Behavioral/Systems/Cognitive

Humans Mimicking Animals: A Cortical Hierarchy for Human Vocal Communication Sounds

William J. Talkington,[1] Kristina M. Rapuano,[1] Laura A. Hitt,[2] Chris A. Frum,[1] and James W. Lewis[1]

[1]Center for Neuroscience, Center for Advanced Imaging, Departments of Physiology and Pharmacology, and [2]School of Theatre and Dance, College of Creative Arts, West Virginia University, Morgantown, West Virginia 26506

Numerous species possess cortical regions that are most sensitive to vocalizations produced by their own kind (conspecifics). In humans, the superior temporal sulci (STSs) putatively represent homologous voice-sensitive areas of cortex. However, superior temporal sulcus (STS) regions have recently been reported to represent auditory experience or "expertise" in general rather than showing exclusive sensitivity to human vocalizations per se. Using functional magnetic resonance imaging and a unique non-stereotypical category of complex human non-verbal vocalizations— human-mimicked versions of animal vocalizations—we found a cortical hierarchy in humans optimized for processing meaningful conspecific utterances. This left-lateralized hierarchy originated near primary auditory cortices and progressed into traditional speech-sensitive areas. Our results suggest that the cortical regions supporting vocalization perception are initially organized by sensitivity to the human vocal tract in stages before the STS. Additionally, these findings have implications for the developmental time course of conspecific vocalization processing in humans as well as its evolutionary origins.

Introduction

In early childhood, numerous communication disorders develop or manifest as inadequate processing of vocalization sounds in the CNS (Abrams et al., 2009). Cortical regions in several animals have been identified that are most sensitive to vocalizations produced by their own species (conspecifics) including some bird species, marmosets and cats, macaque, chimpanzee, and humans (Belin et al., 2000; Tian et al., 2001; Wang and Kadia, 2001; Hauber et al., 2007; Petkov et al., 2008; Taglialatela et al., 2009). Voice-sensitive regions in humans have been traditionally identified bilaterally within the superior temporal sulci (STSs) (Belin et al., 2000, 2002; Lewis et al., 2009). However, by showing preferential superior temporal sulcus (STS) activity to artificial nonvocal sounds after perceptual training, recent studies consider these regions to be "higher-order" auditory cortices that function as substrates for more general auditory experience— contrary to these areas behaving in a domain-specific manner solely for vocalization processing (Leech et al., 2009; Liebenthal et al., 2010). Thus, we questioned whether preferential cortical sensitivity to intrinsic human vocal tract sounds, those uniquely produced by human source-and-filter articulatory structures (Fitch et al., 2002), could be revealed in earlier "low-level" acoustic signal processing stages closer to frequency-sensitive primary auditory cortices (PACs).

Within human auditory cortices, we predicted that there should be a categorical hierarchy reflecting an increasing sensitivity to one's conspecific vocalizations and utterances. Previous studies investigating cortical voice-sensitivity in humans have compared responses to stereotypical speech and non-speech vocalizations with responses to other sounds categories, including animal vocalizations and environmental sounds (Belin et al., 2000, 2002; Fecteau et al., 2004). However, these comparisons did not always represent gradual categorical differences, especially when using broadly defined samples of "environmental sounds." Thus, in the current study, we incorporated naturally produced human-mimicked versions of animal vocalizations (Lass et al., 1983). Human-mimicked animal vocalizations acted as a crucial intermediate vocalization category of human-produced stimuli, acoustically and conceptually bridging between animal vocalizations and stereotypical human vocalizations. We therefore avoided confounds associated with using overlearned acoustic stimuli when characterizing these early vocalization processing networks (e.g., activation of acoustic schemata; Alain, 2007). Using high-resolution functional magnetic resonance imaging (fMRI), our findings suggest that the cortical networks mediating vocalization processing are not only organized by verbal and prosodic non-verbal information processing (left and right hemispheres, respectively), but also that the left hemisphere processing hierarchy becomes organized along an acoustic dimension that reflects increasingly meaningful conspecific communication content.

Received March 6, 2012; revised April 26, 2012; accepted April 28, 2012.

Author contributions: W.J.T., K.M.R., and J.W.L. designed research; W.J.T., K.M.R., C.A.F., and J.W.L. performed research; L.A.H. contributed unpublished reagents/analytic tools; W.J.T., analyzed data; W.J.T., K.M.R., L.A.H., C.A.F., and J.W.L. wrote the paper.

This work was supported by the NCRR NIH Centers of Biomedical Research Excellence Grant E15524 [to the Sensory Neuroscience Research Center of West Virginia University (WVU)] and the National Science Foundation under Grant No. EPS-1003907, in the form of the WVU High-Performance Computing Initiative. W.J.T. was supported by the United States Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG). We thank Bachelors of Fine Arts Acting students from the Division of Theatre and Dance, College of Creative Arts (WVU) as well as foreign language speakers for their assistance in creating verbal stimuli. We also thank Drs. Brandi Snyder, George Spirou, Glenn Marrs, Aric Agmon, and Jack Gallant for critically reading earlier versions of the manuscript. Additionally, we thank Dr. James P. Lewis and Adam Dorsey at the High Performance Computing facilities of WVU.

Correspondence should be addressed to Dr. James W. Lewis, Department of Physiology and Pharmacology, PO Box 9229, West Virginia University, Morgantown, WV 26506. E-mail: jwlewis@hsc.wvu.edu.

DOI:10.1523/JNEUROSCI.1118-12.2012

Materials and Methods

Participants. We studied 22 right-handed participants (11 female; average age: 27.14 years ± 5.07 years SD). All participants were native English speakers with no previous history of neurological, psychiatric disorders, or auditory impairment, and had self-reported normal ranges of hearing. Each participant had typical structural MRI scans, was free of medical disorders contraindicative to MRI, and was paid for his or her participation. Informed consent was obtained from each participant following procedures approved by the West Virginia University Institutional Review Board.

Vocalization sound stimulus creation and acoustic attributes. We prepared 256 vocalization sound stimuli. Sixty-four stimuli were in each of four sound categories, including human-mimicked animal vocalizations, corresponding real-world animal vocalizations, foreign speech samples (details below), and nine predetermined English speech examples with neutral affect (performed by 13 native-English speaking theatre students). The animal vocalizations were sourced from professionally recorded compilations of sounds (Sound Ideas, Inc; 44.1 kHz, 16-bit). The three remaining vocalization categories were digitally recorded in our laboratory within a sound-isolated chamber (Industrial Acoustics Company) using a Sony PCM-D1 Linear PCM recorder (sampled at 44.1 kHz, 16-bit).

Six non-imaging volunteers recorded human-mimicked versions of corresponding animal vocalization stimuli. Each mimicker attempted to match the spectrotemporal qualities of the real-world animal vocalizations. A group of four listeners then assessed the acoustic similarity of each animal-mimic pair until reaching a consensus for the optimal mimicked recordings. A subset of our fMRI subjects ($n = 18/22$) psychophysically rated all of the animal vocalization and human-mimics after their respective scanning sessions. Subjects were asked to rate each stimulus (button response) along a 5-point Likert-scale continuum to assess the "animal-ness" (low-score, 1 or 2) or "human-ness" (high score, 4 or 5) quality of the recording. Stimuli rated ambiguously along this dimension were given a score of three (3). The number of subjects who correctly categorized each animal or human-mimicked vocalization are displayed in Table 1.

The foreign speech samples used in this study were performed by native speakers of six different non-Romance and non-Germanic languages: (1) Akan, (2) Farsi, (3) Hebrew, (4) Hindi, (5) Mandarin, and (6) Yoruban. The Hindi, Farsi, and Yoruban speech samples were produced by female speakers and the Mandarin, Hebrew, and Akan speech samples were produced by male speakers. The foreign speakers were asked to record short phrases with communicative content in a neutral tone. The speech content was determined by the speakers. However, it was suggested that they discuss everyday situations to help ensure a neutral emotional valence in the speech samples.

The English vocalizations were modified versions of complete sentences used in an earlier study (Robins et al., 2009); additional phrasing was added to each stimulus to increase its overall length so that it could be spoken over a long enough timeframe (see below) with neutral emotional valence. All sound stimuli were edited to within 2.0 ± 0.5 s duration, matched for average root mean square power, and a linear onset/offset ramp of 25 ms was applied to each sound (Adobe Audition 2.0, Adobe Inc.). All stimuli were recorded in stereo, but subsequently converted to mono (44.1 kHz, 16-bit) and presented to both ears, thereby removing any binaural spatial cues present in the signals.

All of the sound stimuli were quantitatively analyzed; the primary motivation for these analyses was to acoustically compare the stimuli in each animal-mimic pair (Table 1). The harmonic content in each stimulus was quantified with a harmonics-to-noise ratio (HNR) using Praat software (http://www.fon.hum.uva.nl/praat/) (Boersma, 1993). HNR algorithm parameters were the default settings in Praat (Time step (seconds): 0.01; Minimum Pitch (Hz): 75; Silence threshold: 0.1; Periods per window: 1.0). Weiner entropy and spectral structure variation (SSV) were also calculated for each sound stimulus (Reddy et al., 2009; Lewis et al., 2012). We used a freely available custom Praat script to calculate Weiner entropy values (http://www.gbeckers.nl/; Gabriel J.L. Beckers, Ph.D.); the script was modified to additionally calculate SSV values, which are derived from Weiner entropy values.

Scanning paradigms. Participants were presented with 256 sound stimuli and 64 silent events as baseline controls using an event-related fMRI paradigm (Lewis et al., 2004). All sound stimuli were presented during fMRI scanning runs via a Windows PC (CDX01, Digital Audio sound card interface) installed with Presentation software (version 11.1, Neurobehavioral Systems) through a sound mixer (1642VLZ pro mixer, Mackie) and high-fidelity MR-compatible electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax Ltd.), worn under sound-attenuating ear muffs. The frequency response of the ear buds was relatively flat out to 20 kHz (±4 dB) and the sound delivery system imparted 75 Hz high-pass filtering (18 dB/octave) to the sound stimuli.

The scanning session consisted of eight distinct functional imaging runs; the 256 vocalization and 64 silent stimuli were presented in pseudorandom order (with no consecutive silent event presentations) and counterbalanced by category across all runs. Participants were instructed to listen to each sound stimulus and press a predetermined button on an MRI-compatible response pad as close to the end of the sound as possible ("End-of-Sound" task). This task aimed to ensure that the participants were closely attending to the sound stimuli, but not necessarily making any overt and/or instructed cognitive discrimination.

Using techniques described previously from our laboratory, a subset of participants ($n = 5$) participated in an fMRI paradigm designed to tonotopically map auditory cortices (Lewis et al., 2009). Briefly, tonotopic gradients were delineated in each subject's hemispheres using a "Winner-Take-All" (WTA) algorithm for calculating preferential blood-oxygenated level-dependent (BOLD) responses to three different frequencies of pure-tones and one-octave bandpass noises relative to "silent" events: 250 Hz (Low), 2000 Hz (Medium), and 12,000 Hz (High). An uncorrected node-wise statistical threshold of $p < 0.001$ was applied to each subject's WTA cortical maps; tonotopic gradients were then spatially defined in regions that exhibited contiguous Low-Medium-High progressions of preferential frequency responses along the cortical mantle. The tonotopic gradients of all subjects were then spatially averaged, regardless of gradient direction, on the common group cortical surface model (created by averaging the surface coordinates of all 22 fMRI participants, see below). This effectively created a probabilistic estimate of PACs for our group of participants to be used as a functional landmark. These results were in agreement with anatomical studies that implicate the likely location of human primary auditory cortex (PAC) to be along or near the medial two-thirds of Heschl's gyrus (HG) (Morosan et al., 2001; Rademacher et al., 2001).

Magnetic resonance imaging data collection and preprocessing. Stimuli were presented during relative silent periods without functional scanner noise by using a clustered-acquisition fMRI design (Edmister et al., 1999; Hall et al., 1999). Whole-head, spiral in-and-out images (Glover and Law, 2001) of the BOLD signals were acquired on all trials during functional sessions including silent events as a control condition using a 3T GE Signa MRI scanner. A stimulus or silent event was presented every 9.3 s, and 6.8 s after event onset BOLD signals were collected as 28 axial brain slices approximately centered on the posterior superior temporal gyrus (STG) with $1.875 \times 1.875 \times 2.00\ \text{mm}^3$ spatial resolution (TE = 36 ms, Operational TR = 2.3 s volume acquisition, FOV = 24 mm). The presentation of each stimulus event was triggered by the MRI scanner via a TTL pulse. At the end of functional scanning, whole brain T1-weighted anatomical MR images were acquired with a spoiled gradient recalled acquisition in steady state pulse sequence (1.2 mm slices with $0.9375 \times 0.9375\ \text{mm}^2$ in plane resolution). Both paradigms used identical functional and structural scanning sequences.

All functional datasets were preprocessed with Analysis of Functional NeuroImages (AFNI) and associated software plug-in packages (http://afni.nimh.nih.gov/) (Cox, 1996). The 20th volume of the final scan, closest to the anatomical image acquisition, was used as a common registration image to globally correct motion artifacts due to head translations and rotations.

Individual subject analysis. Three-dimensional cortical surface reconstructions were created for each subject from their respective anatomical data using Freesurfer (http://surfer.nmr.mgh.harvard.edu) (Dale et al.,

Table 1. Acoustic attributes and psychophysical results for real-world animal vocalizations and their corresponding human-mimicked versions

	HNR (dB)		Weiner entropy		SSV		Number correctly categorized (*n* = 18 max)	
Description	Animal	Mimic	Animal	Mimic	Animal	Mimic	Animal	Mimic
Baboon groan	5.264	**11.204**	−7.554	**−8.410**	0.747	**4.286**	7	**15**
Baboon grunt #1	8.721	**10.847**	−7.834	**−7.693**	2.047	**2.254**	10	**17**
Baboon grunt #2	4.915	**10.605**	−7.253	**−6.412**	7.030	**1.792**	14	**17**
Baboon grunt #3	7.440	**11.868**	−9.668	**−7.456**	1.293	**2.119**	3	**18**
Baboon grunt #4	7.009	**7.362**	−10.281	**−7.871**	2.170	**3.955**	10	**18**
Baboon scream	6.568	**10.785**	−8.156	**−7.108**	1.541	**1.006**	14	**18**
Bear roar #1	4.711	**15.294**	−10.347	**−6.480**	4.272	**1.680**	15	**15**
Bear roar #2	11.079	**18.570**	−9.748	**−6.014**	6.672	**3.457**	13	**16**
Boar grunt	0.776	**5.203**	−7.096	**−5.553**	2.237	**1.881**	18	**14**
Bull bellow	20.348	**17.915**	−9.862	**−8.041**	1.925	**1.235**	15	**16**
Camel groan	15.726	**14.431**	−11.696	**−7.169**	2.696	**1.864**	15	**17**
Cat growl	2.424	**1.423**	−8.342	**−6.455**	2.050	**3.078**	15	**14**
Cat meow	12.681	**24.050**	−7.149	**−7.444**	7.577	**6.256**	16	**11**
Cat purr	1.039	**1.238**	−7.717	**−5.166**	0.799	**0.676**	15	**16**
Cattle bellow	13.386	**23.204**	−9.341	**−8.776**	4.532	**1.982**	18	**16**
Cattle cry	3.954	**11.117**	−7.148	**−5.963**	5.846	**0.611**	17	**16**
Chimp chatter #1	16.442	**16.858**	−5.816	**−5.491**	5.279	**5.519**	16	**10**
Chimp chatter #2	11.088	**16.777**	−3.940	**−5.444**	3.345	**2.757**	16	**9**
Chimp chatter #3	5.567	**9.266**	−4.422	**−5.969**	3.219	**1.345**	12	**18**
Chimp grunting #1	1.691	**5.432**	−4.974	**−5.449**	3.840	**1.539**	9	**17**
Chimp grunting #2	6.962	**3.504**	−4.695	**−5.192**	1.474	**0.879**	6	**14**
Chimp scream #1	22.260	**19.542**	−6.697	**−5.286**	5.172	**3.649**	14	**6**
Chimp scream #2	4.861	**22.104**	−5.133	**−5.538**	2.504	**6.852**	18	**1**
Cougar scream	2.369	**1.161**	−10.671	**−6.892**	3.103	**2.160**	18	**14**
Coyote howl #1	25.805	**28.485**	−10.264	**−9.060**	2.549	**0.273**	16	**12**
Coyote howl #2	27.335	**16.599**	−11.434	**−9.702**	2.043	**2.640**	12	**9**
Dog whimper	11.748	**18.342**	−5.874	**−7.160**	2.283	**1.522**	12	**15**
Dog bark #1	6.141	**9.715**	−8.108	**−5.573**	3.236	**4.961**	16	**17**
Dog bark #2	1.740	**3.084**	−5.885	**−3.831**	3.544	**1.785**	17	**18**
Dog bark #3	2.134	**2.685**	−5.850	**−6.816**	6.637	**1.179**	16	**15**
Dog bark #4	3.789	**3.873**	−7.763	**−6.191**	1.024	**1.740**	16	**16**
Dog bark #5	7.269	**11.672**	−7.181	**−5.610**	9.326	**4.782**	16	**14**
Dog bark #6	12.261	**7.754**	−4.681	**−4.716**	4.469	**2.870**	7	**12**
Dog cry	7.183	**8.491**	−8.275	**−6.157**	5.348	**1.535**	17	**15**
Dog growl #1	3.611	**7.622**	−8.231	**−6.803**	2.878	**1.580**	16	**16**
Dog growl #2	0.066	**5.967**	−4.482	**−4.506**	0.440	**0.625**	17	**17**
Dog growl #3	4.685	**6.185**	−8.829	**−6.677**	3.450	**2.176**	18	**12**
Dog moan	13.099	**13.796**	−8.845	**−8.830**	3.640	**0.885**	16	**18**
Donkey bray	7.934	**8.712**	−7.257	**−6.653**	1.752	**2.043**	16	**14**
Gibbon call #1	12.282	**26.346**	−8.592	**−6.804**	1.968	**9.076**	15	**11**
Gibbon call #2	29.316	**20.676**	−10.063	**−6.627**	1.439	**1.320**	13	**11**
Gibbon call #3	19.485	**23.449**	−9.488	**−7.700**	2.287	**6.012**	14	**17**
Goat bleat #1	0.331	**5.688**	−6.721	**−6.045**	0.827	**3.214**	15	**18**
Goat bleat #2	3.042	**10.723**	−7.565	**−5.373**	10.231	**1.473**	10	**10**
Grizzly roar #1	2.056	**1.439**	−5.702	**−5.302**	0.774	**3.304**	18	**15**
Grizzly roar #2	2.868	**8.850**	−5.760	**−6.017**	1.126	**2.914**	18	**16**
Hippo grunt	3.550	**7.648**	−9.218	**−6.560**	4.377	**5.261**	15	**16**
Hyena bark	23.345	**23.635**	−9.963	**−9.170**	3.556	**1.742**	12	**5**
Monkey chitter #1	11.443	**13.596**	−5.678	**−6.231**	4.118	**4.652**	11	**14**
Monkey chitter #2	2.459	**4.303**	−6.583	**−5.649**	3.653	**2.203**	12	**14**
Moose grunt	6.232	**11.818**	−8.201	**−5.254**	3.564	**0.656**	17	**16**
Panda bleat	3.546	**12.635**	−5.668	**−7.493**	0.582	**2.270**	15	**16**
Panda cub #1	19.358	**18.818**	−5.701	**−5.861**	3.649	**3.854**	16	**13**
Panda cub #2	19.799	**18.826**	−5.626	**−5.234**	4.478	**3.531**	11	**3**
Panther cub	6.819	**9.460**	−7.545	**−5.168**	0.896	**1.272**	10	**16**
Pig grunt	1.222	**2.911**	−4.826	**−5.284**	5.737	**2.035**	14	**14**
Pig squeal	1.069	**6.936**	−5.774	**−6.139**	3.872	**2.596**	18	**12**
Primate call #1	15.225	**18.292**	−7.029	**−4.849**	7.780	**3.251**	15	**16**
Primate call #2	15.636	**25.304**	−4.604	**−5.031**	8.532	**5.418**	17	**13**
Sheep bleat	9.470	**22.502**	−9.015	**−7.564**	3.198	**1.347**	11	**12**
Wildcat growl	6.158	**6.081**	−10.876	**−6.765**	6.979	**0.511**	16	**17**
Wolf howl	37.723	**23.689**	−10.885	**−9.168**	1.744	**1.538**	3	**13**
Average	9.428	**12.361**	−7.574	**−6.465**	3.538	**2.627**	14.000	**14.048**
SD	8.169	**7.363**	2.020	**1.286**	2.286	**1.771**	3.590	**3.641**

Each animal-mimic pair is listed (mimic values are in bold type) along with each sound's respective acoustic measurements including HNR, Weiner entropy, and SSV. The last column displays the proportion of subjects for each stimulus who correctly categorized the vocalizations as animal- or human-produced (i.e. animal vocalizations given a 1 or 2 score, human vocalizations given a 4 or 5 score). The acoustic attributes were also calculated for the foreign and English stimulus categories (SDs in parentheses) for comparison, though we did not include these measures in any detailed analyses: HNR, English: 8.708dB (4.220), Foreign: 7.864dB (5.077); Weiner Entropy, English: −5.891 (0.959), Foreign: −5.861 (1.152); SSV, English: 4.077, (1.717), Foreign: 3.500 (1.975).

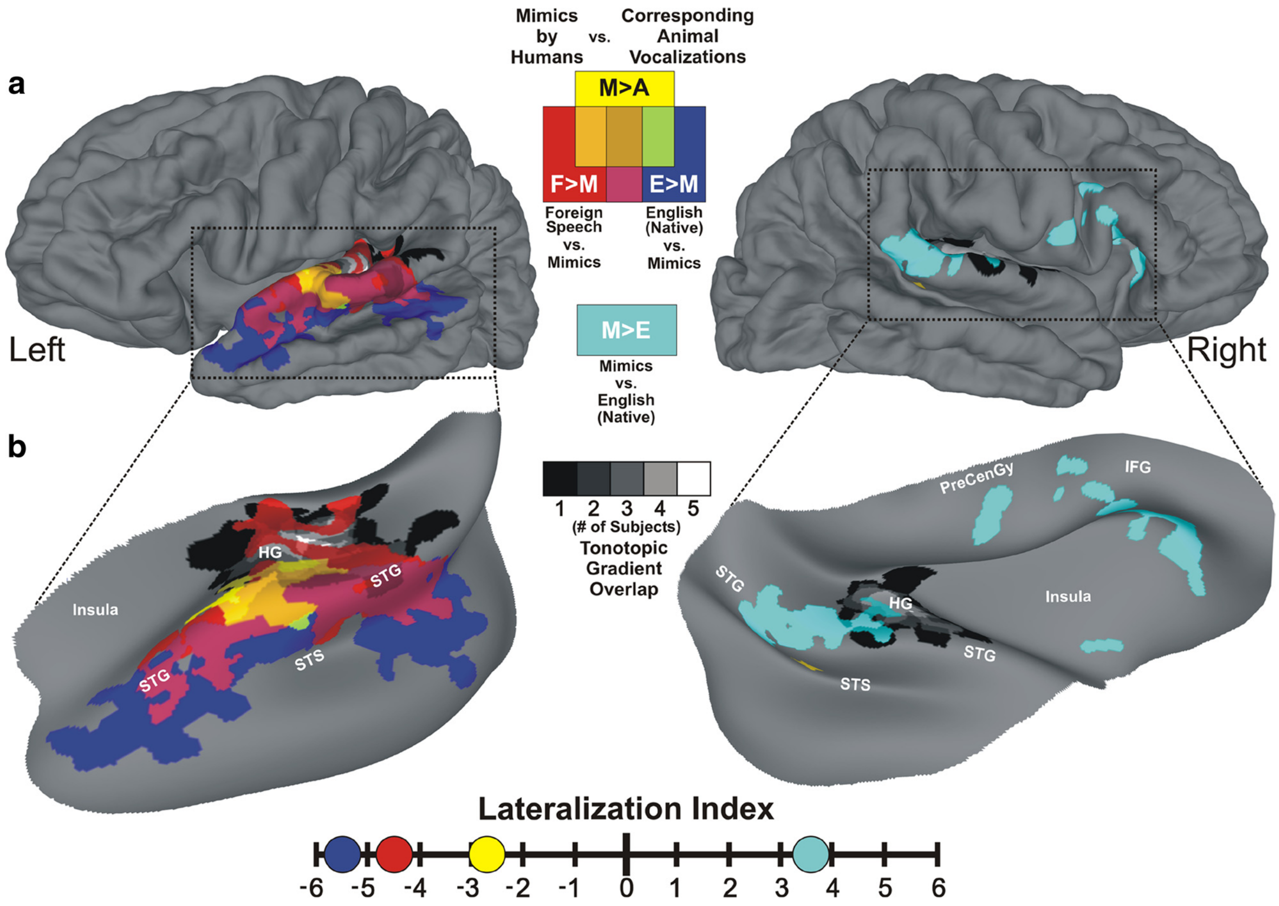

Figure 1. Conspecific vocalization processing hierarchy in human auditory cortex. ***a***, Group-averaged ($n = 22$) functional activation maps displayed on composite hemispheric surface reconstructions derived from all subjects. ***b***, To better visualize the data, we inflated and rotated cortical projections within the dotted outlines in ***a***. The spatial locations of tonotopic gradients from five subjects were averaged (black-to-white gradients) and located along HG. Mimic-sensitive regions (M>A) are depicted by yellow hues, sensitivity to foreign speech samples versus mimic vocalizations (F>M) are depicted by red hues, and sensitivity to native English speech versus mimic vocalizations (E>M) is depicted by dark blue. Regions preferentially responsive to mimic vocalizations versus English speech samples (M>E) are depicted by cyan hues. Corresponding colors indicating functional overlaps are shown in the figure key. TFCE was applied to all data and they were permutation-corrected for multiple comparisons to $p < 0.05$. To quantify the laterality of these functions, we calculated and plotted lateralization indices using threshold- and whole-brain region of interest (ROI)-free methods. Lateralization indices showed increasingly left-lateralized function (negative values indicate a leftward bias) for processing conspecific vocalizations with increasing amounts of communicative content; $LI_{M>A} = -2.68$, $LI_{F>M} = -4.47$, $LI_{E>M} = -5.48$, $LI_{M>E} = +3.59$. Additional anatomy: precentral gyrus (PreCenGy), and inferior frontal gyrus (IFG).

1999; Fischl et al., 1999). These surfaces were then ported to the AFNI-affiliated surface-based functional analysis package Surface Mapping with AFNI (SUMA) for further functional analyses (http://afni.nimh.nih.gov/afni/suma) (Saad et al., 2006). BOLD time-series data were volume-registered, motion-corrected, and corrected for linear baseline drifts. Data were subsequently mapped to each subject's cortical surface model using the SUMA program 3dVol2Surf; data were then smoothed to 4 mm FWHM on the surface using SurfSmooth which implements a heat-kernel smoothing algorithm (Chung et al., 2005). Time-series data were converted to percentage signal change values relative to the average of silent-event responses for each scanning run on a node-wise basis. Functional runs were then concatenated into one contiguous time series and modeled using a GLM-based analysis with AFNI's 3dDeconvolve. Regression coefficients for each subject were extracted from functional contrasts (e.g., MvsA, FvsM, etc.) to be used in group-level analyses (see below). Group analyses were further initiated by standardizing each subject's surface and corresponding functional data to a common spherical space with icosahedral tessellation and projection using SUMA's MapIcosahedron (Argall et al., 2006).

Group-level analyses. Regression coefficients for relevant functional contrasts generated with AFNI/SUMA were grouped across the entire subject pool and entered into two-tailed t tests. These results were then corrected for multiple comparisons in the following manner using Caret6 (Van Essen et al., 2001; Hill et al., 2010): (1) permutation-based corrections were initiated by creating 5000 random permutations of each contrast's t-score map; (2) permuted t-maps were smoothed by an average neighbors algorithm with four iterations (0.5 strength per iteration); (3) threshold-free cluster enhancement (TFCE) was applied to each permutation map (Smith and Nichols, 2009), optimized for use on cortical surface models with parameters: E = 1.0, H = 2.0 (Hill et al., 2010); (4) a distribution ranking maximum TFCE scores was created to find the 95th percentile statistical cutoff value; (5) this value was then applied to the original t-score map to produce the dataset in Figure 1.

Lateralization indices were calculated for each of the functional contrasts described within this manuscript (Fig. 1, M>A, F>M, E>M, and M>E). We accomplished this using a threshold- and whole-brain region of interest (ROI)-free method (Jones et al., 2011). For each functional contrast, we created distributions of non-thresholded t test scores within each hemisphere. After log-transforming these distributions, the centers of each ($-4 \leq t \leq 4$) were fit with parabolic equations to approximate noise in the distributions. Subtracting these noise-approximations from the original score distributions and integrating the results provided a quantitative measure for an individual contrast's strength of activation within a hemisphere. Left and right hemisphere scores were then plotted against one another; the absolute distances of these points from the zero-difference "bilateral" line (slope = 1) represented the relative lateralization of a given function (Fig. 1 illustrates these scores graphically).

Psychophysical affective assessments of sound stimuli. A cohort of non-imaged individuals ($n = 6$) were asked to rate all of the paradigm's stimuli along the affective dimension of emotional potency, or intensity. In our sound isolation booth, participants were seated and asked to rate each stimulus along a 5-point Likert scale: (1) Little or no emotional content, to (5) High levels of emotional content. Note that this scale does not discriminate between positive and negative valence within the stimuli; this scale simply provides a measure of total emotional content (Aeschlimann et al., 2008). Cronbach's α scores were calculated to ensure the reliability of this measure (Cronbach, 1951); the entire set of subjects produced a value of 0.8846 and subsequent removal of each subject individually from the group data consistently produced values between 0.8458 and 0.894, well above the accepted consistency score of 0.7 (Nunnally, 1978). Response means were compared pairwise between each category with nonparametric Kruskal–Wallis tests. These aforementioned tests helped to ensure consistent perceptual effects of our stimuli classes among participants.

Results

Twenty-two native English-speaking (monolingual) right-handed adults were recruited for the fMRI-phase of this project which used a clustered acquisition imaging paradigm in which subjects pressed a button as quickly as possible to indicate the end of each sound. Sound stimuli (2.0 ± 0.5 s) originated from one of four vocalization categories: (1) real-world animal vocalizations, (2) human-mimicked versions of those animal vocalizations, (3) emotionally neutral conversational foreign speech samples that were incomprehensible to our participants, and (4) emotionally neutral English phrases. To create functional landmarks, we mapped the PACs of a subset of participants ($n = 5$) using a modified tonotopy paradigm from our previous work (Lewis et al., 2009). The anatomical extent of each subject's estimated tonotopically sensitive cortices were combined into a group spatial average and depicted by a "heat-map" representation (Fig. 1, gray-scale gradient, see also Fig. 3 for individual maps). The intensity gradient of these averaged data represents the degree of spatial overlap across subjects, providing a probabilistic estimate of PAC locations within our participants. These results were consistent with previous findings indicating that the location of human PACs can be reliably estimated along or near the medial two-thirds of HG (see Materials and Methods).

To assess our hypothesis that the use of non-stereotypical human vocalizations might reveal earlier stages of species-specific vocalization processing, we sought to identify cortical regions preferentially activated by human-mimicked animal vocalizations. Preferential group-averaged BOLD activity to the human-mimicked stimuli relative to their corresponding animal vocalizations was strongly left-lateralized and confined to a large focus in the group-averaged dataset. This activation encompassed regions from the lateral-most aspects of HG, further extending onto the STG, and marginally entering the STS (M>A; Fig. 1, yellow, $p < 0.05$, TFCE and permutation-corrected; Smith and Nichols, 2009). BOLD values in regions defined by this contrast and others discussed below are highlighted in Figure 2. This mimic-sensitive focus (yellow) was located near and partially overlapping functional estimates of PAC. Even within some individuals, the activation foci for human mimic sounds bordered or partially overlapped their functional PAC estimates (Fig. 3; yellow near or within black dotted outlines). Right-hemisphere mimic-sensitive activity in the group-averaged dataset was confined to a small focus along the upper bank of the STS (Fig. 1, yellow). We also calculated a lateralization index (LI) (Jones et al., 2011) with whole brain threshold- and ROI-independent methods (Fig. 1; $\mathrm{LI}_{\mathrm{M>A}} = -2.68$) that strongly supported this robust left-lateralization at the group level.

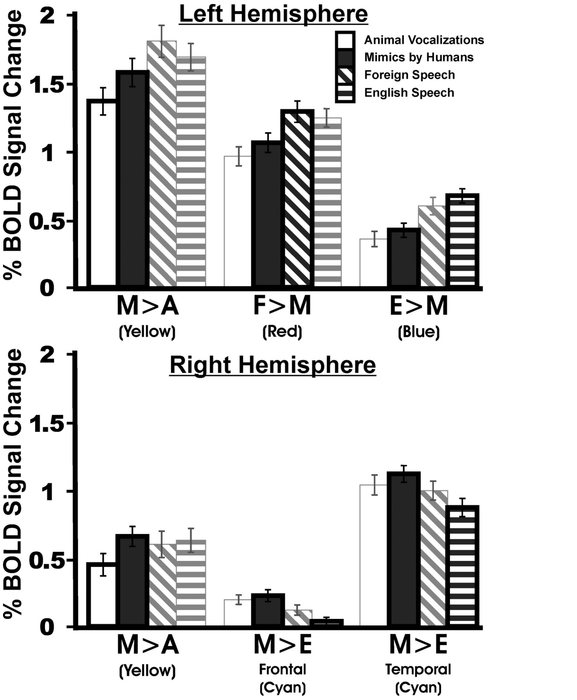

Figure 2. Quantitative representation of BOLD fMRI activation. Mean BOLD signal responses ($n = 22$ subjects) to the four vocalization categories were quantified for each focus or region identified in Figure 1. Data correspond to the means ± SEM. The functional regions identified in Figure 1 are indicated under each four-bar cluster. Left hemisphere regions from Figure 1: M>A (yellow), F>M (red), and E>M (dark blue); right hemisphere regions from Figure 1: M>A (yellow), M>E-Temporal and M>E-Frontal (cyan).

When contrasted with the animal vocalizations, the corresponding human mimic vocalizations were generally well matched for low-level acoustic features such as rhythm, cadence, loudness, and duration. Acoustic and psychophysical attributes were also derived to quantify some of the differences between the mimic-animal vocalizations at sound-pair and categorical levels. One acoustic attribute we measured is related to harmonic content, a signal quality that is significantly represented in vocalizations (Riede et al., 2001; Lewis et al., 2005); this was accomplished by quantifying an HNR value for each stimulus (see Materials and Methods). We previously reported harmonic processing as a distinct intermediate stage in human auditory cortices by showing cortical regions that were parametrically sensitive to the harmonic content of artificial iterated rippled noise stimuli and real-world animal vocalizations (Lewis et al., 2009). In the present study, HNR values for human-mimicked vocalizations were typically greater than their corresponding animal vocalizations; these differences persisted at the categorical level (*t* test, $p < 0.05$) (Table 1).

Two other acoustic attributes we calculated were related to signal entropy measures. Also known as the spectral flatness measure, Weiner entropy quantifies the spectral density in an acoustic signal in the form of resolvable spectral bands (Reddy et al., 2009). Consequently, white noise ("simple" diffuse spectrum) and pure tones (infinite spectral power or density at one frequency) lie at the extreme ends of this attribute's range (white noise: 0, pure tone: $-\infty$). This attribute has been used previously to characterize environmental sounds (Reddy et al., 2009; Lewis et al., 2012). Generally, vocalizations produce the most negative

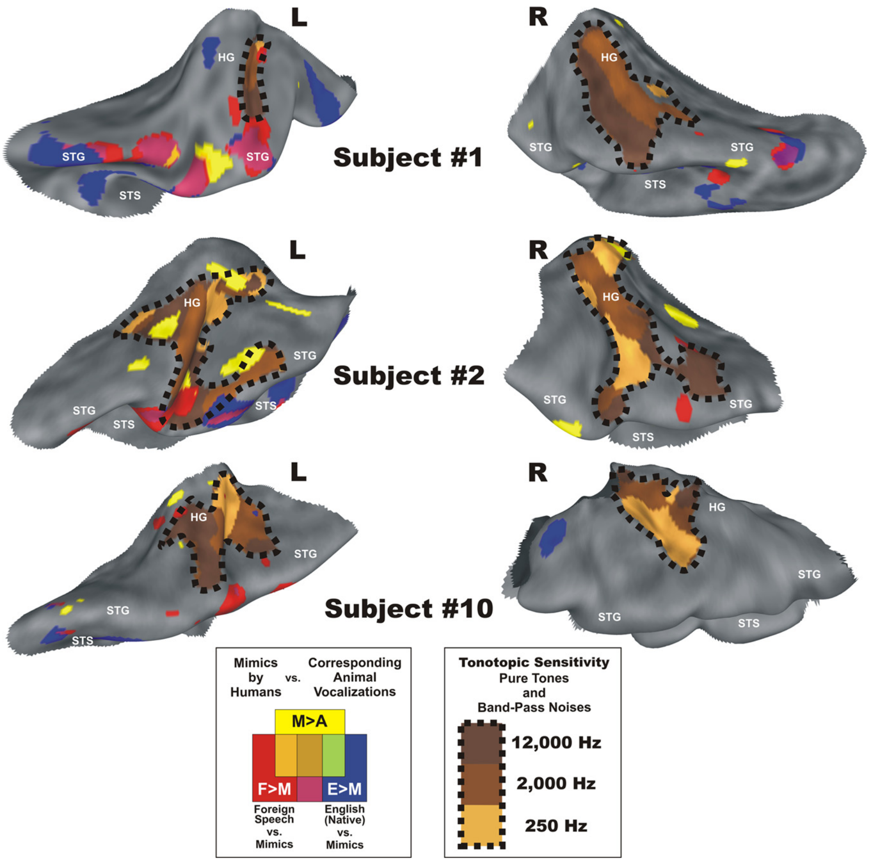
L
R
HG
STG
STS
Subject #1
Subject #2
Subject #10
Mimics by Humans vs. Corresponding Animal Vocalizations
M>A
F>M
E>M
Foreign Speech vs. Mimics
English (Native) vs. Mimics
Tonotopic Sensitivity
Pure Tones and Band-Pass Noises
12,000 Hz
2,000 Hz
250 Hz

Figure 3. Vocalization-sensitive regions near primary auditory cortices in individual participants. Individual cortical maps showing the locations of vocalization-sensitive cortices with respect to tonotopically organized regions (PAC). Tonotopic organization (dotted outlines) occurred primarily along regions within and surrounding HG. Areas activated by the M>A, F>M, and E>M functional contrasts from Figure 1 are highlighted. Cortex that was preferentially sensitive to mimicked versions of animal vocalizations versus corresponding real-world animal vocalizations (M>A, yellow) often occurred near or in some instances overlapped an individual's PAC. More lateral regions along the STG and within the STS often showed preference to neutral foreign (F>M, red) and English phrases (E>M, blue) over non-stereotypical human-mimicked animal vocalizations.

values since they usually contain very specifically structured spectral content (often a fundamental frequency and a few formants). Human-mimicked animal vocalizations from the current study typically had less negative entropy values than their animal vocalization counterparts (Table 1), implying that mimics possessed relatively less ordered acoustic structure. Group-level analysis confirmed the Weiner entropy differences between these two categories (*t* test, $p < 0.0005$).

SSV, derived from Weiner entropy measures, is a measure of the dynamicity of an acoustic signal's spectral distribution over time. Using this measure, white noise and pure tone signal produced similar values near zero, reflecting the stationarity of these signals. Sounds containing dynamic spectral statistics such as vocalizations (especially speech and object-like action sounds) are reported to produce greater SSV values (Reddy et al., 2009; Lewis et al., 2012). Comparing animal vocalizations and corresponding human-mimicked sounds, SSV values were generally lower for human mimics (*t* test, $p < 0.05$). Together, the acoustic signal changes (increases in HNR and Weiner entropy and decrease in SSV) seen between these two categories of sounds were suggestive of the human-mimics having more "simplified" spectrotemporal dynamics (see Discussion).

After scanning sessions, a subset ($n = 18/22$) of our fMRI participants psychophysically rated the animal and human-mimicked stimuli. Each stimulus was rated along a 5-point Likert-scale for the perceived animal-ness (low-score, 1 or 2) or human-ness (high-score, 4 or 5). Ambiguous stimuli that were not perceived as distinctly human- or animal-produced were rated medially along this dimension with a score of three (3). Participants, who were naive to the stimuli during scanning sessions, were relatively proficient at correctly categorizing sounds after being informed of the presence of animal and mimic categories. The numbers of subjects that were able to correctly categorize each stimulus are listed in Table 1 (i.e., animal vocalizations given a 1 or 2 score, human vocalizations given a 4 or 5 score). The accuracy for correctly categorizing both animal vocalizations and human-mimicked versions were comparable across both categories (*t* test, $p = 0.941$; animal vocalizations: 77.95%; human-mimicked vocalizations: 78.56%). An analysis of the fMRI data including BOLD responses only to the correctly categorized stimuli did not produce any qualitative differences from the group-averaged responses to all of the experimental stimuli (data not shown). The relatively low numbers of errors and the fact that the BOLD data and psychophysical data were collected under different conditions (naive and in the scanner vs non-naive and outside of the scanner) also precluded a rigorous "error trials" analysis.

To further identify where these human vocal tract-sensitive regions (M>A) were located in the auditory processing hierarchy, we also compared the responses to mimic stimuli with responses to foreign and English speech samples. Preferential activation to unfamiliar foreign speech, which is incomprehensible with respect to locutionary (semantic) content (Austin, 1975), relative to human-mimicked animal vocalizations (F>M) should reflect the general processing of dynamic spectrotemporal acoustic features typical of spoken languages and utterances at later auditory stages. Cortical regions preferentially responding to foreign speech from six different non-Romance and non-Germanic languages (Akan, Farsi, Hebrew, Hindi, Mandarin, and Yoruban) predominantly radiated posterolaterally out from the left-hemisphere mimic-sensitive focus and into the left STS, as well as medially onto HG (Fig. 1, red).

In addition to basic speech sensitivity, contrasting the BOLD activity between English speech vocalizations and the mimic stimuli (E>M) specifically highlighted cortical sensitivity to the subtle differences in acoustic cues that convey comprehensible locutionary communication in one's native language relative to more fundamental vocal tract sound signals. This condition revealed a strongly left-lateralized expanse of activity (Fig. 1, dark blue) situated further into the STS than foreign-vs-mimic sensitive regions. Responses to the spoken verbal stimuli in our paradigm, whether foreign or native, produced the most strongly left-lateralized networks along the STG and STS (Fig. 1; $LI_{F>M} = -4.47$; $LI_{E>M} = -5.48$). Importantly, our experimental design emphasized conspecific vocal-tract sensitivity and thus did not incorporate any overt phonological, syntactic, or semantic "language tasks" that may have produced more bilateral activation (Hickok and Poeppel, 2007). Collectively, these results form the basis of a left-hemisphere auditory processing hierarchy that is organized by increasingly complex and precise statistical representations of conspecific communication sounds. Directed laterally and anterolaterally along the cortical ribbon (Chevillet et al., 2011), this hierarchy ultimately culminated in the cortical representations of conspecific utterances that express locutionary (semantic) information, similar to models of intelligible speech processing (Scott et al., 2000; Davis and Johnsrude, 2003; Friederici et al., 2010). Individual subjects revealed similar left hemisphere hierarchies that seemed to emerge from around PACs (Fig. 3, yellow to red to blue progressions emanating near HG, extending onto the STG, and into the STS).

Although affective cues are typical of many vocal expressions, we used neutral foreign and English speech samples to avoid cortical activity related to the phatic elements of language. However, the animal vocalizations and their corresponding mimicked versions likely possessed some appreciable amounts of emotional prosodic content. A perceptual screening of our sound stimuli by participants not included in the neuroimaging study did indeed indicate that the mimic sounds were significantly higher in emotional valence content (emotional prosody) than the neutral English and foreign speech stimuli ($n = 6, p < 0.0001$, Kruskal–Wallis tests of 1–5 Likert ratings). Right hemisphere networks are proposed to process affective prosodic cues of vocalization stimuli (e.g., slow pitch-contour modulations) rather than locutionary content (Zatorre and Belin, 2001; Friederici and Alter, 2004; Ethofer et al., 2006; Kotz et al., 2006; Ross and Monnot, 2008; Grossmann et al., 2010). Concordantly, there was a distinct expanse of strongly right-lateralized hemisphere activity that responded preferentially to the mimic stimuli relative to the English speech samples (M>E; Fig. 1, cyan). This included a temporal cluster along the right posterior STG that extended into the planum temporale and onto posterior and lateral aspects of HG near functionally estimated PACs. Additionally, another cluster of foci was revealed within the right inferior frontal cortices along the inferior frontal gyri that extended into the anterior insula. Together, these results strengthen and further specify the purported hemispheric biases for vocal information processing.

Discussion

Lateralized cortical sensitivity to the human vocal tract

The present data revealed a cortical hierarchy in human auditory cortices for processing meaningful conspecific utterances; this hierarchy emerged near left primary auditory cortices in a region that showed species-specific sensitivity to the acoustics of the human vocal tract. We accomplished this by using human-mimicked animal vocalizations—making the animal vocalizations a highly precise control condition because they were generally well matched for numerous low-level acoustic signal attributes. Both of these sound categories lacked familiarity relative to stereotypical human vocalizations and the human-mimic sounds were not within the normal repertoire of frequently encountered or produced human communicative sounds. This minimized any possible effects related to initiating overestablished acoustic schemata (Alain, 2007). As a result, we satisfied our primary aim to create an intermediate non-stereotypical category of complex human vocalizations that was "naturally produced" yet contained little or no human-specific communicative content.

The human-mimicked animal vocalizations used in this study varied qualitatively in their overall imitation "accuracy" when compared with their corresponding animal vocalizations. Attempting to match the pitches of the animal vocalization, the mimickers likely relied more upon the use of their vocal folds (cords). Furthermore, straining the limits of their vocal abilities and the physical limitations of their vocal structures likely emphasized additional nonlinear acoustic elements that are unique to or characteristic of the human vocal tract (Fitch et al., 2002). While a definitive analysis of human-specific vocal acoustics was beyond the objectives of the current study, we nonetheless aimed to acoustically describe our animal vocalization and human-mimic stimuli in part by using three quantitative measures: HNR, Weiner entropy, and SSV. Each of these attributes has been used previously to describe various categories of sound including vocalizations, tool sounds, and other environmental sounds (Riede et al., 2001; Lewis et al., 2009, 2012; Reddy et al., 2009). The increased harmonic content (greater HNR values) seen for human-mimics versus animal vocalizations may reflect a greater reliance on the vocal folds when attempting to match pitches. This effect may further be paralleled by the relative increases in the signal entropy of the mimics. Nonlinear acoustic phenomena emphasized during vocal strain would effectively spread the overall spectral density of human-mimicked vocalizations (Fitch et al., 2002). In addition to the harmonic and spectral entropy changes we observed, human-mimicked versions of animal vocalizations further revealed a decrease in their spectral dynamicity over time (SSV measures). This may reflect limitations not only of human articulatory structures, but also the ability to adapt highly learned vocal routines (e.g., speech) for the purpose of mimicking animals. Overall, the quantitative changes observed within these three acoustic signal attributes (Table 1) are consistent with the notion that the human-mimicked animal vocalizations generally represented acoustically "simplified" versions of their real-world counterparts, simplified in a manner that emphasized acoustic phenomena that are unique to the human vocal tract.

A notable sound category not included in the current experiment was stereotypical non-verbal human vocalizations; this category would include sounds such as humming, coughing, crying, yawning, etc. Our rationale for not including this category reflected both experimental limitations (longer scanning sessions) and theoretical considerations. Scientifically, our primary goal was to describe a cortical network in auditory cortex that reflected increasing representation of locutionary information—information in vocal utterances that reflects their ostensible meaning. While many non-verbal human vocalizations can be produced using prosodic cues that express specific intentions (a questioning "hmm?," coughing conspicuously to gain someone's attention, etc.), the same stimuli can oftentimes be purely reflexive and produced with no overt communicative motivation. We felt that our chosen spectrum of sounds—animal vocalizations, their human-mimicked counterparts, foreign and English speech—represented a straight forward and incremental progression along a dimension of communicative expression that culminated in discernible locutionary content, an utterance mechanism presumably unique to humans (Austin, 1975).

The results of our experiment newly suggest that "voice sensitivity" for humans predominantly emerges along the boundary between the left HG and STG in close proximity to primary auditory cortices. These results, reporting a left-lateralized conspecific vocalization hierarchy near PACs, contrasts with previous studies showing either bilateral voice-sensitivity located more laterally along the STG/STS (Belin et al., 2000; Binder et al., 2000) or right-hemisphere biased effects when using stereotypical verbal or non-verbal vocalizations (Belin et al., 2002). The present findings have significant implications germane to both the evolutionary and developmental trajectory of this cortical function in human and non-human primates, as addressed in the following sections.

The evolution of conspecific voice sensitivity

The evolution of cortical networks that mediate vocal communication and language functions, and more specifically the lateralization of these functions and supporting anatomical structures, is a burgeoning area of research (for review, see Wilson and Petkov, 2011). Specific anatomical differences between primate species point to left-hemisphere biases for the structures that would putatively support the emergence of language functions. For instance, diffusion tensor imaging (DTI) tractography results demonstrate a striking expansion and increasing connectivity of the left arcuate fasciculus between macaques, chimpanzees and humans (Rilling et al., 2008). Additionally, posterior temporal lobe regions such as the planum temporale display asymmetries in gross anatomical structure, similar to those seen in humans (Gannon et al., 1998; Hopkins et al., 1998). Neuroimaging techniques are increasingly being used to describe whole-brain networks for vocalization processing in lower primates (Gil-da-Costa et al., 2006). Functional neuroimaging (fMRI) in Old World monkeys (macaques) has demonstrated bilateral foci showing preference for species-specific vocalizations with a more selective focus occurring along the right anterior superior temporal plane (Petkov et al., 2008). Another macaque study using positron emission tomography (PET) revealed preferential left anterior temporal lobe activity to species-specific vocalizations (Poremba et al., 2004). PET neuroimaging in great apes (chimpanzees) has revealed a right hemisphere preference for certain conspecific vocalizations and utterances (Taglialatela et al., 2009). However, the responses to conspecific vocalizations in chimpanzees were not directly compared with those produced by other species thereby precluding interpretations regarding species-specificity per se. While the auditory pathways in lower and higher non-human primates that process vocalization sounds require further study, findings hitherto support the presence of at least some lateralized functions in networks for processing conspecific vocalizations.

With regard to vocal communication networks in primates, the present data suggest that early left hemisphere auditory networks in humans are hierarchically organized to efficiently extract locutionary (semantic) content from conspecific speech utterances. By contrasting neuronal activity to a given species' (e.g., chimpanzee) conspecific utterances versus human-mimicked versions, subtle differences may be revealed along various intermediate cortical processing stages. We propose that a hierarchy of "proto-network" homologues similar to the one we have described may be revealed in other primates, especially the great apes, by using a similar experimental rationale. This may further our understanding of the evolutionary underpinnings of vocal communication processing.

Vocalization processing in a neurodevelopmental context

The present findings also have significant implications for language development in children. Early stages in human auditory processing pathways may be, or develop through experience to become, optimized to process the statistically representative qualities unique to the human vocal tract. This arrangement would promote maximal extraction of conspecific communicative content from complex auditory scenes (i.e., socially relevant vocal communication from other humans). Seminal steps of this process would likely involve encoding the fundamental acoustic signatures of personally significant vocal tracts, initially including the voices of one's caretakers', one's own voice and, for social animals, the voices of other conspecifics. For example, human infants generally produce more positive and preferential responses to "motherese" and other baby-directed vocalizations (Cooper and Aslin, 1990; Mastropieri and Turkewitz, 1999). Those responses may be driven heavily by the relatively stable statistical structure of basic simplified vocalizations (Fernald, 1989), notably vowels and other utterances possessing relatively simple amplitude envelopes, elevated pitch and strong harmonic content, the latter being a hallmark acoustic attribute of vocal communication sounds across most species (Riede et al., 2001; Lewis et al., 2005, 2009). While it remains unclear whether sensitivity to intrinsic human vocal tract sounds reflects domain-specific functions (nature) or auditory experience (nurture), the auditory experiences that initiate or influence these functions may begin *in utero* (Decasper et al., 1994) while a fetus experiences harmonically structured vocal sounds. A longer developmental timeframe would ostensibly follow this initial sensitivity, during which an emergent sensitivity to more subtle and complex socially relevant acoustic signal cues appears as more advanced communicative and language abilities develop (Wang, 2000).

Near-infrared spectroscopy has been implemented to demonstrate the emergence of voice-sensitivity in infant auditory cortices between 4 and 7 months of age, showing a right hemisphere bias when processing emotional prosody (Grossmann et al., 2010). Recently, an fMRI study involving infant participants ranging in age from 3 to 7 months revealed regions along the right anterior temporal cortices that were preferentially activated by stereotypical non-speech human vocalizations versus common environmental sounds (Blasi et al., 2011). In conjunction with the results from infant studies, our findings lead us to posit that the right hemisphere possesses greater vocalization sensitivity during early development due to its propensity for processing acoustically "simpler" prosodic cues (when compared with complexly adjoined speech sound). Left hemisphere structures subsequently follow, reflecting a combination of cortical development constraints (Leroy et al., 2011) and the behavioral need to perform the more rapid spectrotemporal analyses (Zatorre and Belin, 2001; Obleser et al., 2008) required to extract more specific communicative information from locutionary vocalizations and other communicative utterances (Austin, 1975). This developmental paradigm may also reflect the increasing cortical influences by social and attention-related cortical networks (Kuhl, 2007, 2010). Regardless, we believe that testing immature auditory systems using the current experiment's rationale will help clarify the typical developmental trajectory of auditory circuits that become optimized for extracting conspecific communication content. This will help provide insight into the etiology of various language and social-affective communication disorders that may manifest during early stages of a child's language development including specific language impairments and autism (Gervais et al., 2004; Shafer and Sussman, 2011).

Notes

Supplemental sample sound files for this article are available at http://www.hsc.wvu.edu/wvucn/Faculty-Lab-Personnel/Lewis/Recent-Publications departmental web page. This material has not been peer reviewed.

References

Abrams DA, Nicol T, Zecker S, Kraus N (2009) Abnormal cortical processing of the syllable rate of speech in poor readers. J Neurosci 29:7686–7693.

Aeschlimann M, Knebel JF, Murray MM, Clarke S (2008) Emotional pre-eminence of human vocalizations. Brain Topogr 20:239–248.

Alain C (2007) Breaking the wave: effects of attention and learning on concurrent sound perception. Hear Res 229:225–236.

Argall BD, Saad ZS, Beauchamp MS (2006) Simplified intersubject averaging on the cortical surface using SUMA. Hum Brain Mapp 27:14–27.

Austin JL (1975) How to do things with words, Ed 2. Cambridge, MA: Harvard UP.

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

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Cogn Brain Res 13:17–26.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Blasi A, Mercure E, Lloyd-Fox S, Thomson A, Brammer M, Sauter D, Deeley Q, Barker GJ, Renvall V, Deoni S, Gasston D, Williams SC, Johnson MH, Simmons A, Murphy DG (2011) Early specialization for voice and emotion processing in the infant brain. Curr Biol 21:1220–1224.

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 15:97–110.

Chevillet M, Riesenhuber M, Rauschecker JP (2011) Functional correlates of the anterolateral processing hierarchy in human auditory cortex. J Neurosci 31:9345–9352.

Chung MK, Robbins SM, Dalton KM, Davidson RJ, Alexander AL, Evans AC (2005) Cortical thickness analysis in autism with heat kernel smoothing. Neuroimage 25:1256–1265.

Cooper RP, Aslin RN (1990) Preference for infant-directed speech in the first month after birth. Child Dev 61:1584–1595.

Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Decasper AJ, Lecanuet JP, Busnel MC, Granierdeferre C, Maugeais R (1994) Fetal reactions to recurrent maternal speech. Infant Behav Dev 17:159–164.

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

Ethofer T, Anders S, Erb M, Herbert C, Wiethoff S, Kissler J, Grodd W, Wildgruber D (2006) Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. Neuroimage 30:580–587.

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.

Fernald A (1989) Intonation and communicative intent in mothers' speech to infants: is the melody the message? Child Dev 60:1497–1510.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Fitch WT, Neubauer J, Herzel H (2002) Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. Anim Behav 63:407–418.

Friederici AD, Alter K (2004) Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang 89:267–276.

Friederici AD, Kotz SA, Scott SK, Obleser J (2010) Disentangling syntax and intelligibility in auditory language comprehension. Hum Brain Mapp 31:448–457.

Gannon PJ, Holloway RL, Broadfield DC, Braun AR (1998) Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. Science 279:220–222.

Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthélémy C, Brunelle F, Samson Y, Zilbovicius M (2004) Abnormal cortical voice processing in autism. Nat Neurosci 7:801–802.

Gil-da-Costa R, Martin A, Lopes MA, Muñoz M, Fritz JB, Braun AR (2006) Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. Nat Neurosci 9:1064–1070.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515–522.

Grossmann T, Oberecker R, Koch SP, Friederici AD (2010) The developmental origins of voice processing in the human brain. Neuron 65:852–858.

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.

Hauber ME, Cassey P, Woolley SM, Theunissen FE (2007) Neurophysiological response selectivity for conspecific songs over synthetic sounds in the auditory forebrain of non-singing female songbirds. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 193:765–774.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hill J, Dierker D, Neil J, Inder T, Knutsen A, Harwell J, Coalson T, Van Essen D (2010) A surface-based analysis of hemispheric asymmetries and folding of cerebral cortex in term-born human infants. J Neurosci 30:2268–2276.

Hopkins WD, Marino L, Rilling JK, MacGregor LA (1998) Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). Neuroreport 9:2913–2918.

Jones SE, Mahmoud SY, Phillips MD (2011) A practical clinical method to quantify language lateralization in fMRI using whole-brain analysis. Neuroimage 54:2937–2949.

Kotz SA, Meyer M, Paulmann S (2006) Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. Prog Brain Res 156:285–294.

Kuhl PK (2007) Is speech learning 'gated' by the social brain? Dev Sci 10:110–120.

Kuhl PK (2010) Brain mechanisms in early language acquisition. Neuron 67:713–727.

Lass NJ, Eastham SK, Wright TL, Hinzman AR, Mills KJ, Hefferin AL (1983) Listeners' identification of human-imitated animal sounds. Percept Mot Skills 57:995–998.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial non-speech sounds recruits speech-sensitive cortical regions. J Neurosci 29:5234–5239.

Leroy F, Glasel H, Dubois J, Hertz-Pannier L, Thirion B, Mangin JF, Dehaene-Lambertz G (2011) Early maturation of the linguistic dorsal pathway in human infants. J Neurosci 31:1500–1506.

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. Cereb Cortex 14:1008–1021.

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, Brefczynski-Lewis JA (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, Tallaksen KC, Frum CA (2012) Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. Front Syst Neurosci 6:1–15. doi:10.3389/fnsys.2012.00027.

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20:2958–2970.

Mastropieri D, Turkewitz G (1999) Prenatal experience and neonatal responsiveness to vocal expressions of emotion. Dev Psychobiol 35:204–214.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13:684–701.

Nunnally JC (1978) Psychometric theory. New York: McGraw-Hill.

Obleser J, Eisner F, Kotz SA (2008) Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci 28:8116–8123.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11:367–374.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427:448–451.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669–683.

Reddy RK, Ramachandra V, Kumar N, Singh NC (2009) Categorization of environmental sounds. Biol Cybern 100:299–306.

Riede T, Herzel H, Hammerschmidt K, Brunnberg L, Tembrock G (2001) The harmonic-to-noise ratio applied to dog barks. J Acoust Soc Am 110:2191–2197.

Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TE (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. Nat Neurosci 11:426–428.

Robins DL, Hunyadi E, Schultz RT (2009) Superior temporal activation in response to dynamic audio-visual emotional cues. Brain Cogn 69:269–278.

Ross ED, Monnot M (2008) Neurology of affective prosody and its functional-anatomic organization in right hemisphere. Brain Lang 104:51–74.

Saad ZS, Chen G, Reynolds RC, Christidis PP, Hammett KR, Bellgowan PSF, Cox RW (2006) Functional Imaging Analysis Contest (FIAC) analysis according to AFNI and SUMA. Hum Brain Mapp 27:417–424.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shafer VL, Sussman E (2011) Predicting the future: ERP markers of language risk in infancy. Clin Neurophysiol 122:213–214.

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

Taglialatela JP, Russell JL, Schaeffer JA, Hopkins WD (2009) Visualizing vocal perception in the chimpanzee brain. Cereb Cortex 19: 1151–1157.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc 8:443–459.

Wang X (2000) On cortical coding of vocal communication sounds in primates. Proc Natl Acad Sci U S A 97:11843–11849.

Wang X, Kadia SC (2001) Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. J Neurophysiol 86:2616–2620.

Wilson B, Petkov CI (2011) Communication and the primate brain: insights from neuroimaging studies in humans, chimpanzees and macaques. Hum Biol 83:175–189.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.