	Female cry	Flute
Dog sound	Nasal sound	Church bell
Cricket	Male laugh	School gong
Mosquito	High-pitched laugh	Instrumental music
Cicada	Infant babble	Train
Bird	Baby crying	Cycle bell
Elephant	Male singing	Cuckoo clock
Cow moo	Gurgling	Button bell
Bee	Deep Male singing	Factory siren
Bird call	Throaty rolling	Ambulance siren
Fly	Humming	Cuckoo doorbell
	Child Singing	Screech
	Tongue rolling	Police siren
		Telephone
		Doorbell

REFERENCES

Bamiou DE, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Res Rev* 42:143–154.

Bartlett EL, Wang X (2005) Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 94:83–104.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.

Belin P, Fecteau S, Bédard C (2004) Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 8:129–135.

Belin P (2006) Voice processing in human and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361:2091–2107.

Binder JR, Rao SM, Hammeke TA, Yetkin FZ, Jesmanowicz A, Bandettini PA, et al. (1994) Functional magnetic resonance imaging of human auditory cortex. *Ann Neurol* 35:662–672.

Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc Inst Phon Sci* 17:97–110.

Brett M, Anton J, Valabregue R, Poline J (2002) Region of Interest analysis using an SPM toolbox. In: 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. Available on CD-Rom in NeuroImage, Vol. 16(2).

Brown S, Martinez MJ (2007) Activation of premotor vocal areas during musical discrimination. *Brain Cogn* 63:59–69.

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

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* 7:89–97.

- Engel LR, Frum C, Puce A, Walker NA, Lewis JW (2009) Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage* 47:1778–1791.
- Engelien A, Silbersweig D, Stern E, Huber W, Döring W, Frith C, Frackowiak RS (1995) The functional anatomy of recovery from auditory agnosia. A PET study of sound categorization in a neurological patient and normal controls. *Brain* 118: 1395–1409.
- Fecteau S, Armony JL, Joanette Y, Belin P (2004) Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23:840–848.
- Ghazanfar AA (2009) The multisensory roles for auditory cortex in primate vocal communication. *Hear Res* 258:113–120.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213–223.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci* 12:698–701.
- Langers DRM, Melcher JR (2011) Hearing without listening: functional connectivity reveals the engagement of multiple non auditory networks during basic sound processing. *Brain Connectivity* 1(3):233–244.
- Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30:7604–7612.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148–5158.
- Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C (2009) Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J Neurosci* 29:2283–2296.
- Lewis JW, Talkington WJ, Puce A, Engel LR, Frum C (2011) Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *J Cogn Neurosci* 23:2079–2101.
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath-Logothetis M, Logothetis NK (2008) A voice region in the monkey brain. *Nat Neurosci* 11:367–374.
- Rauschecker JP (1998) Cortical processing of complex sounds. *Curr Opin Neurobiol* 8:516–521.
- Remedios R, Logothetis NK, Kayser C (2009) An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J Neurosci* 29:1034–1045.
- Samson F, Zeffiro T, Toussaint A, Belin P (2011) Stimulus complexity and categorical effects in human auditory cortex: an activation likelihood estimation meta-analysis. *Front Psychol* 1:1–23.
- Saygin AP, Leech R, Dick F (2010) Nonverbal auditory agnosia with lesion to Wernicke’s area. *Neuropsychologia* 48:107–113.
- Tian B, Rauschecker JP (2004) Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 92:2993–3013.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1–7.

(Accepted 22 March 2012)
(Available online 20 April 2012)