

8-1-2018

Hearing and orally mimicking different acoustic-semantic categories of natural sound engage distinct left hemisphere cortical regions

James W. Lewis
West Virginia University

Magenta J. Silberman
West Virginia University

Jeremy J. Donai
West Virginia University

Chris A. Frum
West Virginia University

Julie A. Brefczynski-Lewis
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/ctsi>



Part of the [Medicine and Health Sciences Commons](#)

Digital Commons Citation

Lewis, James W.; Silberman, Magenta J.; Donai, Jeremy J.; Frum, Chris A.; and Brefczynski-Lewis, Julie A., "Hearing and orally mimicking different acoustic-semantic categories of natural sound engage distinct left hemisphere cortical regions" (2018). *Clinical and Translational Science Institute*. 980.
<https://researchrepository.wvu.edu/ctsi/980>

This Article is brought to you for free and open access by the Centers at The Research Repository @ WVU. It has been accepted for inclusion in Clinical and Translational Science Institute by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.



Published in final edited form as:

Brain Lang. 2018 August ; 183: 64–78. doi:10.1016/j.bandl.2018.05.002.

Hearing and orally mimicking different acoustic-semantic categories of natural sound engage distinct left hemisphere cortical regions

James W. Lewis^{a,*}, Magenta J. Silberman^a, Jeremy J. Donai^b, Chris A. Frum^a, and Julie A. Brefczynski-Lewis^a

^aDepartment of Physiology, Rockefeller Neurosciences Institute, Pharmacology & Neuroscience, West Virginia University, Morgantown, WV 26506, USA

^bDepartment of Communication Sciences and Disorders, Rockefeller Neurosciences Institute, West Virginia University, Morgantown, WV 26506, USA

Abstract

Oral mimicry is thought to represent an essential process for the neurodevelopment of spoken language systems in infants, the evolution of language in hominins, and a process that could possibly aid recovery in stroke patients. Using functional magnetic resonance imaging (fMRI), we previously reported a divergence of auditory cortical pathways mediating perception of specific categories of natural sounds. However, it remained unclear if or how this fundamental sensory organization by the brain might relate to motor output, such as sound mimicry. Here, using fMRI, we revealed a dissociation of activated brain regions preferential for hearing with the intent to imitate and the oral mimicry of animal action sounds versus animal vocalizations as distinct acoustic-semantic categories. This functional dissociation may reflect components of a rudimentary cortical architecture that links systems for processing acoustic-semantic universals of natural sound with motor-related systems mediating oral mimicry at a category level. The observation of different brain regions involved in different aspects of oral mimicry may inform targeted therapies for rehabilitation of functional abilities after stroke.

Keywords

Acoustic communication; fMRI; Echo-mirror neuron system; Sound symbolism; Language evolution; Acoustic-semantic categories; Categorical perception; Stroke rehabilitation

1. Introduction

Anthropological theories of glottogenesis (evolution of spoken language systems) propose a ‘default’ mouth-gesture hypotheses behind transitions from episodic to mimetic cultures in hominins (Condillac, 1746 (1947); Darwin, 1871/1981; Donald, 1991; Johannesson, 1950;

*Corresponding author. Rockefeller Neurosciences Institute, Department of Physiology, Pharmacology & Neuroscience, P.O. Box 9229, West Virginia University, Morgantown, WV 26506, USA. jwlewis@hsc.wvu.edu (J.W. Lewis).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016Xj.bandl.2018.05.002>.

Paget, 1944, 1963; Sterelny, 2012; Tylor, 1868; Wallace, 1895). Anthropological and cognitive psychology theories converge on the view that conceptual systems associated with oral mimicry of events of the observed world are likely to have played a central role in both the ontogenesis and phylogenesis of communication and language abilities (Arbib, 2005; Hewes, 1973; Imai, Kita, Nagumo, & Okada, 2008; Jackendoff, 2003; Ramachandran and Hubbard, 2001). As such, oral communication form should show a resemblance to properties of sensory-motor and affective referents, as addressed in theories of sound symbolism (Asano et al., 2015; Imai and Kita, 2014; Kanero, Imai, Okuda, Okada, & Matsuda, 2014; Sapir, 1929; Taylor and Taylor, 1962; Weiss, 1964) and iconicity (Perniss and Vigliocco, 2014). Recent theories further posit that the ability to both perceive and orally mimic events depicting incidental sounds of locomotion and tool-use (action sounds), as well as mimicry of animal calls (vocalizations), were likely to have represented some of the most rudimentary semantic categories of natural sound that contributed to the early stages of hominin oral communication (Falk, 2004; Larsson, 2014, 2015). Moreover, in stroke recovery models, observation therapies (observation with intent to imitate or mirror) can facilitate the voluntary production of movement: However, there remains a need for advances in neuroscientific frameworks of goal-directed motor production and communication to enable rigorous testing of rehabilitation hypotheses (Garrison, Aziz-Zadeh, Wong, Liew, & Winstein, 2013; Garrison, Winstein, & Aziz-Zadeh, 2010; Pomeroy et al., 2005). In sum, it remains unclear if, or the extent to which, neuronal systems mediating oral mimicry might be rooted in networks associated with sensory systems (e.g. auditory perception), reflecting potential vestiges of earlier modes of communication at a semantic category level.

From the perspective of hearing perception, we recently developed a neurobiological model for the processing of different acoustic-semantic categories of real-world natural sounds that may apply to all social mammals with hearing ability (Brefczynski-Lewis and Lewis, 2017): This model (Fig. 1) was based largely on neuroimaging results from both human adults (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Clarke, Bellmann, de Ribaupierre, & Assal, 1996; Engel, Frum, Puce, Walker, & Lewis, 2009; Engelien et al., 2006; Lewis et al., 2009; Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005; Lewis, Talkington, Puce, Engel, & Frum, 2011; Lewis, Talkington, Tallaksen, & Frum, 2012; Webster et al., 2017) and infants (Geangu, Quadrelli, Lewis, Macchi Cassia, & Turati, 2015). This included three basic categories of soundsource: (1) action sounds (non-vocalizations) produced by ‘living things’, with human (conspecific) and non-human animal sources representing two subcategories; (2) action sounds produced by ‘nonliving things’, including environmental sounds and human-made machinery; and (3) vocalizations (‘living things’), with human versus nonhuman animals as two subcategories therein. This model was supported in a study that utilized non-human animal action sounds and vocalizations (also used in the present study), which minimized potential confounds related to the processing of deeper semantic encodings in meaning conveyed by commonly experienced (“over-learned”) human conspecific sounds (Webster et al., 2017). The goal of the present study was to determine if this same basic organizational principle, namely the processing along separable cortical pathways, might also be respected in some of the cortical regions involved in planning and orchestrating oral mimicry of these same sounds at a categorical level.

Our first (null) hypothesis was that brain regions involved in oral mimicry of the two distinct categories would show no major network-level differences in activation (given the resolution of fMRI), apart from potential differences in primary motor cortices involved directly in laryngeal versus oral-facial muscle control or from issues of mimicry difficulty. Our second (main, alternative) hypothesis was that, similar to the double-dissociation of brain networks we recently revealed mediating activation preferential for perception of one versus another acoustic-semantic category of natural sounds, there would also be a dissociation of brain regions showing specificity for oral mimicry for one versus another category of natural event type. Evidence in support of this latter hypothesis would potentially identify gross-level network mechanisms for how different types or aspects of semantic information representations are routed. This could reflect different form-meaning mappings or working memory operations between hearing acoustic-semantic universals characteristic of different categories of natural sound events and oral mimicry of those same categories of events. Identifying such mechanisms would have significant implications for future studies designed to further understand models of oral communication acquisition and production, both in terms of neurodevelopment of mimicry during infancy and potentially in rehabilitative strategies to facilitate recovery from aphasia after stroke or traumatic brain injury.

2. Methods

2.1. Participants

We recruited 16 English speaking participants (19–26 years of age, 9 female, 15 right-handed, and 15 native English speakers). Participants had no previous history of major neurological or psychiatric disorders, and a self-reported normal range of hearing with no auditory or vocal production impairments. Informed consent was obtained for all participants following guidelines approved by the West Virginia University Institutional Review Board.

2.2. Sound stimuli

The sound stimulus set consisted of 20 animal vocalizations and 20 animal action sounds, which were a subset of those used in our earlier study of categorical sound processing (Webster et al., 2017). Rationale for selecting these sound stimuli, which were professional recordings of isolated animals (Sound Ideas, Inc., Richmond Hill, Ontario, Canada), was both the perceived ability for them to be mimicked orally and the clarity of their semantic category membership (clearly non-human animal sounds, and being a vocalization or non-vocal action sound, as detailed in our earlier study). After fMRI participant interviews from the present study, two sound stimuli per category were deemed as being perceived as slightly ambiguous as to category, and were subsequently censored from all analyses, retaining 18 stimuli in each category (Table 1) for use in the data analyses described below. The two categories of sound stimuli were matched for duration (2.7 ± 0.2 s) and total root mean squared power (-17.6 ± 0.5 dB), converted to one channel (mono, 44.1kHz, 16-bit; Adobe Audition 3.0, Adobe Systems Inc.) but presented to both ears, thereby removing any binaural spatial cues. Because emphasis of the study was based on sound categories representing ethologically valid events, they necessarily differed in several acoustic signal attributes, including those summarized in Table 1. The animal vocalizations were psychophysically

assessed as having a relatively negative emotional valence overall ($n = 15$, 8 female; Likert scale: $-2 =$ very negative, $0 =$ neutral, $+2 =$ very positive: $\text{Avg} \pm \text{SD} = -0.84 \pm 0.72$), while the ratings of the animal action sounds were more neutral ($+0.39 \pm 0.35$), which were ratings that significantly differed from one another (single factor ANOVA, $F_{1,34} = 42.7$, $p < 10^{-7}$).

2.3. Scanning paradigms

2.3.1. Scanning preparation—All participants practiced the oral mimicry task under two or three listening conditions. This first included practicing the mimicry of all sounds while seated in a sound isolation booth (Model 800A-RF shielded, Industrial Acoustics Co., North Aurora, IL, USA) and repeating practice with difficult sounds as needed. A second practice session involved lying down inside the bore of a simulation MRI scanner (Model PST-100444; Psychology Software tools, Inc., Sharpsburg, PA, USA) with a microphone apparatus positioned near their mouth until participants were comfortable with performing mimicry of all sound stimuli. This practice regimen cycle was repeated as necessary (1 or 2 sessions, and repeating individual sounds) to minimize activation of networks that might simply be associated with sound novelty, attentional demands, laughter, or potential motoric mimicry difficulty across the two sound categories, especially for the action sounds.

2.3.2. Sound mimicry paradigm—The sound mimicry paradigm consisted of 4 separate runs, in which the 40 sound stimuli plus 24 silent events were presented in pseudorandom order (64 TRs per run), avoiding three or more silent events/periods in a row (e.g. see Fig. 2). Events that immediately followed a sound presentation were silent, indicating times when the participant would orally mimic the sound they had just heard. The high fidelity sound stimuli were delivered using a Windows PC computer, with Presentation software (version 11.1, Neurobehavioral Systems Inc.) through a sound mixer and MR compatible ear buds (Model S14, Sensimetrics Corp., Malden MA). Stimulus loudness was set to a comfortable level for each participant (initially set at 82 ± 5 dB Leq to a 1 kHz pure tone; fast A-weighted; Brüel & Kjær 2239a sound meter), as assessed immediately after each scanning session. Orally mimicked sounds were recorded during the scanning sessions using an MR compatible recording system (FOMRI-III, OptoAcoustics; Or-Yehuda, Israel) and related recording software (OptiMRI 3.1). Each individual's mimicry recordings were extracted, and ON/OFF ramped at 25 ms in the same manner as with the animal sound stimuli, and then analyzed for a direct comparison of acoustic signal attributes with the original animal sounds (Table 1). Shortly after the scanning session, most ($n = 14$ of 16) participants provided psychophysical ratings assessing how well they personally thought they had mimicked each of the animal sounds using a Likert scale (1 = very poor, 5 = very well), which were ratings subsequently used to test for brain regions showing parametric sensitivity to perceived mimic quality. Ratings of how difficult they thought the oral mimicry would be was also assessed using a 1–5 Likert scale.

2.3.3. Magnetic resonance imaging and data analysis—The imaging was conducted on a 3 Tesla Siemens Verio MRI scanner using an 8-channel head coil. We acquired whole-head brain volumes using a cardiac-gated, event related design (35 axial slices at 4 mm slice thickness, no gap, descending sequence, 3.75×3.75 mm² in-plane

resolution). Blood oxygen-level dependent (BOLD) signals were collected using a clustered acquisition echo planar pulse sequence (ep2d: TR = 10,000 ms, TE = 30 ms, FOV = 240 mm, 75 degree flip angle, 2.165 s slice packet). The stimulus computer was triggered by each TR, thereby ensuring accurate time stamps. The time between sound onset and fMRI scanning acquisition was 7.5 s plus time until the peak of the participant's next cardiac cycle (R-wave), thereby introducing a temporal jitter in the interstimulus intervals. Whole brain T1-weighted anatomical MR images were collected using a standard MPRAGE pulse sequence ($0.9 \times 0.6 \times 1.5 \text{ mm}^3$ resolution, TI = 900 ms).

Functional datasets were processed using Analysis of Functional NeuroImages (AFNI; <http://afni.nimh.nih.gov/>) and associated software packages (Cox, 1996). Functional scans were globally corrected for minor motion artifacts due to head translations and rotations (software 3dVolreg), and subjected to a 4 mm Gaussian spatial blur (Mikl et al., 2008). Because this was a clustered acquisition design, the effects of differences in head and facial movements were minimized since brain images were collected after mimicry movements had occurred (see Fig. 2). BOLD signals were converted to percent signal change on a voxel-wise basis relative to the averaged BOLD signals attained from the silent events for each scanning run for each participant.

Group-level analyses involved manual volumetric alignment to standardized Talairach space (Talairach and Tournoux, 1988). Scanning runs were concatenated, corrected for baseline linear drifts, and multiple linear regression analyses were performed. Cross-conditions identified voxels showing differential BOLD signal “activation” to the action sounds versus vocalizations, and differential activation to oral mimicry of the action sounds versus vocalizations. For group averaging, the multiple regression coefficients were spatially low-pass filtered (6 mm box filter). Across the 16 participants, the fMRI data were subjected to a repeated-measures, mixed effects two factor ANOVA (3dANOVA3 software; type 4), comparing responses to each category of sound (Category: Action sounds, Vocalizations) versus task condition (Condition: Hear sound, Orally mimic sound), both relative to activation during the separate silent event trials. The data were also subjected to *t*-tests to reveal brain region subdivisions that showed positive BOLD signals (relative to silent events) to both task conditions for a given event category of interest. With consideration to multiple comparisons (Eklund, Nichols, & Knutsson, 2016; Forman et al., 1995), an analysis of the functional noise in the BOLD signal across voxels was used, estimating the full-width half-max Gaussian filter widths at $x = y = z = 5.8 \text{ mm}$ spatial smoothness (using software packages 3dDeconvolve and 3dClustSim, version 16.2.06). Applying a minimum cluster size of 411 mm^3 voxels together with $p_{\text{uncorrected}} < 0.001$ individual voxel-wise ANOVA yielded a whole-brain correction for multiple comparisons at $p_{\text{corr}} < 0.01$. A $p_{\text{uncorrected}} < 0.01$ voxel-wise *t*-test with 1007 mm^3 voxel minimum cluster size yielded a whole-brain correction at $p_{\text{corr}} < 0.01$. For visualization purposes, data were projected onto PALS atlas cortical surface models using Caret software (<http://brainmap.wustl.edu>) (Van Essen, 2005).

To reveal regions showing parametric sensitivity to each individual's mimic quality rating, one analysis entailed modeling each individual's fMRI data using their Likert ratings as regressors as the first level analysis, and processing the data in a manner similar to that described above to regions parametrically sensitive to how well they thought they mimicked

the sounds and to how difficult they thought the sounds were to mimic. A second analysis entailed using regions of interest (ROIs, from Fig. 4) already identified at a group-level and then transforming those coordinates from Talairach coordinate space back to the participants' original brain space, and extracting each individual's time series data from the corresponding voxels (3dfractionize software). This second analysis used category specificity as the first level analysis and parametric sensitivity as a second level, thereby incurring some statistical circularity: thus those results serve more to further characterize the brain region response profiles. Binned Likert ratings were charted against percent signal changes of BOLD signals using a voxel-wise whole-brain analysis and plotted by category.

2.3.4. Auditory cortex localizer paradigm—Nine of the participants additionally performed a paradigm to identify primary auditory cortices (two ~ 8 min fMRI scans), including functionally defined core, belt and parabelt cortices (Chevillet, Riesenhuber, & Rauschecker, 2011): With eyes closed they listened to pure tones (PT), band pass noises (BPN), English phonemes, and silent events. The PT and BPN stimuli had center frequencies of 500, 2000, 8000 Hz, and BPN stimuli had a bandwidth of 1 octave. Spoken phonemes (e.g. “ga, ga, ga, ga, ga”) were derived from ten different individuals (5 female), expressed using a neutral tone of voice from a database of recordings (Dr. Mark Chevillet, personal communication). Each scanning run consisted of 15 events of each sound category counter-balanced and randomly presented together with 15 silent events as a baseline control. Whole-brain volumes of BOLD signal were also collected using a clustered acquisition design, but with the TR set so that the time between stimulus onset and imaging onset was 4.5 s, as opposed to 7.5 s for the sound mimicry paradigm, to reveal early auditory processing stages (Baumann et al., 2010). As described by Chevillet et al. (2011), datasets were subjected to an ANOVA and *t-tests* to identify group-averaged regions showing significant activation to PT > silence thresholded until reaching a roughly 1800 mm³ set of bilateral foci to identify auditory “core” regions. Similar threshold settings revealed BPN > PT, which operationally defined auditory “belt” regions (at $p < 0.001$, uncorrected), and also Phonemes > BPN and PT ($p < 0.001$, uncorrected), which defined auditory “parabelt” regions.

3. Results

3.1. Oral mimicry systems in the human brain

Using an event-related, clustered acquisition fMRI paradigm we imaged participants' brains immediately after they listened to recordings of animal action sounds, animal vocalizations, and silent events, and imaged them immediately after they subsequently orally mimicked the corresponding sounds (see Methods and Fig. 2), which were well practiced prior to scanning procedures. The mimicked sounds were recorded using an MR-compatible microphone during fMRI scanning to verify task compliance and to assess potential oral-facial and vocal mechanisms each participant was using when mimicking the sounds. After scanning procedures, the participants' rated (using 1–5 Likert scales) both their perceived quality of mimicry attempts and their estimated difficulty level for mimicking each sound.

Category-specific effects were revealed for both the hearing task (perception) condition (Fig. 3B) and separately the oral mimicry (motor output) task condition (Fig. 3C). We previously identified brain regions that were preferential for processing and categorizing animal action sounds (Fig. 3A, yellow) versus animal vocalizations (red), using a continuous acquisition fMRI paradigm (Webster et al., 2017). This earlier paradigm was designed to detect cortical processing by collecting and modeling BOLD signal changes continuously over time, and thus was optimized to reveal earlier hemodynamic BOLD signals of and surrounding primary auditory cortices (i.e. Fig. 3A, PAC core plus belt estimates outlines): The present study involved use of a subset of the same sound stimuli, but used fMRI scanning and timing parameters that were designed for clustered acquisitions (Fig. 2), which were optimized to (1) allow sounds to be clearly heard during scanner silence, and (2) allow for audio recording of participant's mimicry performance during scanner silence. Thus, whole-brain imaging occurred after the estimated optimal BOLD signal peaks to sound onset for primary auditory cortices (e.g. 4–5 s delay, versus 7.5 s delay for later processing stages), consistent with a non-human primate fMRI study (Baumann et al., 2010). The mapping of primary auditory cortices was thus conducted as a separate PAC localizer scan in a subset of the participants. Brain slice acquisitions (BOLD signals) for the mimicry paradigm were judiciously timed to capture the relatively later hemodynamic peaks more closely associated with higher stages of hearing perception processing and cognition (Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005; Lewis, Wightman, et al., 2004) and to motor production at relatively later stages of processing in cortex, with imaging timing initiating at roughly 7.5 s after sound onset (see Methods). The task in the present study was to listen to the sound and prepare to mimic the sound after the next scanner sound cycle (~10s later), rather than overtly categorize the sound using a two-alternative forced choice response. Despite differences in both task and in fMRI timing parameters, the present study data similarly revealed activation (cf. Fig. 3A and B) along the left and right lateral STG regions (red) that was preferential for animal vocalizations. The bilateral posterior insulae preferential for action sounds (Fig. 3A, yellow) only showed preferential activation at reduced threshold settings in the present paradigm (not shown). However, similar activation patterns between the two studies for animal action sounds (yellow) was readily evident in the left-lateralized frontal and parietal regions, including the ventro-lateral paracentral (vIPC) lobule, with a focus overlapping the estimated dorsal laryngeal motor cortex (dLMC) as expected for an oral-motor task (Brown, Ngan, & Liotti, 2008; Penfield and Boldrey, 1937; Simonyan and Horwitz, 2011; Simonyan, 2014), the left anterior inferior frontal sulcus (aIFS), bilateral ventral LMC (vLMC), left inferior postcentral gyrus (overlapping primary somatosensory cortex; "S1"), plus the left posterior middle temporal gyrus (pMTG).

Combined, the two oral mimicry category conditions collectively led to substantially greater activation in ventral portions of the posterior frontal cortices (Fig. 3C, orange), most notably overlapping ventrolateral prefrontal regions bilaterally. This included activation of the inferior frontal gyrus along *pars opercularis* and *pars triangularis*, which in the left hemisphere is an estimated location of "Broca's" area (also see Fig. 3C, histogram). The estimated location of Broca's area (ventral inferior frontal gyrus from the PALS atlas database) showed activation levels comparable in magnitude across all conditions (Fig. 3C, histogram), illustrating significant activation relative to baseline silent events, yet showing

no preferential activity at a category level. Differential activation evoked by oral mimicry of the animal action sounds versus animal vocalizations revealed several foci. For mimicry of animal vocalizations, this included the left aSTG region (Fig. 3C, red), which was not present in the right hemisphere, even at lowered significance threshold settings. Vocalization mimicry also preferentially activated the left anterior insula and right hemisphere posterior superior temporal sulcus. Mimicry of animal action sounds (Fig. 3C, yellow) predominantly included bilateral dorsal networks, the vIPC lobule, bilateral inferior postcentral cortices that at least roughly overlapped with somatosensory S1 face and laryngeal representations (Penfield and Boldrey, 1937), bilateral dorsomedial precentral cortices, and a left lateralization bias for activation of inferior frontal regions including the left aIFS. Thus, in addition to the dissociation of cortical pathways for processing animal action sounds versus animal vocalizations, these results similarly identified differences in network recruitment for oral mimicry of those natural acoustic-semantic categories of sound.

To reveal brain regions showing sensitivity to both hearing and mimicking the respective categories of sound, we performed both an ANOVA and a series of *t-test* conjunctions with the four conditions: mixed effects model with the type of task condition (hearing, oral mimicry) and type of stimulus (action sound, vocalization) as within-subject factors, testing for regions that were preferential for hearing either category of sound, and preferential for orally mimicking that particular category of sound. The ANOVA revealed a significant effect across conditions by category, with either positive or negative activation relative to silent events (Fig. 4A–B, pale yellow and pink hues). The use of clustered acquisition and relatively late timing likely resulted in a baseline shift in sampling of the timing of peak BOLD signal (i.e. leading to differential magnitudes closer to baseline levels). Thus a more restrictive analysis was performed by using a conjunction of pair-wise *t-tests* further constrained to only reveal activation foci that exhibited *positive* BOLD signal in responses to sound processing and to oral mimicry for a given category relative to the baseline control silent events (Fig. 4A–B, saturated yellow and saturated red hues; Table 2). The response characteristics of several regions of interest (ROI) from the overlap condition, such as the left aSTG (Fig. 4A–B, histograms), showed significantly greater BOLD responses when hearing animal action sounds (pale yellow outlined bars) versus hearing animal vocalizations (white bars with red outlines) and when orally mimicking those corresponding animal action sounds (yellow filled bars) versus the animal vocalizations (red filled bars), respectively.

Adapting a previously published auditory cortex localizer scan (Chevillet et al., 2011), functionally-defined locations of primary auditory cortices were also mapped in a subset of our participants (Fig. 4, blue hues). This localizer paradigm revealed an estimated location of (1) the auditory core (not illustrated), which is sensitive to pure tones relative to silent events, (2) the auditory “belt” cortices (light blue), which are sensitive to band pass noises (BPNs) relative to tones, and (3) the auditory “parabelt” cortices (dark blue), which are sensitive to human spoken phonemes relative to BPNs (*ibid*). Auditory belt cortex showed greater activation to the heard sounds compared to oral mimicry, and showed no differential activity at a categorical level (e.g. Fig. 4A, light blue, histogram). Importantly, the left aSTG region preferential for both hearing and mimicking animal vocalizations partially overlapped (purple) with the left parabelt auditory cortex (Fig. 4A–B, aSTG histogram). The homologous portion of the right aSTG showed activation preferential for hearing

vocalizations versus animal action sounds and overlapped with parabelt cortex, but did not show significant activation during oral mimicry of the animal vocalizations, even at lowered threshold settings of significance (cf. Figs. 4A and 3C, right hemisphere data). Thus, overlap between cortices preferential for phoneme-processing and for animal vocalization perception and oral mimicry (as opposed to animal action sound oral mimicry) was only present in the left cortical hemisphere aSTG region in this analysis.

To further characterize the anticipated sound mimicry networks, participants had rated each sound after scanning as to how difficult they thought it was to mimic. The animal action sounds were rated overall as being more difficult to mimic relative to the animal vocalizations (Action sound mimicry, Avg \pm SD = 3.30 ± 0.54 ; Vocalization mimicry, 2.52 ± 0.37 ; t -test = 5.78, $p < 7 \times 10^{-6}$). To critically test for the potential confound that the effects of oral mimicry by category (i.e. Fig. 4A–B) could have simply been due to task difficulty, we conducted a second analysis using five animal vocalizations rated as most difficult to mimic (Avg. = 2.90 ± 0.02) and the five animal action sounds rated as easiest to mimic (Avg = 2.62 ± 0.22), which can be heard online (hear Supplementary Audio and view 3D Neuroimaging data). The difficulty ratings of these subsets of sounds were not statistically different from one another (one tailed t -test = 2.13, $p > 0.188$), but were reverse-biased in the opposite direction for perceived task difficulty relative to the analysis using all of the retained data. Using the same ANOVA and t -test analyses, three of the foci survived (cf. Fig. 4A–B and D–E; Table 3), with vocalizations recruiting 63% of the total volume of differentially activated cortex. This included the left aSTG and left-lateralized retrosplenial cortex for hearing and mimicking animal vocalizations, versus the left vIPC region for hearing and orally mimicking animal action sounds. The left vIPC in both analyses spanned the central sulcus to involve somatosensory cortex (S1) and motor cortices overlapping the estimated locations for representations of laryngeal and oral-facial muscles (Penfield and Boldrey, 1937). Portions of the cerebellum also showed preferential activation to mimicry of the action sounds (data not shown). Together, these analyses revealed a network of operationally defined “echo-mirror neuron system” (ENS) regions (Rizzolatti and Craighero, 2004), including a region preferential for perceiving and mimicking animal action sounds (Fig. 4D, yellow cortex) and regions preferential for perceiving and mimicking animal vocalizations (red cortex).

3.2. Cortex parametrically sensitive to perceived mimicry quality

To assess the effects of the participant’s perception of performance of the oral mimicry task, measures of “how well” they thought they had orally mimicked each sound stimulus, collected immediately after fMRI scanning, were also assessed (see Methods). At a group-level, the brain regions showing the strongest parametric correlations with mimic quality self-ratings as a primary level analysis (Fig. 5A, green) included the right precentral gyri and right middle-cingulate regions, which were more strongly activated by the more “poorly” mimicked sounds (Fig. 5A, charts). In a second analysis, ROIs from the respective group-averaged action sound mimicry and vocalization mimicry regions (from Fig. 4A) were separately assessed further for sensitivity to the mimic quality self-ratings (Fig. 5B). Some regions showed parametric sensitivity to one versus the other category of sound mimicry, respecting the boundary of preferred semantic category. Moreover, most regions showed at

least a trend, if not significant correlation, of greater activation when sounds of the preferred category being mimicked were rated as increasingly poorer in quality, consistent with other action observation and imitation fMRI studies (Garrison et al., 2013; Liew, Han, & Aziz-Zadeh, 2011; Vogt et al., 2007). In other words, the sounds perceived as being more poorly mimicked generally led to greater activation, both in the regions of interest showing category specificity (Fig. 5B), as well as a few other cortical regions (Fig. 5A) distinct from the ENS-related foci.

3.3. Acoustic signal attributes of mimicry recordings

When participants orally mimicked the animal sounds, they tended to produce sounds with slightly shorter durations for both categories (Table 1). Additionally, the mimicked vocalization recordings were generally louder than the mimicked action sounds (by ~ 6 dB SPL) relative to the original animal sound stimuli, the latter of which had been carefully balanced for duration and intensity across categories. Consistent with an earlier study examining sound mimicry (Talkington, Rapuano, Hitt, Frum, & Lewis, 2012), human mimic recordings for both categories of sound showed greater harmonic content (Table 1, HNR entries) relative to the original animal sounds. The participant's mimicry recordings also showed on average a relatively greater degree of spectral flatness (Wiener entropy, WE) and relatively less spectral structure variation (SSV). Based in part on post-scanning interviews, this "degradation" in entropy-based measures and increase in harmonic content was likely due to (1) limitations in laryngeal and oral-facial movements by human participants when attempting to articulate the idiosyncrasies of sounds made by other animal's vocalizations, as well as (2) challenges in orally mimicking the action sounds that were originally produced by non-oral mechanisms (e.g. a horse trotting on dirt), and (3) challenges in producing sound while lying still in the MRI scanner. Nonetheless, these oral mimicry sound recording results demonstrated that all participants were able to comply with the task instructions of audibly mimicking all the sound stimuli during the fMRI scanning session.

Another acoustic attribute of the mimicry recordings we analyzed included voicing (vocal cord vibration), which was readily evident in spectrograms and clearly necessary when mimicking animal vocalizations. However, there was relatively little voicing during mimicry of the animal action sounds. Specifically, of the 288 recorded action sounds, roughly only 15% of the signals had any quantifiable voicing segments, which were defined as having at least two successive segments measured by an average magnitude difference function (AMDF: using a 40 ms window, 25% overlap, and 0.5 threshold). Upon inspection, many of the voicing elements appeared to be more accidental/incidental than intentional given the short durations of voicing segments relative to the ~ 2–3 s mimic events. These results indicated that intrinsic laryngeal muscle manipulation (voicing) was a predominant factor when mimicking animal vocalizations while oral-facial mechanisms other than voicing were predominant when mimicking the animal action sounds.

4. Discussion

The main finding of the present study was evidence for the existence of two left-lateralized cortical processing subsystems (i.e. Fig. 4D) that mediate category-specific aspects of both

the perception and oral mimicry of two distinct acoustic-semantic categories of natural sounds: animal action sounds versus animal vocalizations. The finding of a “double-dissociation” of oral mimicry networks, with consideration given to perceived mimicry difficulty and quality, was suggestive of separable neuronal architectures, and thus potentially separable cognitive processes (Shallice, 1988) related to aspects of auditory working memory, motor planning, and/or motor execution during oral mimicry at a categorical level. Importantly, these findings paralleled the dissociated cortical network organization reported in a recent neurobiological model of natural sound processing and perception (Brefczynski-Lewis and Lewis, 2017; Webster et al., 2017), which had been developed in part using some of the same ethologically validated, non-verbal sound stimuli. Because no overt speech or verbal processing was involved in either the task or stimuli, the present results may reflect cortical network vestiges underlying rudimentary elements of oral communication systems that are directly related to the statistics of physical attributes inherent to different semantic categories of real-world sound-producing events. Thus, these results provide a unique perspective on various models of oral communication in primates, spoken language processing and production in humans, as well as rehabilitation strategies for stroke patients with impaired movement execution systems affecting spoken language, as addressed in the following sections.

4.1. The neuroanatomy of sound mimicry in primates

In primates, the brain regions coordinating the 100 or so muscles involved in voice and oral-motor production are reported to include three major systems (Goldberg, 1992; Passingham, 1987; Simonyan and Horwitz, 2011): One is a system in the brainstem for controlling innate vocalizations (e.g. crying); a second involving limbic and medial prefrontal cortices (MPC) is for controlling voluntary emotional vocalizations (e.g. wince from pain, fear cries, warning calls); and a third involving lateral prefrontal cortices (LPC), plus a more highly developed “Broca’s area” in humans, which constitute systems that are under a greater degree of voluntary control, and are ultimately involved in controlling speech and song production (Démonet, Chollet, & Ramsay, 1992; Geschwind, Quadfasel, & Segarra, 1968; Liberman and Mattingly, 1985; Zatorre, Evans, Meyer, & Gjedde, 1992). In the present study, both of the non-verbal sound mimicry systems identified, operationally defined here as subsystems for animal action sounds and for animal vocalizations, involved recruitment from the lateral cortical control systems, and both were heavily lateralized to the left hemisphere (Fig. 4, yellow and red regions). Thus, at a gross anatomical level, the present findings were generally consistent with earlier neuropsychological studies and theoretical frameworks identifying lateral premotor systems as a part of the cognitive-level control systems for complex oral sound communication (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015; Goldberg, 1992; MacNeilage, 1998).

While the present paradigm could not distinguish between activation associated with direct sensory-motor control (some low-level acoustic or motoric features) versus phonological loops and other forms of auditory working memory (Baddeley, 1986; Wager and Smith, 2003), the results indicate that the oral mimicry of non-verbal, real-world natural sounds have a significant hemisphere lateralization bias, similar to aspects of lateralization biases observed in both right- and left-handed individuals for overt and covert speech production

(Devlin, Matthews, & Rushworth, 2003; Hickok and Poeppel, 2004; Shuster and Lemieux, 2005; Szaflarski et al., 2002), as well as for manual tool use and auditory perception of uni-manual tool-use sounds (Johnson-Frey, 2004; Johnson-Frey, Newman-Norlund, & Grafton, 2005; Lewis, 2006; Lewis, Phinney, Brefczynski-Lewis, & DeYoe, 2006). Brief summaries of the proposed functional roles of the three main cortical regions of interest (from Fig. 4D–E; also see Supplemental 3D neuroimaging data online) are addressed next, followed by discussion of their possible role(s) in the context of models of spoken language processing and production.

4.2. Functional roles of the main regions of interest

4.2.1. Left aSTG—The subsystem for perception and oral mimicry of animal vocalization notably involved the left aSTG (Fig. 4D–E, red), which is traditionally considered as auditory cortex rather than motor-related cortex (Belin et al., 2000; Binder et al., 2000; Chevillet et al., 2011; Rauschecker and Scott, 2009). Indeed, this focus directly overlapped cortex independently identified as phoneme-sensitive parabelt auditory cortex (Fig. 4, dark blue). Conceivably, this left aSTG activation during mimicry could have reflected some form of sensory feedback (hearing one’s own voice) or efferent copy related to vocal production (Curio, Neuloh, Numminen, Jousmaki, & Hari, 2000; Eliades and Wang, 2005; Fu et al., 2006; Gunji, Hoshiyama, & Kakigi, 2001; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Ventura, Nagarajan, & Houde, 2009; Wise, Greene, Buchel, & Scott, 1999). A simple peripheral sensory-feedback explanation alone, however, seemed unlikely in the present study for several reasons. First, the participant’s own voice during mimicry events could only be heard muffled through the hearing protection and through bone conduction, while the headphone-delivered animal vocalization (and action sound) stimuli were heard with much greater clarity and intensity: fMRI studies of hearing self-vocalizations in isolation generally report a relative reduction in activation during vocalizations compared to hearing one’s own vocal playback (Christoffels, Formisano, & Schiller, 2007), and thus the effects of sound attenuation biases against this simple sensory-feedback explanation given the robust dissociated activation profile in the left aSTG during oral mimicry (Fig. 4, histograms). Secondly, feedback during one’s own speech typically affects auditory cortices in both hemispheres (Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Price et al., 1996), rather than strongly lateralized effects as observed in the present study. Thirdly, the animal vocalizations were assessed as having greater prosodic and emotional elements: Thus, if participants were effectively engaging a sound-feedback strategy for accurate oral emulation, then by some accounts this might be expected to engage bilateral or even right hemisphere biased activation (Friederici and Alter, 2004; Grossmann, Oberecker, Koch, & Friederici, 2010; Schirmer and Kotz, 2006), though see discussion on language production models (Section 4.4). Collectively, these observations suggest that the left aSTG activation during oral mimicry of animal vocalizations was not consistent with simple peripheral sensory feedback effects alone. Rather, this finding was more consistent with reflecting functions related to matching sensory-to-motor mappings, indexing predictive errors of one’s own voice during imitation, and/or motor preparation/execution related to vocalizations as a distinct category of natural sound, as addressed in later sections.

4.2.2. Left retrosplenial cortex—The left-lateralized retrosplenial region (also variably termed as posterior cingulate, precuneate cortex, Brodmann area 31, and medial parietal cortex) was preferentially activated during both hearing and orally mimicking animal vocalizations relative to animal action sounds (Fig. 4D, red). Cortex in this vicinity has a wide range of reported functions related to episodic memory, imagery, and attention (Fletcher et al., 1995; Raichle et al., 2001; Shulman et al., 1997), but perhaps most pertinent to the present study is its proposed role in the processing of emotional states (Maddock, 1999). In particular, the right retrosplenial cortex is commonly activated in studies involving emotionally salient pictorial stimuli, while the left retrosplenial cortex is activated more prominently with emotionally salient verbal stimuli (Maddock and Buonocore, 1997). In general, the animal vocalizations of the present study were rated as significantly more emotionally valent than the relatively neutral ratings for the animal action sounds (Table 1). Earlier attempts to balance sound stimuli across the two categories for emotional valence proved to be prohibitively challenging given the distributions of both negative and positive rating extremes for animal vocalizations and individual variability in ratings. Perceptual features related to emotional valence and prosody in vocal calls may thus prove to statistically reflect acoustic-semantic attributes that inherently lend themselves to processing as a distinct category of natural sound, which is an issue addressed later in the context of language system evolution (Section 4.5).

4.2.3. Left vlPC lobule—The left-lateralized ventro-lateral paracentral (vlPC) lobule, overlapping the estimated location of primary somatosensory cortex (“S1”) and dorsal portions of laryngeal motor cortex (dLMC) (Penfield and Boldrey, 1937), were generally consistent with representing a somatosensory-motor feedback circuit involved in controlling oral-facial sound production (Baddeley, 1986). The data analyses addressing perceived mimic difficulty and mimic quality further attested to this interpretation in that the sounds rated as more poorly mimicked (and for the action sounds typically rated more difficult to mimic) generally led to parametrically greater degrees of activation in the vlPC as well as in other ROIs, especially for the region’s preferred category of sound. The left posterior portion of the vlPC lobule (involving “S1”) appeared to partially overlap with a previously reported focus associated with poor versus good mimickers of foreign language accents (Reiterer, Hu, Sumathi, & Singh, 2013), and thus a region that has been previously associated with oral mimicry of speech sounds.

Working memory theories of brain function indicate that left lateralized portions of the ventrolateral prefrontal cortex (i.e. near Broca’s area) have functions especially related to simple storage tasks related to verbal working memory (Wager and Smith, 2003). In this regard, the left vlPC lobule activation focus of the present study was also consistent with a function related to rehearsal and/or preparation for oral production. Relative to the vocal calls, the action sounds of the present study were characterized by greater low frequency amplitude modulations and temporal sequences that may have had to be remembered, potentially rehearsed, and then orally produced on cue, requiring working memory operations on a time scale commensurate with temporal cadences characteristic of oral and manual gestures by motoric systems (MacNeilage, 1998). Although the sound stimuli and task of oral mimicry were all “non-verbal”, and in some regards could be considered

independent of language systems (e.g. a parrot or minor bird could in principle have performed our paradigm), these results should, of course, also be interpreted in the context of spoken language models, as addressed next.

4.3. Relation to models of speech perception and production

4.3.1. Echo-mirror neuron subsystems—Contemporary gestural theories of language evolution have often been addressed in the context of mirror neuron systems (MNS), and related echo-mirror neuron systems (ENS). ENS systems are defined as brain regions showing activation both when hearing and when mimicking sounds produced during behaviorally meaningful events, and reported to be present both in humans (Molenberghs, Cunnington, & Mattingley, 2012; Rizzolatti and Craighero, 2004, 2007) and macaque monkeys (Keysers et al., 2003; Kohler et al., 2002). More specifically, ENS-like networks are thought to function to provide a probabilistic match between cortical representations of heard action sounds with those associated with representing the listener’s own repertoire of sound-producing actions (e.g. walking quickly through leaves), thereby providing a sense of meaning or intention behind the sound and soundsource when produced by other agents (Buccino et al., 2001; Engel et al., 2009; Galati et al., 2008; Gazzola, Aziz-Zadeh, & Keysers, 2006; Lahav, Saltzman, & Schlaug, 2007; Lewis et al., 2011). The present study supported this general framework as a component of oral mimicry, especially for the animal action sounds, which could arguably be readily “embodied” through observation visually and/or acoustically.

Based on the present findings, the left vIPC lobule may be considered as part of an ENS subsystem (Fig. 4D, yellow), more generally utilized for orally mimicking or producing action sounds, which inherently contain more temporally distinctive spectro-temporal acoustic attributes relative to vocalizations—wherein there are typically far greater intermodal invariant audio-visual associations with this category of real-world action event. Similarly, the left aSTG and retro- splenial region (Fig. 4D, red) may be considered as part of another ENS- like subsystem that is more generally utilized for embodying vocalizations, which are predominantly or uniquely processed in the auditory domain (rather than visual domain)—having notably fewer intermodal invariant audio-visual associations, since vocal cords and surrounding vocal tract structure movements are typically not directly seen. Thus, audio-visual attributes, or a lack thereof, may reflect another level of sensory-semantic universal features that contribute statistically to the recruitment (and evolution) of separable pathways related to oral communication of these different categories of natural events. Future studies in non-human primates with oral mimicry ability, though challenging, might reveal more about the generalizability of this architecture in regard to mimicry and oral communication hypotheses.

4.3.2. Dorsal-ventral streams for sound processing and production—Sound processing pathways in cortex are proposed to utilize two major divisions, a dorsal “what is it” stream and a ventral “where is it” stream (Bornkessel-Schlesewsky et al., 2015; Rauschecker and Scott, 2009, 2015; Rauschecker, 1998; Romanski and Averbach, 2009; Saur et al., 2008; Ueno and Lambon Ralph, 2013). The results of the present study fit reasonably well with the hypothesized dorsal-ventral divisions of sound processing (and

production) in the primate auditory system. Briefly, ventral streams are thought to have developed to convey increasing degrees of communicative content in terms of spectral forms, which include analyses of vocal sound signals characterized by strong harmonic content and spectro-temporal variations therein (Lewis et al., 2009; Medvedev, Chiao, & Kanwal, 2002; Zatorre and Belin, 2001; Zatorre, Bouffard, & Belin, 2004). Conversely, dorsal streams, which more heavily interconnect postero-medial auditory cortices with sensory-motor parietal and frontal cortices, are proposed to play a role in temporal dynamic processing, including processing related to speech sound framing, sequencing, and articulation in humans (Arbib, 2005; Corballis, 1999; MacNeilage, 1998). These temporal sequencing functions are generally consistent with the notion that the dorsal auditory pathways were archaically more fundamentally suited for conveying information about (1) where in space a sound-source is located dynamically relative to the listeners' body and limbs—for purposes of movements to avoid or engage the sound-source, and (2) what the sound source is in terms of whether or how well an action sound sequence fits with the listeners' own repertoire of sound-producing motor actions—thereby providing a sense of meaning through “embodiment” (addressed earlier in the context of ENS systems; Section 4.3.1). However, to mesh with the dorsal-ventral stream hypothesis, the left aSTG focus of the present study would effectively need to be considered as part of the “dorsal” pathway in humans, which is anatomically conceivable given the interconnections of the arcuate fasciculus, as addressed next.

The STG cortices of humans, relative to great apes and monkeys, have disproportionately greater connections with prefrontal regions via the arcuate fasciculus, which is a major white matter tract that is especially prominent in the left hemisphere (Rilling et al., 2008). This fiber tract is also reported to be modulated by long term vocal-motor training (Halwani, Loui, Ruber, & Schlaug, 2011), and lesions to this tract may lead to conduction aphasia, wherein an individual's ability to repeat words (speech repetition) is disrupted despite relatively normal speech comprehension and speech production (Bernal and Ardila, 2009; Tanabe et al., 1987; Zhang et al., 2010). From the human connectome project, cortical territories overlapping the left aSTG region of the present study showed significant resting state functional connectivity with left inferior frontal regions (Glasser et al., 2016; Jakobsen et al., 2016), further supporting these regions as forming a highly interconnected functional subnetwork. The greater left aSTG connectivity with frontal regions in humans, relative to other hominids, may thus have a greater role in directly modulating left-lateralized motor-related articulatory systems necessary for accurate mimicry of complex animal vocalizations, and presumably greater ability for emotional expression and communication through voice.

4.4. Neural theories of speech acquisition and production

Models of speech acquisition and production, such as the DIVA model (directions into velocities of articulators), posit that neural control entails interactions of three subsystems, including: (a) an auditory feedback control system that transforms auditory errors into corrective motor commands; (b) a somatosensory feedback control system that encodes somatosensory targets for speech sounds and corrective motor commands; and (c) a feedforward control system (Guenther and Vladusich, 2012; Tourville and Guenther, 2011).

These systems feed into a speech sound map that controls articulation of oral- vocal sound production. In the context of this neural computational model, an alternative (though not necessarily exclusive) interpretation of the present results also seems plausible. In particular, the left aSTG for vocalization mimicry (together with input from retrosplenial cortex) may have been more heavily engaged in indexing prediction errors made during vocal control (i.e. voicing-weighted feedback with emotional content matching). In contrast, the left vIPC lobule activation may have been more heavily weighted by relatively greater needs for oral-facial somatosensory feedback operations. Note, that both categories of oral mimicry recruited mid-cingulate activation that was correlated with mimicry performance (Fig. 5A), which is near regions thought to be more generally involved in feedback and conflict monitoring (Carter and van Veen, 2007; Christoffels et al., 2007; Sohn, Albert, Jung, Carter, & Anderson, 2007). In this regard, the differential results by category, despite each participant's extensive practice with the task prior to scanning, may relate to life-long differences in experience with oral sound production: The animal vocalization category, for instance, may be argued to contain many sounds that human toddlers are encouraged to mimic from a very early age, while the actions sounds (incidental sounds of locomotion) may contain relatively more difficult and less familiar sounds that are less frequently practiced in modern societies. Future tests of spoken language models using nonverbal natural sound categories based on the rationale behind the present study (though with more systematically controlled low-level acoustic and/or motor output features), may prove useful for elucidating the functional roles of different regions or subsystems, revealing functions that may have antedated modern language systems in hominins.

4.5. Acoustic-semantic universals and glottogenesis theories

The left-lateralized activations related to oral mimicry of different acoustic-semantic categories of natural sound may have reflected underlying nascent circuits and metamodal network organizations (Pascual-Leone and Hamilton, 2001) upon which oral mimicry, oral mimesis, oral communication, and ultimately speech, reside. Alternatively, the already matured networks subserving speech-sound perception and production in our adult participants may have simply been recruited as the most efficient means to orally mimic the different categories of sound (addressed earlier), an alternative that may be best addressed by future studies with prelingual infants. Assuming the former interpretation, however, the potential impact of the present findings on mechanistic theories of the evolution of oral communication is briefly outlined below.

Manual gesture language models suggest that increases in brain size, with judiciously interconnected networks, enabled greater cognitive ability for mimicry and social intelligence in general (Darwin, 1871/1981; Donald, 1991). Greater cortical expanses and specializations in processing led to a need for more efficient local circuit processing and thus to hemispheric asymmetries (Morillon et al., 2010; Preuss, 2011), leading to functional lateralizations such as for handedness and skilled tool use (and tool making) with a dominant hand (Cashmore, Uomini, & Chapelain, 2008; Stout, Toth, Schick, & Chaminade, 2008). The ability to produce increasingly complex oral mimicry and to represent acoustic abstractions related to manual gestures, as well as the ability to decode similar oral mimics by other conspecifics, could confer significant advantages in social communication and thus

survival for some hominin species (Corballis, 2003; Donald, 1991; Hewes, 1973). Vocal-acoustic communication forms have obvious advantages relative to vision-dominated manual gestures and sign languages, in that they are known to be faster, not require a need of being in line of sight, can be communicated in total darkness and over longer distances, and allow mothers rearing infants to have more freedom with the hands for foraging at a distance (Hewes, 1973). Lateralizations for processing certain classes of vocal calls have been reported in other primate species (Poremba et al., 2004; Tagliabata, Russell, Schaeffer, & Hopkins, 2009; Talkington et al., 2012; Talkington, Tagliabata, & Lewis, 2013), supporting the idea that more advanced vocalization processing and decoding is associated with brain asymmetries. However, assuming that improvements in oral mimicry of real-world natural sound-producing events were critical to oral communication evolution, what sounds were most likely needed to be imitated in early stages of hominin evolution?

Two of the most rudimentary semantic categories of natural sound that are proposed to have contributed to the early stages of hominin oral communication include vocalizations and incidental sounds of locomotion (Falk, 2004; Larsson, 2014, 2015). Changes in oral communication ability that could improve, for instance, big-game social hunting (pre-hunt communications) and emotional communication have been proposed to play an important role leading to selective advantages for vocal-sound decoding (Donald, 1991; Hewes, 1973; Szamado, 2011). Transitions from episodic to mimetic cultures are thought to include miming, dancing and vocalizing the motor habits of other species with increasing degrees of abstract mimesis, which reflect acoustic-gestural actions that could convey greater levels of communicative content. Selection pressures for gradual changes conferring greater abilities in decoding and/or encoding such sounds could ostensibly convey a more diverse range of communicated concepts. This could reflect an individual's multi-modal cognitive analyses of environmental actions and events, and enhance communication and planning of increasingly complex cooperative interactions among troop members.

Evolutionary changes in neural architectures that link auditory with motor circuits (e.g. via the arcuate fasciculi) may be used to more effectively represent cross-modal abstractions of different semantic categories of gestures, action events, the unique identity of different sound-sources and objects in the natural environment, and in conveying emotional states. The pairing of regularities particular to sensory-motor representations with sound symbols of intended meaning ("iconicity") may naturally lead to acoustic form-meaning mappings that can thus be communicated orally. Importantly, only relatively small evolutionary changes in oral-facial anatomy and function leading to greater precision or control of oral sound production (and decoding) would be needed for oral communication development, adhering to continuity models and least biological resistance tenets of evolutionary change.

Gradual advances in sound symbolism representations along the lines outlined above may thus have served as initial scaffolding mechanisms for the brain of early hominins to subsequently evolve in capacity to formalize and unify knowledge and thought, which could be communicated by adoption of arbitrary sounds, notably short utterances and words, to more abstractly refer to sound-producing action events and their semantic referents (Gogate and Hollich, 2010; Kanero et al., 2014; Maurer, Pathman, & Mondloch, 2006; Schmidtke, Conrad, & Jacobs, 2014; Yoshida, 2012). In its extreme form, this had been proposed, and

later abandoned, as the “ding-dong” theory on origins of language (Mithen, 2006; Müller, 1866). Though well outside the scope of the present study, greater abstractions from spoken words to communication presumably involved grammar, which may have been further rooted in the sequential and temporal processing forms related to gesture and skilled manual control operations (Chomsky, 2002; Jackendoff, 2003; Kemmerer, 2014; Perlovsky, 2011; Pinker, 1994), and remains a fascinating area for future research. Notwithstanding, the use of acoustic-semantic universals as fundamental sound symbolism referents that could become amenable to mimicry, and potentially mimesis, would reflect a bottom-up, bootstrapping mechanism that supports and advances mouth-gestural continuity theories behind the origin of spoken language systems, and potentially more efficient rehabilitation of such systems after brain injury.

4.6. Neural theories of speech recovery after stroke or brain injury

Among the many treatments for stroke patients are those involving mirror therapy and action observation and imitation, representing a possible means for providing cortical reorganization (Garrison et al., 2010; Pomeroy et al., 2005). Observing actions of others are reported to activate specific motor plans in damaged and penumbral motor-related circuits after stroke (Garrison et al., 2013; Pomeroy et al., 2011), though with critical dependence on what, when (post-stroke), and to what extent patients actively imitate (Cowles et al., 2013; Ertelt and Binkofski, 2012). The present study provides a novel test bed for rehabilitation therapies geared toward recovering different aspects of oral mimicry and communication, with the idea that distinct circuits are preferentially involved in separable cognitive processes related to oral communication.

In sum, the present study provides novel evidence suggestive of dissociable cortical subsystems that mediate aspects of both hearing and oral mimicry of different categories of natural sound-producing events. Specifically, this included mimicry of animal action sounds (e.g. incidental sounds of locomotion) and animal vocalizations, both of which are sound categories independently hypothesized to represent categories of events that would have been especially important to mimic in early stages of hominin oral communication evolution. Based on the present findings we propose that networks for oral mimicry are founded, at least in part, on cortical organizations of sensory systems, including acoustic-semantic and sensory-semantic universals associated with sound-producing events of the natural environment. The brain regions recruited in the present study may thus reflect vestiges of fundamental oral communication architectures that ultimately develop to subserve more abstract representations related to spoken language functions. These findings, in accord with our neurobiological model of sound perception, provides a theoretical framework for further exploration of the underpinnings of oral communication systems and spoken language evolution, as well as advancing clinically relevant models of oral mimicry acquisition in infants, and models of language recovery and speech movement execution system recovery after aphasia(s) resulting from stroke or traumatic brain injury.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Bob Hou (Dept. Radiology at West Virginia University; WVU) for technical support with MRI scanning, and Dr. Mark Chevillet for providing sound stimuli for auditory cortex mapping. This work was supported by the Dept. Radiology, NIGMS NIH COBRE grant GM103503, and affiliated WVU Summer Undergraduate Research Internships.

References

- Arbib MA (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–124 (discussion 125–167). [PubMed: 16201457]
- Asano M, Imai M, Kita S, Kitajo K, Okada H, & Thierry G (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, 63, 196–205. [PubMed: 25282057]
- Baddeley AD (1986). Working memory. In Gazzaniga MS (Ed.). *Principles of neuroscience*. MIT Press: Clarendon Press.
- Baumann S, Griffiths TD, Rees A, Hunter D, Sun L, & Thiele A (2010). Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. *Neuroimage*, 50, 1099–1108. [PubMed: 20053384]
- Belin P, Zatorre RJ, Lafaille P, Ahad P, & Pike B (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312. [PubMed: 10659849]
- Bernal B, & Ardila A (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain*, 132, 2309–2316. [PubMed: 19690094]
- Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, & Possing E (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512–528. [PubMed: 10847601]
- Bornkessel-Schlesewsky I, Schlesewsky M, Small S, & Rauschecker J (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, 19, 142–150. [PubMed: 25600585]
- Brefczynski-Lewis JA, & Lewis JW (2017). Auditory object perception: A neurobiological model and prospective review. *Neuropsychologia*.
- Brown S, Ngan E, & Liotti M (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, 18, 837–845. [PubMed: 17652461]
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, ... Freund H-J (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13, 400–404. [PubMed: 11168545]
- Carter CS, & van Veen V (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, 7, 367–379.
- Cashmore L, Uomini N, & Chapelain A (2008). The evolution of handedness in humans and great apes: A review and current issues. *The Journal of Archaeological Science*, 35, 7–35.
- Chevillet M, Riesenhuber M, & Rauschecker JP (2011). Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *The Journal of Neuroscience*, 31, 9345–9352. [PubMed: 21697384]
- Chomsky N (2002). *On nature and language*. Cambridge, UK: Cambridge University Press.
- Christoffels IK, Formisano E, & Schiller NO (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, 28, 868–879. [PubMed: 17266104]
- Clarke S, Bellmann A, de Ribaupierre F, & Assal G (1996). Non-verbal auditory recognition in normal subjects and brain-damaged patients: Evidence for parallel processing. *Neuropsychologia*, 34, 587–603. [PubMed: 8736571]
- Condillac EB (1746 (1947)). *Essai sur l'origine des connaissances humaines, ouvrage ou l'on réduit a un seul principe tout ce concerne l'entendement*. In: LeRoy G (Ed.), *Oeuvres philosophiques de Condillac* Paris.
- Corballis MC (1999). The gestural origins of language. *American Scientist*, 87, 138–145.

- Corballis MC (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, 26, 199–208. [PubMed: 14621511]
- Cowles T, Clark A, Mares K, Peryer G, Stuck R, & Pomeroy V (2013). Observation- to-imitate plus practice could add little to physical therapy benefits within 31 days of stroke: ranslational randomized controlled trial. *Neurorehabilitation and Neural Repair*, 27, 173–182. [PubMed: 22798151]
- Cox RW (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162–173. [PubMed: 8812068]
- Curio G, Neuloh G, Numminen J, Jousmaki V, & Hari R (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*, 9, 183–191. [PubMed: 10770228]
- Darwin C (1871/1981). *The descent of man, and selection in relation to sex*.
- Démonet J-F, Chollet F, & Ramsay S (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1752–1768.
- Devlin JT, Matthews PM, & Rushworth MF (2003). Semantic processing in the left inferior prefrontal cortex: A combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 15, 71–84. [PubMed: 12590844]
- Donald M (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
- Eklund A, Nichols TE, & Knutsson H (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 7900–7905. [PubMed: 27357684]
- Eliades SJ, & Wang X (2005). Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex*, 15, 1510–1523. [PubMed: 15689521]
- 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. [PubMed: 19465134]
- Engelien A, Tuscher O, Hermans W, Isenberg N, Eidelberg D, Frith C, ... Silbersweig D (2006). Functional neuroanatomy of non-verbal semantic sound processing in humans. *Journal of Neural Transmission*, 113, 599–608. [PubMed: 16075182]
- Ertelt D, & Binkofski F (2012). Action observation as a tool for neurorehabilitation to moderate motor deficits and aphasia following stroke. *Neural Regeneration Research*, 7, 2063–2074. [PubMed: 25624838]
- Falk D (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, 27, 491–503. [PubMed: 15773427]
- Fletcher PC, Frith CD, Baker SC, Shallice T, Frackowiak RS, & Dolan RJ (1995). The mind's eye-precuneus activation in memory-related imagery. *Neuroimage*, 2, 195–200. [PubMed: 9343602]
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, & Noll DC (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33, 636–647. [PubMed: 7596267]
- Friederici AD, & Alter K (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89, 267–276. [PubMed: 15068909]
- Fu CH, Vythelingum GN, Brammer MJ, Williams SC, Amaro E Jr., Andrew CM, ... McGuire PK (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, 16, 969–977. [PubMed: 16195470]
- Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, & Pizzamiglio L (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage*, 40, 1274–1286. [PubMed: 18276163]
- Garrison KA, Aziz-Zadeh L, Wong SW, Liew SL, & Winstein CJ (2013). Modulating the motor system by action observation after stroke. *Stroke*, 44, 2247–2253. [PubMed: 23743974]
- Garrison KA, Winstein CJ, & Aziz-Zadeh L (2010). The mirror neuron system: A neural substrate for methods in stroke rehabilitation. *Neurorehabilitation and Neural Repair*, 24, 404–412. [PubMed: 20207851]
- Gazzola V, Aziz-Zadeh L, & Keysers C (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, 16, 1824–1829. [PubMed: 16979560]

- Geangu E, Quadrelli E, Lewis JW, Macchi Cassia V., & Turati C (2015). By the sound of it. An ERP investigation of human action sound processing in 7-month-old infants. *Developmental Cognitive Neuroscience*, 12, 134–144. [PubMed: 25732377]
- Geschwind N, Quadfasel FA, & Segarra JM (1968). Isolation of speech area. *Neuropsychologia*, 6, 327-+.
- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, ... Van Essen DC (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536, 171–178. [PubMed: 27437579]
- Gogate LJ, & Hollich G (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117, 496–516. [PubMed: 20438235]
- Goldberg G (1992). Premotor systems: attention to action and behavioral choice In Klein J, McCrohan C, & Winlow W (Eds.). *Neurobiology of motor program selection* Pergamon Press.
- Grossmann T, Oberecker R, Koch SP, & Friederici AD (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65, 852–858. [PubMed: 20346760]
- Guenther FH, & Vladusich T (2012). A neural theory of speech acquisition and production. *The Journal of Neurolinguistics*, 25, 408–422. [PubMed: 22711978]
- Gunji A, Hoshiyama M, & Kakigi R (2001). Auditory response following vocalization: A magnetoencephalographic study. *Clinical Neurophysiology*, 112, 514–520. [PubMed: 11222973]
- Halwani GF, Loui P, Ruber T, & Schlaug G (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, 2, 156. [PubMed: 21779271]
- Heinks-Maldonado TH, Mathalon DH, Gray M, & Ford JM (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42, 180–190. [PubMed: 15787855]
- Hewes GW (1973). Primate communication and the gestural origin of language. *Current Anthropology*, 14, 5–24.
- Hickok G, & Poeppel D (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67–99. [PubMed: 15037127]
- Imai M, & Kita S (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 20130298.
- Imai M, Kita S, Nagumo M, & Okada H (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54–65. [PubMed: 18835600]
- Jackendoff R (2003). *Precis of Foundations of language: brain, meaning, grammar, evolution*. *Behavioral and Brain Sciences*, 26, 651–665 (discussion 666–707). [PubMed: 15377127]
- Jakobsen E, Liem F, Klados MA, Bayrak S, Petrides M, & Margulies DS (2016). Automated individual-level parcellation of Broca's region based on functional connectivity. *Neuroimage*.
- Johannesson A (1950). The gestural origin of language; evidence from six 'unrelated' languages. *Nature*, 166, 60–61. [PubMed: 15439119]
- Johnson-Frey SH (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, 8, 71–78. [PubMed: 15588811]
- Johnson-Frey SH, Newman-Norlund R, & Grafton ST (2005). A distributed left hemisphere network active during planning of everyday tool use skills. *Cerebral Cortex*, 15, 681–695. [PubMed: 15342430]
- Kanero J, Imai M, Okuda J, Okada H, & Matsuda T (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS One*, 9, e97905.
- Kemmerer D (2014). Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. *Cortex*, 58C, 27–51.
- Keysers C, Kohler E, Umiltà A, Nanetti L, Fogassi L, & Gallese V (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153, 628–636. [PubMed: 12937876]
- Kohler E, Keysers C, Umiltà A, Fogassi L, Gallese V, & Rizzolatti G (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297, 846–848. [PubMed: 12161656]

- Lahav A, Saltzman E, & Schlaug G (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience*, 27, 308–314. [PubMed: 17215391]
- Larsson M (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Animal Cognition*, 17, 1–14. [PubMed: 23990063]
- Larsson M (2015). Tool-use-associated sound in the evolution of language. *Animal Cognition*, 18, 993–1005. [PubMed: 26118672]
- Lewis JW (2006). Cortical networks related to human use of tools. *Neuroscientist*, 12, 211–231. [PubMed: 16684967]
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, & DeYoe EA (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience*, 25, 5148–5158. [PubMed: 15917455]
- Lewis JW, Phinney RE, Brefczynski-Lewis JA, & DeYoe EA (2006). Lefties get it “right” when hearing tool sounds. *Journal of Cognitive Neuroscience*, 18, 1314–1330. [PubMed: 16859417]
- 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. *Journal of Cognitive Neuroscience*, 23, 2079–2101. [PubMed: 20812786]
- 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. *Frontiers in Systems Neuroscience*, 6(27), 1–16. [PubMed: 22291622]
- 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. *The Journal of Neuroscience*, 29, 2283–2296. [PubMed: 19228981]
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, & DeYoe EA (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, 14, 1008–1021. [PubMed: 15166097]
- Lieberman AM, & Mattingly IG (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36. [PubMed: 4075760]
- Liew SL, Han S, & Aziz-Zadeh L (2011). Familiarity modulates mirror neuron and mentalizing regions during intention understanding. *Human Brain Mapping*, 32, 1986–1997. [PubMed: 20882581]
- Loucks TM, Poletto CJ, Simonyan K, Reynolds CL, & Ludlow CL (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *Neuroimage*, 36, 131–143. [PubMed: 17428683]
- MacNeilage PF (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499–511 (discussion 511–446). [PubMed: 10097020]
- Maddock RJ (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neurosciences*, 22, 310–316. [PubMed: 10370255]
- Maddock RJ, & Buonocore MH (1997). Activation of left posterior cingulate gyrus by the auditory presentation of threat-related words: An fMRI study. *Psychiatry Research*, 75, 1–14. [PubMed: 9287369]
- Maurer D, Pathman T, & Mondloch CJ (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316–322. [PubMed: 16669803]
- Medvedev AV, Chiao F, & Kanwal JS (2002). Modeling complex tone perception: Grouping harmonics with combination-sensitive neurons. *Biological Cybernetics*, 86, 497–505. [PubMed: 12111277]
- Mikl M, Marecek R, Hlustik P, Pavlicova M, Drastich A, Chlebus P, ... Krupa P (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging*, 26, 490–503. [PubMed: 18060720]
- Mithen S (2006). The ‘Singing Neanderthals’: The origins of music, language, mind and body. *Cambridge Archaeological Journal*, 16, 97–112.
- Molenberghs P, Cunnington R, & Mattingley JB (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*, 36, 341–349. [PubMed: 21782846]

- Morillon B, Lehongre K, Frackowiak RS, Ducorps A, Kleinschmidt A, Poeppel D, & Giraud AL (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 18688–18693. [PubMed: 20956297]
- Müller FM (1866). *Lectures on the science of language*. Lugmans, Green and Co.
- Paget RA (1944). The origin of language. *Science*, 99, 14–15. [PubMed: 17844549]
- Paget R (1963). *Humanspeech: Some observations, experiments and conclusions as to the nature, origin, purpose and possible improvement of human speech*. London: Routledge and Kegan Paul.
- Pascual-Leone A, & Hamilton R (2001). The metamodal organization of the brain. *Progress in Brain Research*, 134, 427–445. [PubMed: 11702559]
- Passingham RE (1987). Two cortical systems for directing movement In Bock G, O'Connor M, & Marsh J (Eds.). *Ciba foundation symposium*. Wiley.
- Penfield W, & Boldrey E (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 1x, 389–443.
- Perlovsky L (2011). Language and cognition interaction neural mechanisms. *Computational Intelligence and Neuroscience*, 2011, 454587. [PubMed: 21876687]
- Perniss P, & Vigliocco G (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 20130300. [PubMed: 25092668]
- Pinker S (1994). *The language instinct*. London: Penguin Books.
- Pomeroy V, Aglioti SM, Mark VW, McFarland D, Stinear C, Wolf SL, ... Fitzpatrick SM (2011). Neurological principles and rehabilitation of action disorders: Rehabilitation interventions. *Neurorehabilitation and Neural Repair*, 25, 33S–43S. [PubMed: 21613536]
- Pomeroy VM, Clark CA, Miller JS, Baron JC, Markus HS, & Tallis RC (2005). The potential for utilizing the “mirror neurone system” to enhance recovery of the severely affected upper limb early after stroke: A review and hypothesis. *Neurorehabilitation and Neural Repair*, 19, 4–13. [PubMed: 15673838]
- 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. [PubMed: 14749833]
- Preuss TM (2011). The human brain: Rewired and running hot. *Annals of the New York Academy of Sciences*, 1225(Suppl. 1), E182–E191. [PubMed: 21599696]
- Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, ... Friston KJ (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain*, 119(Pt 3), 919–931. [PubMed: 8673502]
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, & Shulman GL (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 676–682. [PubMed: 11209064]
- Ramachandran VS, & Hubbard EM (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society B: Biological Sciences*, 268, 979–983.
- Rauschecker JP (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neurotology*, 3, 86–103. [PubMed: 9575379]
- Rauschecker JP, & Scott SK (2015). Pathways and streams in the auditory cortex In Hickok GS, & Small SL (Eds.). *Neurobiology of language* (pp. 287–298). Elsevier.
- Rauschecker JP, & Scott SK (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724. [PubMed: 19471271]
- Reiterer SM, Hu X, Sumathi TA, & Singh NC (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology*, 4, 782. [PubMed: 24155739]
- 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. *Nature Neuroscience*, 11, 426–428. [PubMed: 18344993]
- Rizzolatti G, Craighero L (2007). Language and mirror neurons. In: Gaskell MG (Ed.), *The Oxford handbook of psycholinguistics* Oxford.

- Rizzolatti G, & Craighero L (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Romanski LM, & Averbeck BB (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annual Review of Neuroscience*, 32, 315–346.
- Sapir E (1929). A study in phonetic symbolism. *Journal of Experimental psychology*, 12, 225–239.
- Saur D, Kreher BW, Schnell S, Kummerer D, Kellmeyer P, Vry MS, ... Weiller C (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18035–18040. [PubMed: 19004769]
- Schirmer A, & Kotz SA (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10, 24–30. [PubMed: 16321562]
- Schmidtke DS, Conrad M, & Jacobs AM (2014). Phonological iconicity. *Frontiers in Psychology*, 5, 80. [PubMed: 24575062]
- Shallice T (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Shulman GL, Fiez JA, Corbetta M, Buckner RL, Meizen FM, & Raichle ME (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience*, 9, 648–663. [PubMed: 23965122]
- Shuster LI, & Lemieux SK (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, 93, 20–31. [PubMed: 15766765]
- Simonyan K (2014). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurology*, 28, 15–21.
- Simonyan K, & Horwitz B (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist*, 17, 197–208. [PubMed: 21362688]
- Sohn MH, Albert MV, Jung K, Carter CS, & Anderson JR (2007). Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 10330–10334. [PubMed: 17563353]
- Sterelny K (2012). Language, gesture, skill: The co-evolutionary foundations of language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367, 2141–2151. [PubMed: 22734057]
- Stout D, Toth N, Schick K, & Chaminade T (2008). Neural correlates of Early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 1939–1949. [PubMed: 18292067]
- Szaflarski JP, Binder JR, Possing ET, McKiernan KA, Ward BD, & Hammeke TA (2002). Language lateralization in left-handed and ambidextrous people: fMRI data. *Neurology*, 59, 238–244. [PubMed: 12136064]
- Szamado S (2011). Pre-hunt communication provides context for the evolution of early human language. *Biological Theory*, 5, 366–382.
- Tagliavolanti JP, Russell JL, Schaeffer JA, & Hopkins WD (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, 19, 1151–1157. [PubMed: 18787228]
- Talairach J, & Tournoux P (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.
- Talkington WJ, Rapuano KM, Hitt LA, Frum CA, & Lewis JW (2012). Humans mimicking animals: A cortical hierarchy for human vocal communication sounds. *The Journal of Neuroscience*, 32, 8084–8093. [PubMed: 22674283]
- Talkington WJ, Tagliavolanti JP, & Lewis JW (2013). Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. *Hearing Research*, 305, 74–85. [PubMed: 23994296]
- Tanabe H, Sawada T, Inoue N, Ogawa M, Kuriyama Y, & Shiraishi J (1987). Conduction aphasia and arcuate fasciculus. *Acta Neurologica Scandinavica*, 76, 422–427. [PubMed: 3434200]
- Taylor IK, & Taylor MM (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Experimental Psychology*, 16, 344–356.
- Tourville JA, & Guenther FH (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26, 952–981. [PubMed: 23667281]

- Tylor EB (1868). On the origin of language. *Fortnightly Review*, 1, 22.
- Ueno T, & Lambon Ralph M. A. (2013). The roles of the “ventral” semantic and “dorsal” pathways in conduite d’approche: A neuroanatomically-constrained computational modeling investigation. *Frontiers in Human Neuroscience*, 7, 422. [PubMed: 23986670]
- Van Essen DC (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*, 28, 635–662. [PubMed: 16172003]
- Ventura MI, Nagarajan SS, & Houde JF (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10, 58. [PubMed: 19523234]
- Vogt S, Buccino G, Wohlschlager AM, Canessa N, Shah NJ, Zilles K, ... Fink, G. R. (2007). Prefrontal involvement in imitation learning of hand actions: Effects of practice and expertise. *Neuroimage*, 37, 1371–1383. [PubMed: 17698372]
- Wager TD, & Smith EE (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 255–274.
- Wallace A (1895). Expressiveness of speech, or mouth gesture as a factor in the origin of language. *Fortnightly Review*, 64, 528–543.
- Webster PJ, Skipper-Kallal LM, Frum CA, Still HN, Ward BD, & Lewis JW (2017). Divergent human cortical regions for processing distinct acoustic-semantic categories of natural sounds: Animal action sounds vs. vocalizations. *Frontiers in Neuroscience*, 10, 579. [PubMed: 28111538]
- Weiss JH (1964). Phonetic symbolism re-examined. *Psychological Bulletin*, 61, 454–458. [PubMed: 14183608]
- Wise RJ, Greene J, Buchel C, & Scott SK (1999). Brain regions involved in articulation. *Lancet*, 353, 1057–1061. [PubMed: 10199354]
- Yoshida H (2012). A cross-linguistic study of sound-symbolism in children’s verb learning. *Journal of Cognition and Development*, 13, 232–265. [PubMed: 23807870]
- Zatorre RJ, & Belin P (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953. [PubMed: 11549617]
- Zatorre RJ, Bouffard M, & Belin P (2004). Sensitivity to auditory object features in human temporal neocortex. *The Journal of Neuroscience*, 24, 3637–3642. [PubMed: 15071112]
- Zatorre RJ, Evans AC, Meyer E, & Gjedde A (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846–849. [PubMed: 1589767]
- Zhang Y, Wang C, Zhao X, Chen H, Han Z, & Wang Y (2010). Diffusion tensor imaging depicting damage to the arcuate fasciculus in patients with conduction aphasia: A study of the Wernicke-Geschwind model. *Neurological Research*, 32, 775–778. [PubMed: 19825277]

Statement of significance

The results revealed distinct brain regions preferentially involved in the hearing and oral mimicry of animal action sounds versus animal vocalizations. These results may reflect vestiges of rudimentary oral communication networks that support anthropological models of glottogenesis, and advance clinically-relevant models of spoken language neurodevelopment and of recovery after stroke.

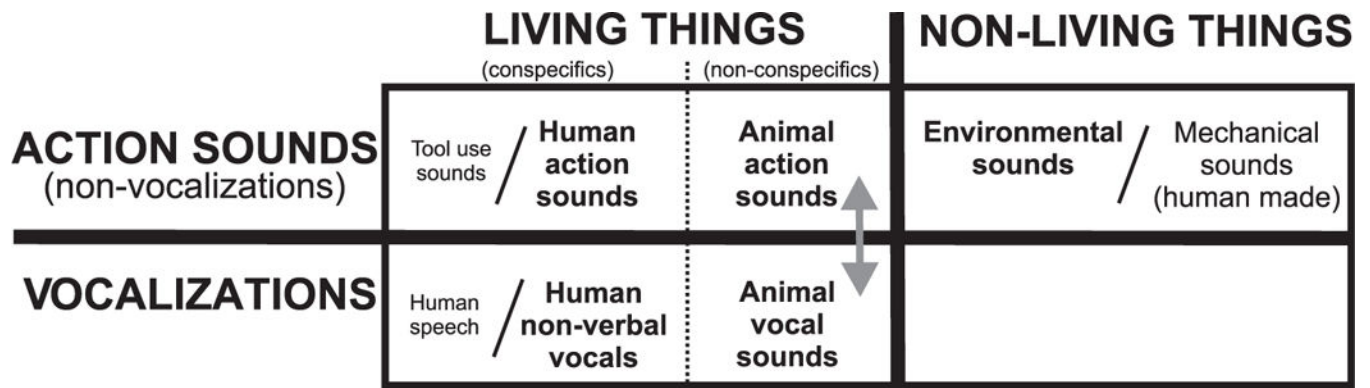


Fig. 1.

A neurobiological model of the organization of the human brain for processing and recognizing different acoustic-semantic categories of natural sounds [from Breczynski-Lewis and Lewis (2017)]. Bold text in the boxed regions depict rudimentary sound categories proposed to represent ethologically relevant categories germane to sound recognition for all mammalian species. Human speech, tool use sounds, and human-made machinery sounds are represented as extensions of these categories. Vocal and instrumental music sounds are regarded as higher forms of communication, which rely on other networks. The present study is testing the putative functional boundary (double headed arrow) of cortical networks for mimicking action sounds versus mimicking vocalizations using animal (non-conspecific) sound stimuli. Refer to text for other details.

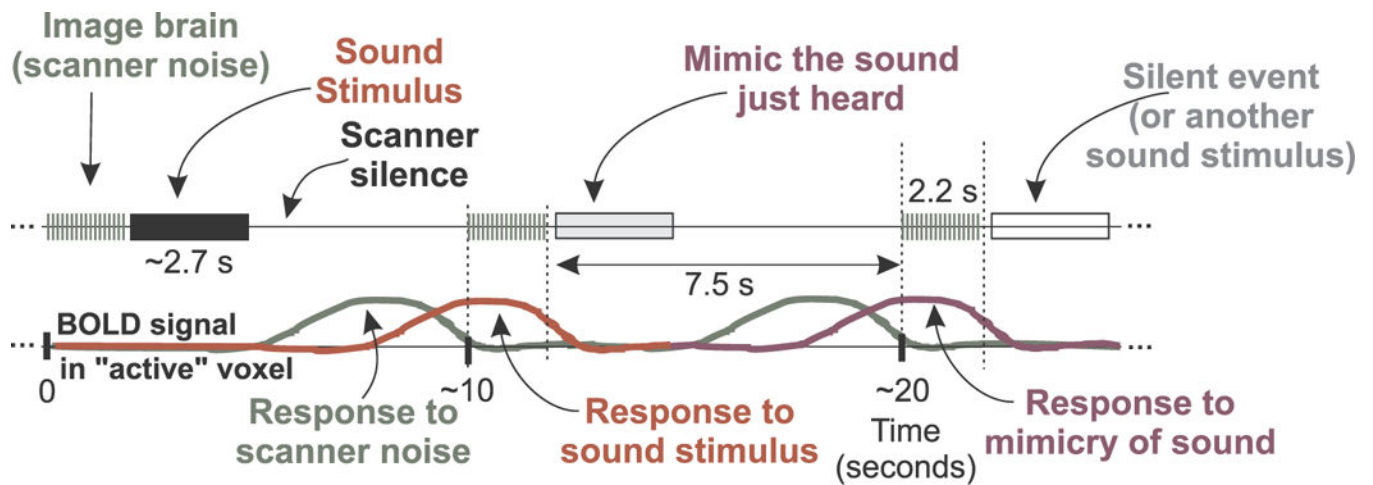


Fig. 2.

Clustered acquisition fMRI imaging design. The animal action sound events, vocalization sound events, and silent events were presented in a pseudo-random order. However, each sound event was followed by a ‘silent period’ wherein the participant mimicked the sound they had just heard, as depicted. Stimulus and mimicry events were triggered every 10 s plus the time until the participant’s next cardiac cycle (R-wave). Refer to text for other details.

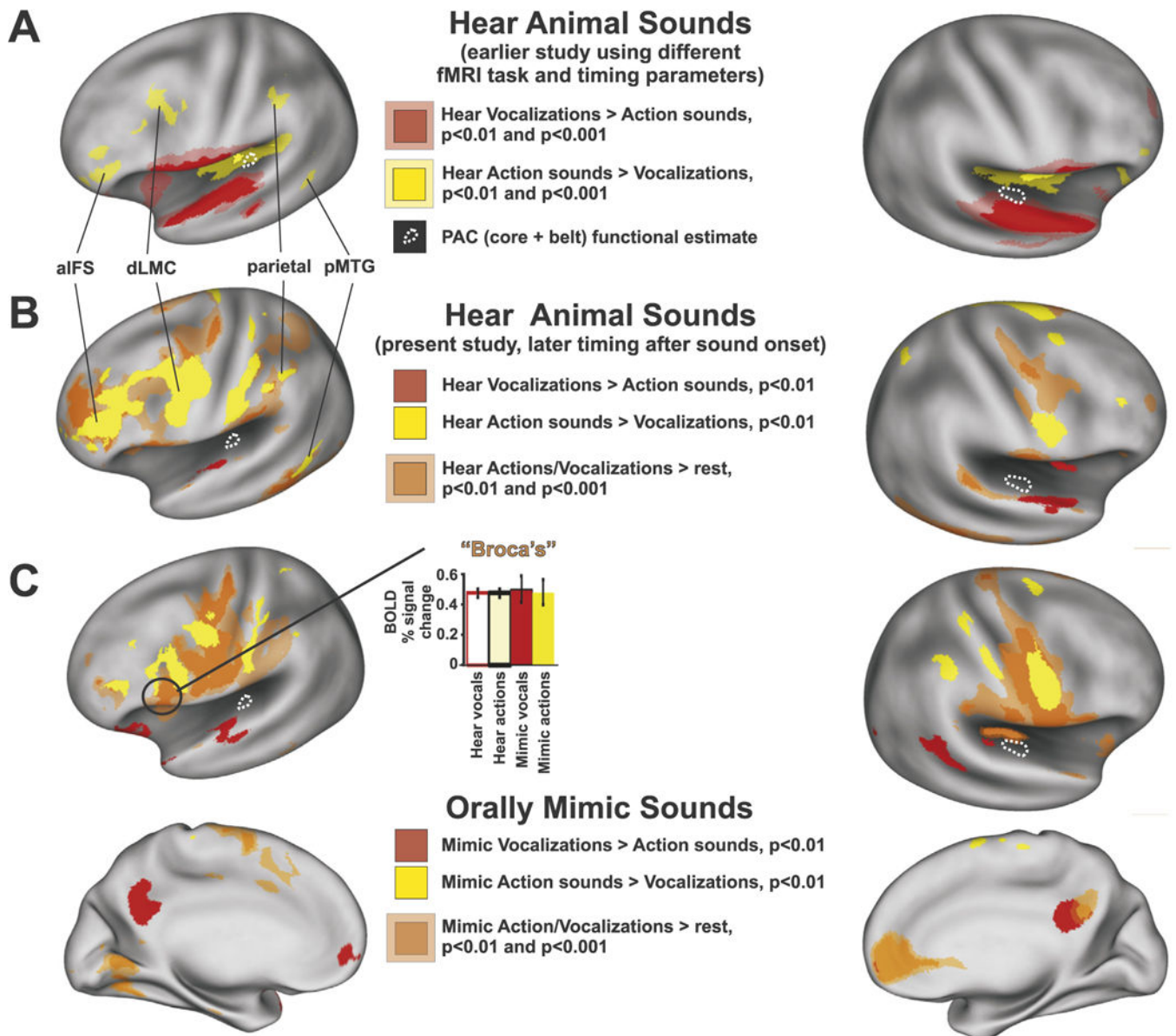


Fig. 3. Cortical networks preferentially activated when (A–B) hearing animal action sounds versus animal vocalizations and when (C) orally mimicking those corresponding sound stimuli. Whited dotted outlines depict functional estimates of core and belt auditory cortices based on the localizer scan. (A) Data from an earlier study with timing parameters optimized for revealing intermediate auditory cortices for processing animal action sounds (yellow, $p_{\text{corr}} < 0.001$; pale yellow, $p_{\text{corr}} < 0.01$) versus animal vocalizations (red, $p_{\text{corr}} < 0.001$; transparent red, $p_{\text{corr}} < 0.01$), illustrated on inflated cortical surface models of the PALS atlas, adapted and reprinted with permission by the publisher. (B) Group-averaged fMRI results ($n = 16$) from the present study preferential for hearing animal actions versus vocalizations, and for (C) orally mimicking those same sounds by category (refer to color keys for corrected threshold settings). Histogram indicates the BOLD percent signal change (average \pm SEM)

in response to each category of sound and to oral mimicry of those corresponding sounds. Refer to text for other details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

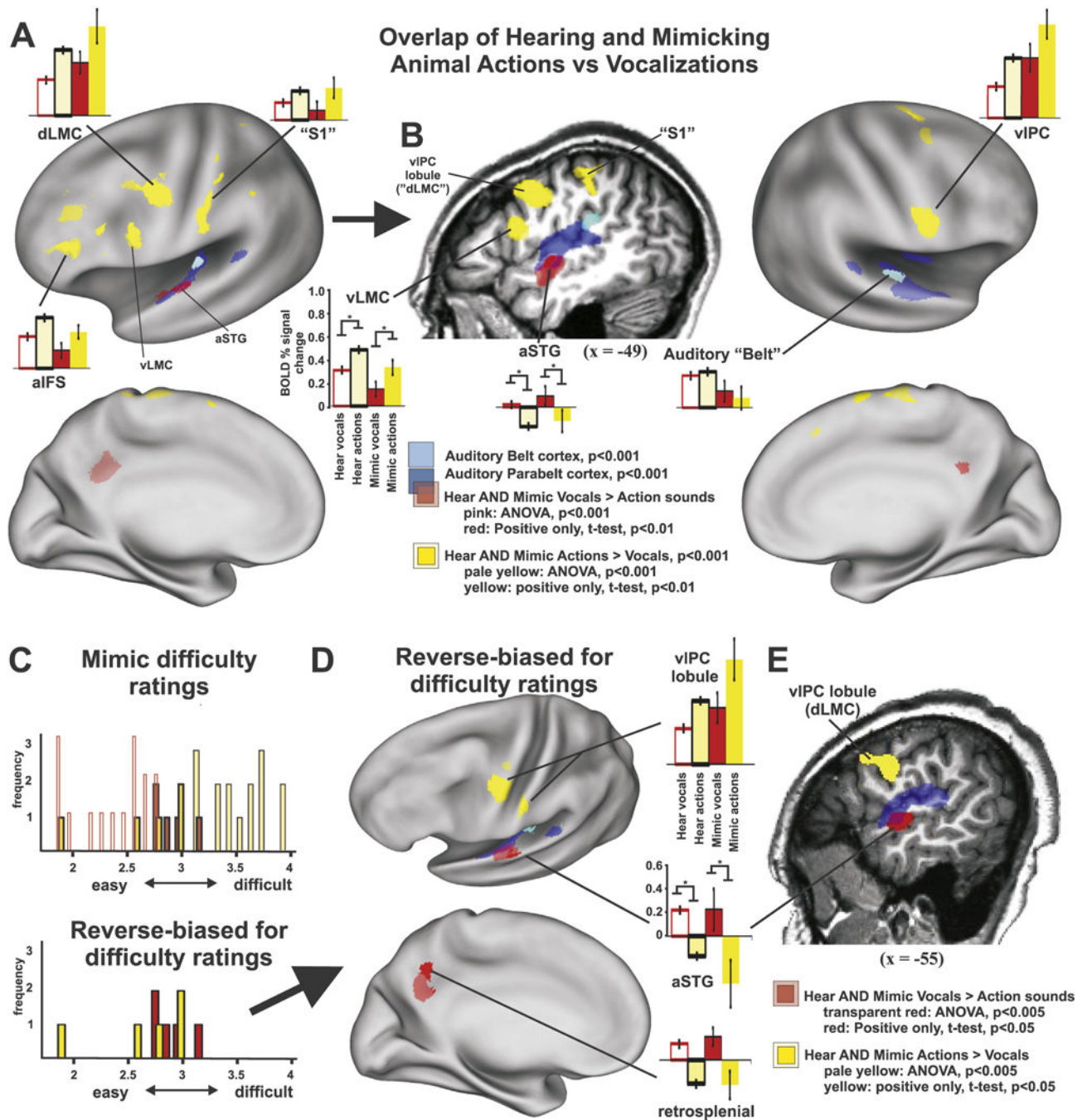


Fig. 4. Group-averaged activation maps resulting from ANOVA and *t*-test analyses revealing regions preferential for both perception of a given category of sound and for oral mimicry of that same category. (A–B) Foci derived from analyses including the 36 retained stimulus event types (from Table 1) showing maps of category- preferential foci relative to functionally derived auditory belt (light blue) and parabelt cortices (dark blue), defined using a separate localizer scan using English phonemes. Histograms illustrate the BOLD percent signal change (average \pm SEM) for various regions of interest in response to each category

of sound and to oral mimicry of those corresponding sounds, both relative to averaged responses to silent events. (C) Charts illustrating subject ratings of perceived difficulty for mimicking each sound stimulus. (D–E) Maps showing preferential processing to hearing and oral mimicry using the same analysis techniques but using only a subset of the sounds (panel C) that were reverse-biased in perceived difficulty to mimic. LMC = laryngeal motor cortex (estimated; overlapping with vlPC); aSTG = anterior superior temporal gyrus; S1 = primary somatosensory cortex (estimated); vlPC = ventro-lateral paracentral lobule. Refer to text for other details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

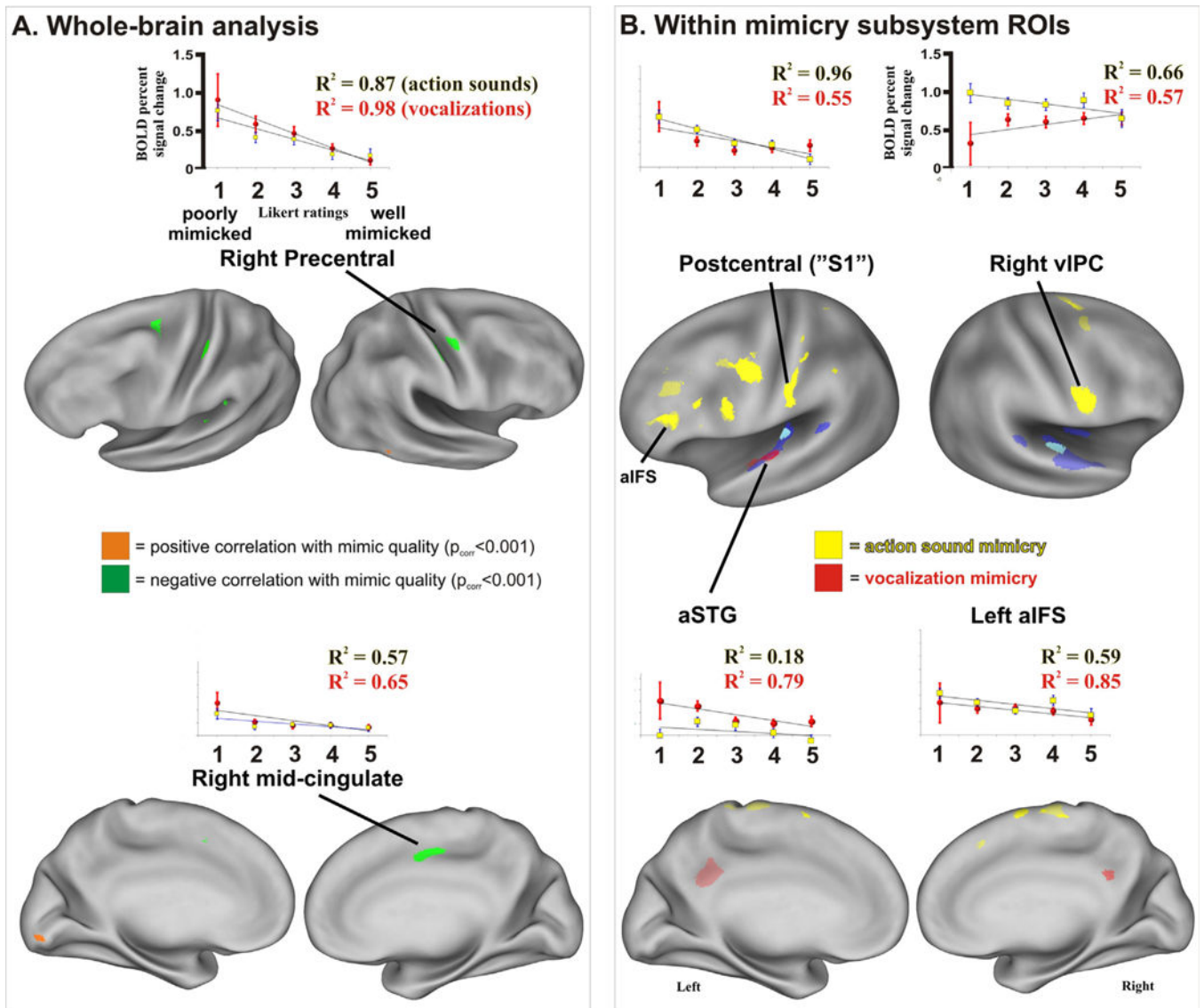


Fig. 5. Brain regions showing parametric sensitivity to the participant's ratings of how well they thought they orally mimicked each sound, by category ($n = 14$ of 16 participants). (A) Whole-brain primary level analysis showing the most strongly activated regions that were parametrically correlated with perceived mimic quality (see color key for thresholds). Poorer mimicry was generally associated with greater activation. The right precentral gyrus focus (Talairach $x = 45, y = -14, z = 52; 835 \text{ mm}^3$) and right middle cingulate ($x = 11, y = -10, z = 42; 1090 \text{ mm}^3$) showed the strongest degree of linear correlation between perceived mimic quality and BOLD signal brain responses. (B) Several group-averaged ROIs from Fig. 4 also showed significant parametric activation correlated with perceived mimic quality, and with some areas showing dependence on the category of sound. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

List of animal sound stimuli heard and orally mimicked during fMRI scanning, including duration (seconds), intensity (RMS power, dB), and several higher-order acoustic signal attributes of those sounds relative to human mimicked versions of those corresponding sounds (average of n = 14 participants). HNR = harmonics-to-noise ratio (dB_{HNR}); WE = Wiener entropy; SSV = spectral structure variation. Bolded entries depict sounds retained for the difficulty-matched mimicry analysis of Fig. 4C–E. Emotion = average ratings of emotional valence. The fMRI participants rated how “Difficult” each sound was to mimic, and rated “How Well” they felt they mimicked each sound during scanning procedures. Refer to Methods for other details.

Description	Action sounds			Human mimicks of animal sounds			How Well						
	Duration	Intensity	HNR	WE	SSV	Emotion		Duration	Intensity	HNR	WE	SSV	Difficulty
Action sounds													
dog drinking water from bowl	2.52	-17.20	1.27	-4.28	2.12	0.80	2.80	-37.32	1.79	-2.08	1.12	3.64	3.07
dog footsteps	2.67	-18.27	-0.68	-1.94	0.38	0.33	2.05	-38.69	1.40	-2.12	0.80	3.38	2.79
dog licking/lapping #2	2.89	-17.65	-1.52	-2.49	1.53	0.07	2.34	-43.67	1.39	-2.31	0.70	3.71	2.79
dog panting heavily	2.80	-17.80	-0.463	-4.418	0.62	1.33	2.22	-35.65	0.50	-2.19	0.52	1.86	4.07
dog snuffle and digging	3.03	-17.03	-1.01	-2.96	0.80	0.60	2.17	-38.48	1.16	-2.40	0.78	2.54	3.71
dog trotting, nails on floor #1	2.86	-18.67	-0.64	-2.53	0.41	0.00	2.10	-33.06	2.11	-2.03	0.93	4.00	2.50
dog trotting, toe nails on floor	2.68	-18.05	-1.24	-3.59	0.73	0.07	2.11	-35.95	1.44	-1.93	0.72	3.79	2.50
dog, large eating from bowl	2.84	-17.46	-1.71	-3.21	1.81	0.40	2.39	-40.52	1.07	-2.16	0.79	3.14	2.36
dog/animal, licking & chewing	2.55	-17.48	-1.40	-2.54	1.52	0.53	2.06	-33.94	1.52	-2.46	0.71	3.14	2.86
horse galloping #1	2.70	-17.60	-1.65	-6.17	1.91	0.67	3.09	-30.88	1.16	-2.17	0.82	3.57	2.50
horse galloping #2	2.75	-17.25	3.16	-2.97	2.03	0.27	2.71	-32.57	1.48	-2.08	0.84	3.43	2.93
horse galloping #3	2.70	-17.45	1.52	-2.60	1.72	0.33	2.19	-34.15	1.59	-2.12	0.86	3.69	2.86
horse galloping #4	2.85	-17.21	3.88	-3.28	2.28	0.40	2.58	-31.98	1.76	-2.20	0.82	3.36	2.71
horse trotting	2.60	-16.09	2.80	-3.38	2.19	0.60	2.27	-34.91	1.66	-2.00	0.98	3.00	3.14
horse, bridaled walking	2.77	-17.98	2.33	-2.87	0.80	0.53	2.16	-35.37	0.98	-1.95	0.77	3.43	3.14
horse, on cobblestone	2.77	-17.47	1.14	-4.88	1.77	0.00	2.75	-33.40	1.15	-1.94	0.83	3.00	3.21
horse, walking	2.68	-17.94	7.72	-6.02	2.95	-0.07	1.88	-37.79	1.74	-1.93	0.88	2.71	3.14
zebra, trotting on dirt	2.85	-17.77	3.59	-3.21	2.54	0.13	2.37	-33.60	1.62	-2.11	0.75	4.00	2.50
Average	2.75	-17.58	0.95	-3.52	1.56	0.39	2.35	-35.66	1.42	-2.12	0.81	3.30	2.93
StDev	0.13	0.55	2.58	1.19	0.77	0.35	0.32	3.26	0.37	0.15	0.13	0.54	0.44

Description	Action sounds			Human mimicks of animal sounds			SSV	WE	HNR	SSV	Difficulty	How Well
	Duration	Intensity	HNR	WE	HNR	SSV						
Vocalizations												
baby bear bark	2.72	-18.39	-1.43	-8.64	1.509	-1.00	1.84	-28.36	10.28	2.241	2.29	3.50
bear call #1	2.84	-17.97	1.64	-8.89	2.876	-0.73	2.73	-26.64	11.79	2.069	2.71	3.79
bear call #2	2.74	-17.36	4.50	-7.96	2.946	-0.93	2.25	-29.82	10.85	2.028	2.71	3.36
bobcat growl	2.74	-17.88	-1.14	-6.47	1.338	-1.40	2.62	-30.99	9.13	1.643	2.36	3.57
bull call #1	2.93	-17.40	3.84	-6.90	4.289	-0.93	2.10	-34.10	14.49	0.819	1.86	4.21
bull call #2	2.85	-18.20	11.18	-9.86	3.248	-0.47	1.93	-29.16	15.45	0.776	2.00	3.71
camel moo	2.70	-17.73	13.01	-10.00	0.778	-0.87	2.56	-29.45	12.29	0.63	2.64	3.14
cougar angry	2.58	-18.16	1.83	-7.41	3.709	-1.73	2.00	-30.87	6.88	1.274	2.85	3.29
dog bark	2.57	-17.69	4.03	-7.51	3.207	-0.20	2.07	-23.49	9.31	3.729	2.14	3.79
dog whine and bark	2.60	-17.77	6.27	-7.44	2.412	-1.33	1.96	-30.00	11.03	2.688	2.57	3.00
dog growl-sigh	2.50	-17.81	4.43	-7.01	3.871	-0.87	1.85	-30.18	9.31	1.932	2.57	3.14
horse winney	2.59	-17.13	8.64	-6.14	1.033	-1.07	1.99	-32.76	12.51	1.096	2.93	3.21
lion roar in distance	2.85	-17.29	1.69	-10.82	2.304	-1.00	2.04	-32.88	6.97	1.181	2.50	3.14
monkey, angry call	2.20	-17.55	14.96	-6.20	4.394	-0.93	2.15	-20.61	13.69	2.963	2.57	3.36
monkey, vocal	2.75	-17.35	14.07	-7.37	6.758	1.00	2.29	-28.00	15.06	3.619	3.14	3.43
pig squeal	2.80	-18.19	6.70	-5.25	0.691	-1.73	2.15	-27.75	11.72	1.975	1.86	3.00
sheep baa	2.75	-17.35	8.46	-5.70	16.66	0.60	2.18	-26.13	5.04	1.404	2.79	3.64
wolf growl	2.59	-17.12	1.13	-5.80	0.61	-1.47	2.48	-43.38	5.17	-3.879	2.79	3.29
Average	2.68	-17.69	5.77	-7.52	3.48	-0.84	2.18	-29.70	10.61	1.85	2.52	3.42
StDev	0.17	0.39	5.04	1.59	3.66	0.72	0.26	4.74	3.17	0.93	0.37	0.32

Group activation centroids in Talairach coordinate space for *t*-test cortical ROI foci from Fig. 4A–B data at $p_{\text{corr}} < 0.01$.

Table 2

Anatomical location	Talairach coordinates			volume (mm ³)
	x	y	z	
Action sound mimicry subsystem				
Left aIFC	-38	30	21	1726
Left aIFG	-47	8	21	860
Left precentral g. (LMC)	-51	-4	37	4295
Left somatosensory cortex	-52	-28	45	676
Right precentral g. (LMC)	56	-1	30	1542
Right SMA (cluster)	7	-7	65	1223
Vocalization mimicry subsystem				
Left mST G	-53	-11	4	233

Group activation centroids in Talairach coordinate space for *t*-test cortical ROI foci from Fig. 4D–E data at $p_{\text{corr}} < 0.01$.

Table 3

Table 3 (Anova)	Talairach coordinates			volume (mm ³)
	x	y	z	
Anatomical location				
Action sound mimicry subsystem				
Left precentral g. (LMC)	-51	-4	33	2441
Vocalization mimicry subsystem				
Left retrosplenial	-3	-52	-10	1643
Left aSTG	-60	-12	1	2475