Divergent Human Cortical Regions for Processing Distinct Acoustic-Semantic Categories of Natural Sounds: Animal Action Sounds vs. Vocalizations

Paula J. Webster[1], Laura M. Skipper-Kallal[1,2], Chris A. Frum[3], Hayley N. Still[1], B. Douglas Ward[4] and James W. Lewis[1]*

[1] Blanchette Rockefeller Neurosciences Institute, Department of Neurobiology & Anatomy, West Virginia University, Morgantown, WV, USA, [2] Department of Neurology, Georgetown University Medical Campus, Washington, DC, USA, [3] Department of Physiology and Pharmacology, West Virginia University, Morgantown, WV, USA, [4] Department of Biophysics, Medical College of Wisconsin, Milwaukee, WI, USA

A major gap in our understanding of natural sound processing is knowledge of where or how in a cortical hierarchy differential processing leads to categorical perception at a semantic level. Here, using functional magnetic resonance imaging (fMRI) we sought to determine if and where cortical pathways in humans might diverge for processing action sounds vs. vocalizations as distinct acoustic-semantic categories of real-world sound when matched for duration and intensity. This was tested by using relatively less semantically complex natural sounds produced by non-conspecific animals rather than humans. Our results revealed a striking double-dissociation of activated networks bilaterally. This included a previously well described pathway preferential for processing vocalization signals directed laterally from functionally defined primary auditory cortices to the anterior superior temporal gyri, and a less well-described pathway preferential for processing animal action sounds directed medially to the posterior insulae. We additionally found that some of these regions and associated cortical networks showed parametric sensitivity to high-order quantifiable acoustic signal attributes and/or to perceptual features of the natural stimuli, such as the degree of perceived recognition or intentional understanding. Overall, these results supported a neurobiological theoretical framework for how the mammalian brain may be fundamentally organized to process acoustically and acoustic-semantically distinct categories of ethologically valid, real-world sounds.

Keywords: auditory system pathways, fMRI, mirror neuron system, acoustic communication, grounded cognition

INTRODUCTION

Critical to survival and social organization is our ability to view, hear, and understand the goals and intentions of others, including both human (conspecifics) and non-human (non-conspecific) animals. However, it remains unclear how the human auditory system is fundamentally organized to process the diverse range of natural, biologically relevant sounds, such as speech and action events, and provide the listener with a sense of meaning. In dual-stream models of sound processing
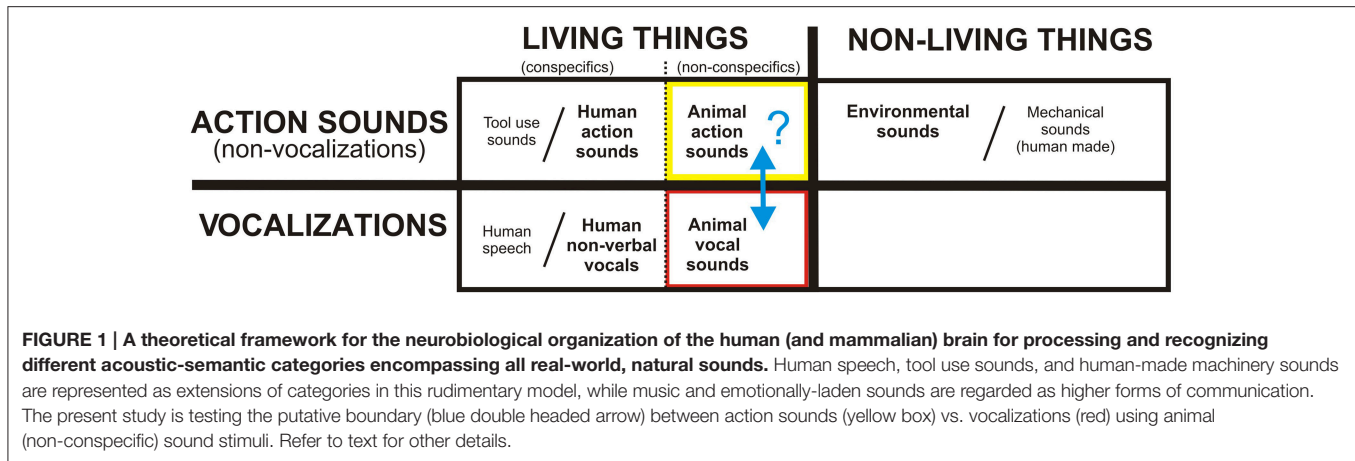
dorsally directed cortical pathways, relative to primary auditory cortex (PAC), are involved in part with processing spatial information, such as localizing sounds and preparing motor movements or action schemas to engage or avoid the sound-source (Romanski et al., 1999; Rauschecker and Tian, 2000; Rauschecker and Scott, 2015). Conversely, ventrally directed cortical pathways, relative to PAC, are involved more in identification of sound patterns for perception of what the source is or its potential communicative content. This dorsal vs. ventral dichotomy has also been applied to models of spoken language processing (MacNeilage, 1998; Rauschecker and Scott, 2009; Arbib, 2010; Perlovsky, 2011; DeWitt and Rauschecker, 2013). In these models, dorsal pathways are generally implicated in temporal processing, such as amplitude or temporal envelope segmentation. In contrast, ventral pathways are involved in pattern matching for content, such as syllables, phonemes, and words. However, the processing that takes place along intermediate stages, between PACs and the higher-level semantic and audio-motor representations that are more commonly affiliated with recognition and behavior remains unclear. Thus, while hypothesized dorsal-ventral organizations for sound processing are reasonably well established, many gaps remain in our understanding of how the putative ventral pathways for sound recognition ("what is it") may be organized to process the acoustic signal attributes that may be characteristic of different semantic or "acoustic-semantic" categories of natural sounds.

With regard to "bottom-up" theories of sound processing and perception, numerous animal models, and more recently human models, have revealed principles for how real-world sounds become processed subsequent to tonotopically organized regions (Rauschecker et al., 1995; Kaas and Hackett, 1998; Petkov et al., 2008; Lewis et al., 2009; Giordano et al., 2013). In particular, hierarchical models indicate that parallel processing pathways along auditory cortices have successive stages containing neurons that show increasing selectivity for complex spectro-temporal attributes (Fecteau et al., 2004; Medvedev and Kanwal, 2004; Kumar et al., 2007; Rauschecker and Scott, 2009; Schönwiesner and Zatorre, 2009; Lewis et al., 2012). Such organizations are suggestive of "matched-filter" tuning mechanisms, which may facilitate or mediate the detection of certain types of sounds or sound attributes that effectively segment sounds of interest from background noise (Suga, 1965; Pollak and Bodenhamer, 1981; Rose and Capranica, 1983; Lewicki, 2002; Woolley et al., 2005). However, there is no consensus as to what low- or high-level acoustic signal attributes comprise the basic building blocks for the processing and reconstruction of real-world, natural sounds as meaningful events to a listener.

Conversely, top-down driven models, such as embodied or grounded cognition models, have led to the development of theories of how various widespread brain regions may mediate semantic knowledge representations at a categorical level (Moore and Price, 1999; Barsalou et al., 2003; Miller et al., 2003; Hickok and Poeppel, 2004; Martin, 2007; Binder et al., 2009; Engel et al., 2009; Lewis et al., 2011; Bornkessel-Schlesewsky et al., 2015). However, many earlier human neuroimaging studies using human-produced sound stimuli, which in adult listeners

tend to inherently entail complex semantic associations, likely recruited brain regions well outside of auditory cortex *proper* for processing higher order cognitive information, including linguistic features, subtle communicative nuances, and audio-motor schemas (Shallice, 1988; Grafton et al., 1997; Binder et al., 2000; Fiebach et al., 2003; Alain, 2007; Canessa et al., 2007; McNamara et al., 2008; Turkeltaub and Coslett, 2010; Meyer et al., 2011; Woods et al., 2011; Beer et al., 2013). Consequently, the cortical processing of highly familiar human-produced sounds may mask some of the more fundamental, or intermediate, stages of real-world, natural sound processing.

In the present study, we sought to further bridge bottom-up and top-down models of fundamental aspects of sound recognition by examining functional networks for processing two categories of semantically (and acoustically) distinct real-world sounds produced by non-human animals: Action sounds vs. vocalizations. Such sounds would arguably be ethologically valid, yet have fewer complex semantic associations than human-produced sounds in general. More specifically, based on a large number of earlier neuropsychological and neuroimaging studies of hearing perception, we sought to test a general theoretical framework for how the mammalian brain may be, or may become, optimized for processing real-world sounds at a semantic or acoustic-semantic categorical level (**Figure 1**). This simple model incorporates the classical living vs. non-living semantic category boundary for word-form knowledge as one major division (Warrington and Shallice, 1984; Hillis and Caramazza, 1991; Capitani et al., 2003) but here being applied more specifically to sound-sources. A second major theorized division, being explicitly tested in the present study, distinguishes vocalizations from action sounds (non-vocalizations). Earlier studies of speech and human vocalization processing have identified the superior temporal gyri (STG) and surrounding territories as principle "intermediate" cortical processing stages for vocal communication sounds (Belin et al., 2000; Turkeltaub and Coslett, 2010), including a left lateralized extended network for intelligible speech sounds (Ohnaka et al., 1995; Scott et al., 2000, 2006; Engelien et al., 2006; Friederici et al., 2010) and right lateralized extended network preferential for processing non-verbal and prosodic communication sounds (Corballis, 1989; Zatorre et al., 1992; Thierry et al., 2003; Gandour et al., 2004). The semantic boundary for processing vocalizations produced by animals vs. humans (i.e., non-conspecifics vs. conspecifics) has received relatively less attention (**Figure 1**, black dotted line next to red box). Nonetheless, at least a subset of the above mentioned STG foci and surrounding cortical regions are reported to be recruited when human listeners process animal (non-conspecific) vocalizations (Maeder et al., 2001; Lewis et al., 2005, 2009; Engelien et al., 2006; Altmann et al., 2007; Belin et al., 2008; Doehrmann et al., 2008; Talkington et al., 2012, 2013). This conspecific vocalization processing boundary concept is generally consistent with neuroimaging studies of other mammalian species, including monkeys (Kohler et al., 2002; Petkov et al., 2008) and canines (Andics et al., 2014), which reveal regions or pathways preferential for processing their respective conspecific vocalizations and calls. Consequently, for the present study we regarded human vocalizations as belonging to a

FIGURE 1 | A theoretical framework for the neurobiological organization of the human (and mammalian) brain for processing and recognizing different acoustic-semantic categories encompassing all real-world, natural sounds. Human speech, tool use sounds, and human-made machinery sounds are represented as extensions of categories in this rudimentary model, while music and emotionally-laden sounds are regarded as higher forms of communication. The present study is testing the putative boundary (blue double headed arrow) between action sounds (yellow box) vs. vocalizations (red) using animal (non-conspecific) sound stimuli. Refer to text for other details.

distinct acoustic-semantic category, which potentially facilitates recruitment of down-stream cortical pathways for additional processing of communicative intent or other complex learned semantic knowledge associations. Therefore, human produced vocalization sounds were not included in order to target more rudimentary or intermediate stages of sound processing.

With regard to action (non-vocalization) sounds, we previously reported the existence of a four-fold dissociation of cortical networks for processing different categories of real-world sounds, distinguishing those produced by human, animal, mechanical, and environmental sources (Engel et al., 2009; Lewis et al., 2011). These studies revealed the left and right posterior insulae as regions preferential for processing non-human animal action sounds. Thus, also included in our theoretical framework (**Figure 1**, black dotted line next to yellow box) is a semantic processing boundary between human vs. animal action sounds. On this front, our model incorporates grounded ("embodied") cognition theories, which at least loosely relate to mirror neuron systems (MNS), which may reflect system-level processing strategies for matching heard action sounds to the listeners own repertoire of sound-producing motor actions; this processing strategy is thought to provide a listener with a sense of meaning or intention behind the sound-producing action (MacNeilage, 1998; Arbib, 2001; Buccino et al., 2001; Barsalou et al., 2003; Rizzolatti and Craighero, 2004; Bidet-Caulet et al., 2005; Calvo-Merino et al., 2005; Gazzola et al., 2006; Lewis et al., 2006; Lahav et al., 2007; Martin, 2007; Mutschler et al., 2007; Barsalou, 2008; Galati et al., 2008; Giordano et al., 2010; Imai et al., 2015). Most studies examining biological motion (visual or auditory) have focused on using human (conspecific) stimuli (for review, see, Chouchourelou et al., 2012). However, one study using point-light animations of both human and animal biological motion, together with brain lesion mapping, reported cases of selective impairment in recognizing human motion but not animal motion (Han et al., 2013), thereby distinguishing these semantic action-event categories. Ostensibly, when processing biological sounds that cannot be fully or accurately mimicked motorically by one's own body or limbs (e.g., a horse galloping or pigeon flapping its wings) the brain must

still attempt to process and decode the acoustic signals for potential meaning behind the heard action. Thus, in the present study we also regarded human vs. animal action sounds as belonging to semantically and neurobiologically distinct categories.

At a gross level, our model of acoustic-semantic category-level sound processing makes a strong prediction for the existence of a major divergence in cortical pathways for processing animal action sounds vs. animal vocalizations (**Figure 1**, blue arrow). As alluded to earlier, higher level cognitive systems tend to show greater hemispheric lateralization for information processing (Preuss, 2011), with some dependence on task and developmental plasticity (Hamzei et al., 2003; Iacoboni et al., 2005; Lewis et al., 2006). This lateralization effect is especially prominent with processing associated with language, vocal and instrumental music, and other complex semantic-level processing that entails potentially higher forms of acoustic communication or thought (Peretz and Zatorre, 2005; Binder et al., 2009; Meyer et al., 2011; Alluri et al., 2013; Cai et al., 2013). Hence, our presumed rudimental model for the acoustic-semantic organization of sound processing further predicts that the processing of non-conspecific animal sounds should recruit divergent pathway activation in *both* cortical hemispheres that likely relate to relatively less specialized processing weighted by bottom-up acoustic signal processing. Thus, the goal of the present study, using fMRI, was to test for the existence of divergent cortical processing streams along human auditory cortices, presumably in both hemispheres, that respect the theorized acoustic-semantic boundary for processing *non-conspecific animal* action sounds vs. vocalizations as ideal stimulus sets to critically test this putative major boundary—which may represent a sensory processing boundary common to most if not all mammals. Exploring and establishing fundamental cortical organizational principles behind natural sound processing is ultimately expected to help advance both anthropological theories of spoken communication evolution, as well as neurodevelopmental models of hearing perception and spoken language acquisition in healthy young children and those with speech or language delays.

MATERIALS AND METHODS

Participants

We tested 17 right-handed participants (Avg. 23.6 ± 2.6 years, range 20–29, 8 female) who performed one or multiple paradigms using functional magnetic resonance imaging (fMRI). All participants were native English speakers with no history of neurological or psychiatric disorders, and a self-reported normal range of hearing with no auditory impairments. Informed consent was obtained for all participants following guidelines approved by the West Virginia University Institutional Review Board and in compliance with the Code of Ethics of the World Medical Association.

Sound Stimulus Creation and Presentation

The sound stimulus set consisted of 29 animal vocalizations and 29 animal action sounds (**Table 1**; also see art piece in Supplemental Figure S4) derived from professional compilations (Sound Ideas, Inc., Richmond Hill, Ontario, Canada), incorporating recordings of animals in isolation with little to no background noise. The animal vocalizations selected were produced by vibrations of the vocal folds, primarily of various land mammals. To facilitate data interpretation, we minimized the use of bird calls and aquatic mammals, due to differences in their mechanisms of sound production via their respective "vocal" systems (e.g., syrinx or blowhole), as a listener's ability to mimic sounds may influence perceptual processing mechanisms (Liberman and Mattingly, 1985). Animal vocalizations were initially derived from another study by our lab examining mimicry (Talkington et al., 2012) in which at least 46 of 50 participants could correctly categorize the sound as non-human (in contrast to animal vocalizations mimicked by a human; two alternative-forced choice [2AFC] task). Animal action sounds were also selected based upon whether they were produced primarily by land animals and were deemed to be devoid of any vocalization content. This set of sounds was initially derived from our earlier study (Engel et al., 2009), which included action sound sources that were categorized as animal (vs. human, mechanical, or environmental; four alternative forced choice [4AFC]) by more than 80% of participants.

Using commercially available software (Adobe Audition 3.0, Adobe Systems Inc.), the two categories of sound stimuli were carefully matched for duration (2.7 ± 0.2 s) and total root mean squared power (−17.6 ± 0.5 dB), which remained matched both before and after *post-hoc* censoring six of the action sounds from data analyses (addressed below). In addition, the onsets/offsets were ramped by 25 ms to avoid "pops" during sound presentation. Sound stimuli were converted to one channel (mono, 44.1 kHz, 16-bit) but presented to both ears via ear buds, thereby removing any binaural spatial cues.

To validate the specificity and clarity of the sound source category, five participants who were not included in the fMRI studies, and were naïve to the purpose of the experiment, assessed the censored sound stimuli to (1) verify that the animal vocalizations and animal action sounds were easily categorized as an action or vocalization, and (2) verify that sounds were clearly perceived as <u>not</u> being produced by a human agent. Sounds that were unclear or retained ambiguous elements that might cross the theorized category boundary/boundaries were replaced by other examples until a complete set of 58 sound stimuli was obtained that all five listeners could unanimously categorize correctly. After obtaining post-scanning feedback from several fMRI participants, we determined that some of the action sounds were perceived as possibly containing overt communicative intent (e.g., rattlesnake rattling to indicate "stay away"), human intervention (e.g., hearing metallic sounds of a dog collar jingling), or were "breathy" (e.g., panting) some participants confused as blending across the conceptual boundary between action events and vocalizations. Consequently, we censored out fMRI responses (blood oxygen-level dependent; BOLD) activation brain responses to six of the animal action sounds *post-hoc* (**Table 1**, italicized entries). Note, that including or excluding these sound stimuli did not qualitatively affect the main results (data not shown), thought these exclusions facilitated data interpretation.

Acoustic Signal Features of the Sound Stimuli

The sound stimuli were assessed for differences in quantifiable high-order acoustic signal attributes and fine scale power spectra using freely available phonetic software (Praat, http://www.fon.hum.uva.nl/praat/, version 5.1.04). The animal action sounds generally had greater power at the lower frequency ranges (long term average spectrum at 10 Hz bandwidth increments). This included a significantly greater prevalence of amplitude fluctuations at rates between 10 and 50 Hz for the animal action sounds [single factor ANOVA $F_{(1, 9)} = 473$, $p < 10^{-8}$; see **Supplemental Figure S1A**], which is a range typically regarded as envelope information (2–50 Hz; also known as "amplitude envelope" or "temporal information") in speech literature (Rosen, 1992). Some of this difference in power at lower frequencies likely reflected temporal envelope events, such as the pace of locomotion sounds (e.g., footsteps and galloping) and temporal harmonics therein. Several low- and high-order acoustic signal attributes could distinguish the two semantic categories of sound on average (see **Table 1**). For instance, the global harmonics-to-noise ratio (HNR) value was determined for each sound using Praat software (10 ms time step; minimum pitch cutoff of 75 Hz, 20 kHz ceiling; and 1 period per window) and on average revealed a significantly greater degree of harmonic structure for the animal vocalizations (animal vocalizations = +7.34 dB$_{HNR}$, animal action sounds = +0.58 dB$_{HNR}$; two-tailed *t-test*, $p < 10^{-7}$), which was expected based on earlier studies (Lewis et al., 2005, 2009). Additionally, as a measure of spectral flatness, which has also been used to study the dynamics of vocal imitation in song birds (Tchernichovski et al., 2001), we assessed the magnitudes of signal entropy (Wiener entropy, WE). Entropy measures, and related spectral structure variation (SSV) measures addressed below, were derived using a freeware script with Praat software ("Wiener

TABLE 1 | List of sound stimuli together with some of their acoustic signal attributes and perceptual features.

Animal Action sounds	HNR	WE	SSV	Intent	Recog	Emotion
Bat flapping wings	2.83	−5.40	1.52	2.00	2.15	−0.07
Bird, flying away						
Cow(s) moving in pen	2.23	−3.39	1.43	2.10	2.65	0.07
Dog (large) eating food from bowl	−1.71	−3.21	1.81	2.90	2.20	0.53
Dog (small) eating food						
Dog drinking water from bowl	1.27	−4.28	2.12	3.20	2.60	0.67
Dog footsteps	−0.68	−1.94	0.38	2.65	2.60	0.07
Dog licking/lapping	−1.52	−2.49	1.53	2.80	2.20	0.33
Dog panting heavily						
Dog shaking head/ears flapping	−0.93	−3.50	0.79	2.80	2.80	0.07
Dog sniffle and digging	−1.01	−2.96	0.80	3.15	3.15	0.53
Dog swimming and shakes	−4.00	−2.37	0.32	2.90	2.95	0.07
Dog trotting, toe nails on floor #1	−0.64	−2.53	0.41	2.75	2.70	0.33
Dog trotting, toe nails on floor #2	−1.24	−3.59	0.73	2.90	3.10	0.80
Dog/animal licking & chewing	−1.40	−2.54	1.52	2.75	2.35	0.27
Horse eating	−0.74	−4.58	1.84	2.30	1.85	0.40
Horse galloping #1	−1.65	−6.17	1.91	2.95	3.20	0.07
Horse galloping #2	3.16	−2.97	2.03	2.90	3.40	−0.07
Horse galloping #3	1.52	−2.60	1.72	3.15	3.45	0.13
Horse galloping #4	3.88	−3.28	2.28	3.15	3.45	0.33
Horse trotting	2.80	−3.38	2.19	3.30	3.65	0.33
Horse walking #1	7.72	−6.02	2.95	3.50	3.75	0.40
Horse walking #2	2.33	−2.87	0.80	3.35	3.40	0.33
Horse walking on cobblestone	1.14	−4.88	1.77	3.25	3.85	0.60
Pigeon flight	−3.67	−2.30	0.36	2.65	2.75	−0.13
Rattlesnake rattling #1						
Rattlesnake rattling #2						
Rattlesnake rattling #3						
Zebra trotting on dirt	3.59	−3.21	2.54	3.05	3.50	0.60
Average	0.58	−3.50	1.47	2.89	2.94	0.29
StDev	2.74	1.18	0.76	0.38	0.57	0.26
Baby bear bark	3.84	−6.90	4.29	2.45	2.30	−1.00
Bear call #1	1.64	−8.89	2.88	2.55	2.50	−0.73
Bear call #2	4.50	−7.96	2.95	2.90	3.25	−0.93
Bear call #3	12.52	−8.25	5.15	2.50	2.75	−0.93
Bear growl #1	−1.43	−8.64	1.51	2.70	2.50	−1.40
Bear growl #2	1.38	−8.40	1.28	2.55	2.20	−1.07
Bear growl #3	1.59	−6.57	0.86	3.05	2.95	−1.60
Bobcat growl	−1.14	−6.47	1.34	3.00	2.70	−1.40
Bull call #1	12.08	−11.25	0.92	2.45	2.75	−0.93
Bull call #2	4.74	−8.19	1.87	2.75	2.65	−1.73
Bull call #3	11.18	−9.86	3.25	2.40	3.50	−0.47
Camel lowing	13.05	−10.00	0.82	1.95	2.60	−1.13
Camel moo	13.01	−10.00	0.78	2.35	2.45	−0.87
Cougar growl	1.83	−7.41	3.71	3.60	3.55	−1.73
Cow moo	16.90	−10.24	1.90	2.05	2.90	−0.80
Dog bark	4.03	−7.51	3.21	3.45	4.00	−0.20
Dog growl-sigh	4.43	−7.01	3.87	2.90	3.10	−0.87
Dog whine and bark	6.27	−7.44	2.41	3.40	3.60	−1.33
Goose (Canadian) honks	9.64	−7.99	9.63	2.45	3.40	0.53
Horse whinny	8.64	−6.14	1.03	3.45	3.90	−1.07
Jackal call	17.79	−7.14	4.29	2.90	3.30	−0.47
Lion roar	1.69	−10.82	2.30	3.00	3.45	−1.00
Monkey scream #1	14.96	−6.20	4.39	3.05	3.75	−0.93
Monkey scream #2	10.08	−5.92	0.88	3.00	3.70	−1.67
Monkey vocal	14.07	−7.37	6.76	2.65	3.75	1.00
Pig squeal	6.70	−5.25	0.69	3.10	3.70	−1.73
Sheep baa	8.46	−5.70	16.66	2.65	3.95	0.60
Wolf bark	9.42	−8.54	2.85	3.35	3.70	−0.60
Wolf growl	1.13	−5.80	0.61	3.35	2.80	−1.47
Average	7.34	−7.86	3.21	2.83	3.16	−0.89
StDev	5.51	1.62	3.30	0.43	0.55	0.69

Italicized entries were censored a priori from fMRI data analyses. HNR = harmonics-to-noise ratio; WE = Weiner entropy; SSV = spectral structure variation; Recog = recognition. Refer to text for other details.

entropy," see online supplemental software script). On average, action sounds showed a significantly greater magnitude in signal entropy: Vocalization $= -3.50$, action sounds $= -7.86$; two-tailed t-test, $p < 10^{-15}$). Using a measure of change in entropy over time (an index of spectral dynamics or rate of change of spectral structure), the animal vocalizations showed on average a greater degree of SSV measures (vocalizations $=$ 3.21, action sounds $= 1.47$; two-tailed t-test, $p < 0.01$). Each of the above signal attributes were assessed and examined here because they have been implicated in auditory stream segregation and auditory object perception in earlier studies (Lewis et al., 2009, 2012; Reddy et al., 2009). We also assessed potential category-level differences in modulation power spectra (MPS), using techniques that have revealed spectro-temporal attributes of sound that may be processed along distinct pathways in auditory cortices in humans and macaques (Woolley et al., 2005; Cohen et al., 2007; Elliott and Theunissen, 2009; Kuśmierek et al., 2012; Herdener et al., 2013; Santoro et al., 2014). We used freely available Matlab software programs to derive MPS measures (http://strfpak.berkeley.edu), using a 32 Hz frequency band and low-pass filter method for the ensemble of 23 animal action sounds and 23 (of the 29) animal vocalizations (**Supplemental Figure S1C**). In short, various quantifiable low- and high-order acoustic signal attributes could distinguish the two semantic categories of sound on average (action sounds vs. vocalizations). Consequently, we referred to the real-world natural categories of sound in **Figure 1** as "acoustic-semantic" categories. We further conducted *post-hoc* analyses of differentially activated brain regions to test for activation that showed parametric sensitivity to several of the above acoustic signal attributes (addressed below), as well as perceptual features, thereby probing for their potential functional roles.

Perceptual Features of the Sound Stimuli

To identify potential perceptual features that may influence cortical processing of the sound stimuli, we assessed the perceived intention behind the sounds, the degree to which they could be recognized, and the emotional valence associated with each sound. Using freely available PsychoPy2 (v1.73.05) software (Peirce, 2007), a separate set of non-imaging listeners ($n =$ 20, 10 female) heard all sounds in a random order while in a sound isolation booth, and rated them on a 1-4 Likert scale with the instructions: "With regard to the sound you heard, your ability to recognize the Intention was: (1) almost certainly not, (2) unlikely, (3) likely, to (4) almost certainly". While there was a full range of ratings, overall there were no significant differences in mean ratings of intent between the two categories of sound (**Table 1**, *t*-test, two-tail $p <$ 0.58). Using a counter-balanced design, these participants also separately rated each sound for their ability to recognize the sound with the instructions: "With regard to the sound you heard, your ability to Recognize it was: (1) almost certainly not, (2) unlikely, (3) likely, to (4) almost certainly." Again, there was a full range of responses but no significant differences in mean ratings between categories (*t*-test, two-tail $p <$ 0.16).

To assess the emotional valence of the sounds, another group of listeners ($n = 15$, 8 female) rated all the sounds, presented in a random order (using PsychoPy software), on a 5-point scale of perceived emotional valence. This ranged from very negative (-2), to neutral (0), to very positive ($+2$) for the degree of emotional content ascribed to the agent (animal) producing the action sound or vocalization. The animal vocalizations were rated (**Supplemental Figure S1B**) as having a relatively negative emotional valence overall (Mean/$SD = -0.89 \pm 0.69$) while animal action sounds were rated closer to neutral valence ($+0.29 \pm 0.26$), which were ratings that on average significantly differed from one another [single factor ANOVA, $F_{(1, 51)} = 61.3$, $p < 10^{-13}$].

Scanning Paradigms
Sound Categorization Paradigm
The main auditory scanning paradigm for fMRI imaging ($n = 17$, 8 female) consisted of two separate runs that were to be presented three times each. Across the two runs, the 58 sound stimuli plus 24 silent events were presented in pseudo-random order, with no more than two silent events presented in a row. This randomized approach was adopted to avoid or minimize activation related to state-dependent effects observed in block designs (Rehme et al., 2013). The high-fidelity sound stimuli were delivered using a Windows PC computer, with Presentation software (version 11.1, Neurobehavioral Systems Inc.) via 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: Immediately after scanning a computer generated 1 kHz pure tone was played through the sound transmission system and the intensity was typically in the range of 80–83 dB C-weighted to each ear (Brüel & Kjær 2239a sound meter). Participants were instructed to determine by 2AFC left hand button response whether the sound heard was (1) an animal vocalization (index finger) or (2) an animal action sound (middle finger). Although none of the participants had any previous exposure to these specific sound stimuli, they were presented with several of the sound stimuli during a brief practice fMRI scan to ensure the instruction were clearly understood, and they were explicitly informed that all the sounds they heard would be animal actions and animal vocalizations.

Button Press Control Paradigms
To identify brain regions that may be involved in the preparation and mechanics of button presses, we conducted two motor output-related control paradigms germane to the main auditory paradigm. As a first control condition, one participant performed the main auditory paradigm using her right hand for button presses for two runs and her left hand for two runs. This allowed for a direct comparison of the anticipated effects of primary motor cortex activation related to the hand used when responding. As a second control condition, two of the participants performed a separate button press paradigm after the main sound categorization task. This involved a simple ON/OFF block design cued by short tone pips every 20 s for 4 ON and 5 OFF blocks. During the ON cycles, they made self-initiated

button box responses with their left index or middle finger at approximately the same rate as during the main auditory scanning paradigm. ON cycles were flanked by OFF cycles during which they would rest their hand.

Magnetic Resonance Imaging and Data Analysis

The fMRI imaging paradigms were all conducted on a 3 Tesla Siemens Verio MRI scanner using a 32-channel head coil. For all paradigms, participants kept their eyes closed. For the auditory paradigm, a sound or silent event was presented every 10 s, and BOLD signals were collected continuously using an echo planar pulse sequence (ep2d: TR = 2000 ms, TE = 30 ms, FOV = 240 mm, 75° flip angle). Whole-head brain volumes were collected, including 32 axial slices at 4 mm slice thickness (plus 0.6 mm gap) and 3.75×3.75 mm^2 in-plane resolution. Continuous scanning, as opposed to sparse sampling, was utilized in order to obtain BOLD signals over time that could be more robustly analyzed using gamma-variate models. Sound stimuli were heard over continuous scanner noise. The presentation of each auditory stimulus event was triggered by the fifth transistor-transistor logic (TTL) pulse from the MRI scanner for each TR (triggering sound or silent events every 10 s), thereby ensuring accurate time stamps for image acquisition relative to stimulus sound onsets and for recording button press reaction times. After the completion of the functional imaging scans, whole brain T1-weighted anatomical MR images were collected using an MPRAGE pulse sequence (1.5 mm sagittal slices, 0.625×0.625 mm^2 in-plane resolution, TI = 1100 ms).

All functional datasets were processed using Analysis of Functional NeuroImages (AFNI) and associated software plug-in packages (http://afni.nimh.nih.gov/) (Cox, 1996). For each paradigm, the 20th volume of the final functional scan, closest to the anatomical image acquisition, was used as a common registration image to globally correct motion artifacts due to head translations and rotations using program 3dvolreg. Voxels were subjected to a Gaussian spatial blurring of 8 mm (Mikl et al., 2008), and BOLD signals were then converted to percent signal change on a voxel-wise basis relative to BOLD signal during silent events for each scanning run for each participant.

For subsequent group-level analyses we conducted both volumetric and surface-based alignment methods. Surface-based alignment methods generally enable greater accuracy in the localization of activated ROIs across the cortical folds, such as along the superior temporal plane and fronto-parietal operculum (Desai et al., 2005). However, both alignment methods yielded qualitatively similar results for our datasets, so only the results from the standard Talairach volumetric alignment method are presented here. Individual anatomical brain volumes were manually aligned to standardized Talairach space (Talairach and Tournoux, 1988). For each individual, scanning runs were concatenated into a single time series and corrected for baseline linear drifts (six runs for fourteen participants, five runs for one participant, and four runs for two participants). A multiple linear regression analysis was performed using a gamma-variate waveform model (using AFNI program 3dDeconvolve) of the
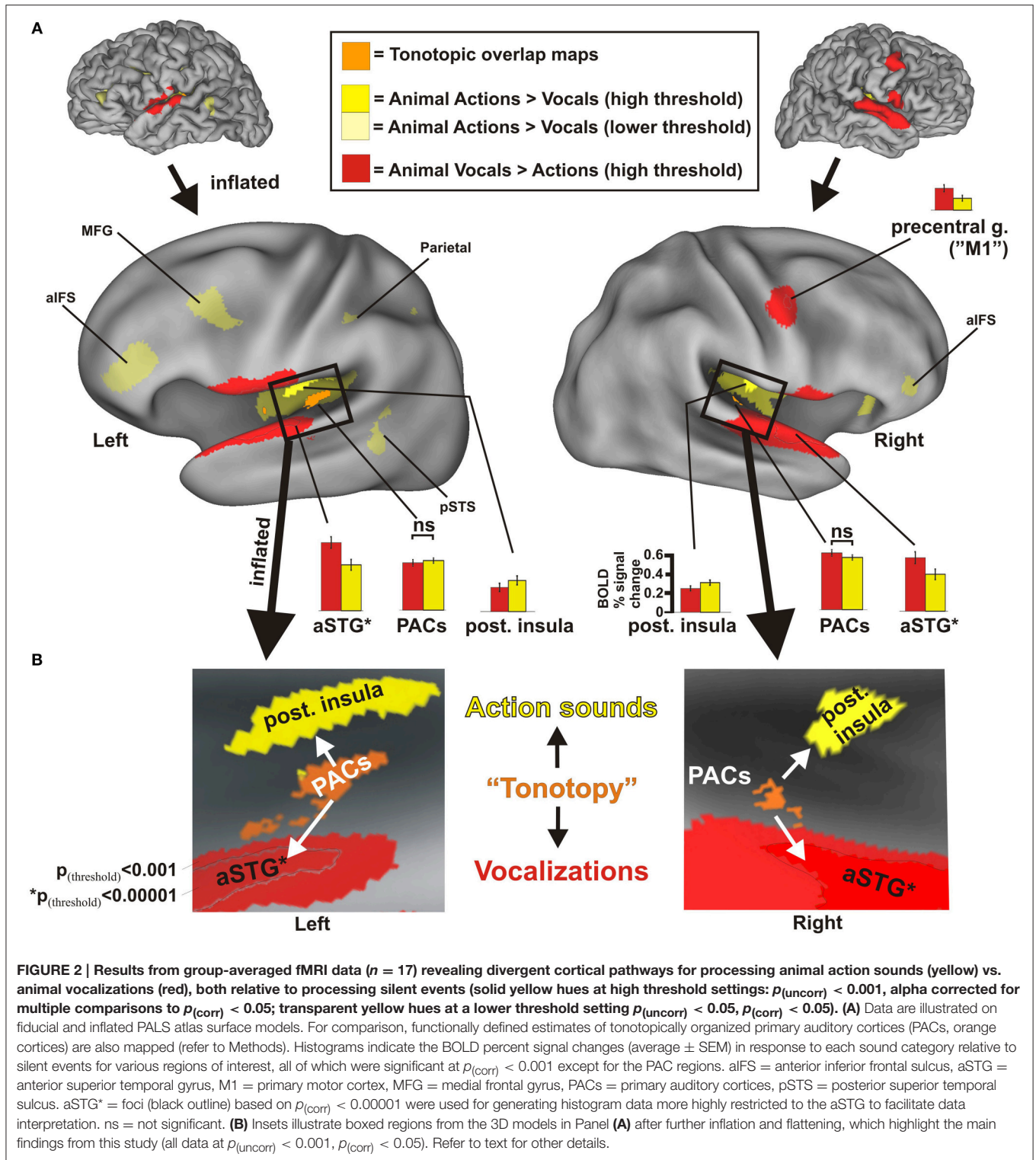
sound onsets for each category of sound stimulus to compare cross-categorical BOLD brain responses (animal vocalizations vs. animal action sounds), both relative to averaged BOLD signals from silent events as a baseline control (used for computing percent signal changes). We further restricted the analysis to reveal only those contrasts where the averaged BOLD activation responses to vocalizations or action sounds, or both, were positive relative to the responses to baseline silent events.

For the primary analyses, only those imaging events that corresponded to correctly categorized sounds for a given participant (as determined by button press accuracy) were retained for analyses (average 96.9% accuracy for vocalizations and 94.3% for action sounds across all scanning runs, rejecting trials with omissions or commissions). There were insufficient numbers of incorrectly categorized sounds for statistical error-trial analyses. The resulting BOLD multiple regression coefficients (statistical maps) were spatially low-pass filtered (6 mm box filter) and subjected to *t-test* and thresholding. We also processed all individual datasets using no spatial blurring, which yielded qualitatively similar results but required lower group-averaged threshold settings (data not shown).

To correct for multiple comparisons (multiple inference problem) (Forman et al., 1995; Eklund et al., 2016), we used AFNI related software. We estimated spatial blurring in our data by examining signal noise from the residual error time series from the full model of two individual's input data runs (3dDeconvolve software), blurring the error signals at 6 mm (roughly 1.5 times the size of acquisition voxels). This yielded estimated full-width half-max Gaussian filter widths of x = 6.5, y = 6.5, and z = 6.4 mm spatial smoothness. A brain mask was created for each of the subjects (3dAutomask software; dilation factor of 2). Using the brain mask and the above spatial filter widths, 10,000 Monte Carlo simulations were applied using 3dClustSim (version 16.2.06) to obtain corrected significance levels. For data in **Figures 2, 3**, a combination of a minimum cluster size threshold (6.4 voxels, or 414 mm^3 volume; two-sided thresholding) and an individual voxel probability threshold ($p_{thr} < 0.001$) achieved a specified overall false detection probability (i.e., family-wise error rate; FWER) of alpha = 0.05, [$p_{(corr)} < 0.05$], which was herein designated as the high threshold setting. A lower threshold setting was also used to reveal subtler effects, which was obtained with a cluster size of 44.6 voxels (2885 mm^3 volume) and $p_{thr} < 0.05$, which also yielded a corrected significance level of $p_{(corr)} < 0.05$. Functional imaging data were transformed into standardized Talairach coordinate space and for visualization purposes were projected onto the PALS atlas cortical surface models (in AFNI-Talairach space) using Caret software (http://brainmap.wustl.edu) (Van Essen et al., 2001; Van Essen, 2005).

Multiple Linear Regression Analyses

Several *post-hoc* linear and multiple linear regressions analyses were used to identify correlations between various acoustic signal

FIGURE 2 | Results from group-averaged fMRI data ($n = 17$) revealing divergent cortical pathways for processing animal action sounds (yellow) vs. animal vocalizations (red), both relative to processing silent events (solid yellow hues at high threshold settings: $p_{(uncorr)} < 0.001$, alpha corrected for multiple comparisons to $p_{(corr)} < 0.05$; transparent yellow hues at a lower threshold setting $p_{(uncorr)} < 0.05$, $p_{(corr)} < 0.05$). (A) Data are illustrated on fiducial and inflated PALS atlas surface models. For comparison, functionally defined estimates of tonotopically organized primary auditory cortices (PACs, orange cortices) are also mapped (refer to Methods). Histograms indicate the BOLD percent signal changes (average ± SEM) in response to each sound category relative to silent events for various regions of interest, all of which were significant at $p_{(corr)} < 0.001$ except for the PAC regions. aIFS = anterior inferior frontal sulcus, aSTG = anterior superior temporal gyrus, M1 = primary motor cortex, MFG = medial frontal gyrus, PACs = primary auditory cortices, pSTS = posterior superior temporal sulcus. aSTG* = foci (black outline) based on $p_{(corr)} < 0.00001$ were used for generating histogram data more highly restricted to the aSTG to facilitate data interpretation. ns = not significant. **(B)** Insets illustrate boxed regions from the 3D models in Panel **(A)** after further inflation and flattening, which highlight the main findings from this study (all data at $p_{(uncorr)} < 0.001$, $p_{(corr)} < 0.05$). Refer to text for other details.

attributes (**Table 1**) or perceptual features of the sound stimuli (within and across categories) with activation strength of a voxel-wise basis for various regions of interest (ROIs). We modeled the BOLD signal as the linear convolution of the input stimulus (whether vocal sounds or action sounds), with the hemodynamic response function (HRF) for that particular stimulus. This BOLD signal was in addition to the baseline signal, which was modeled as a 3rd order polynomial, which was fit separately for each of
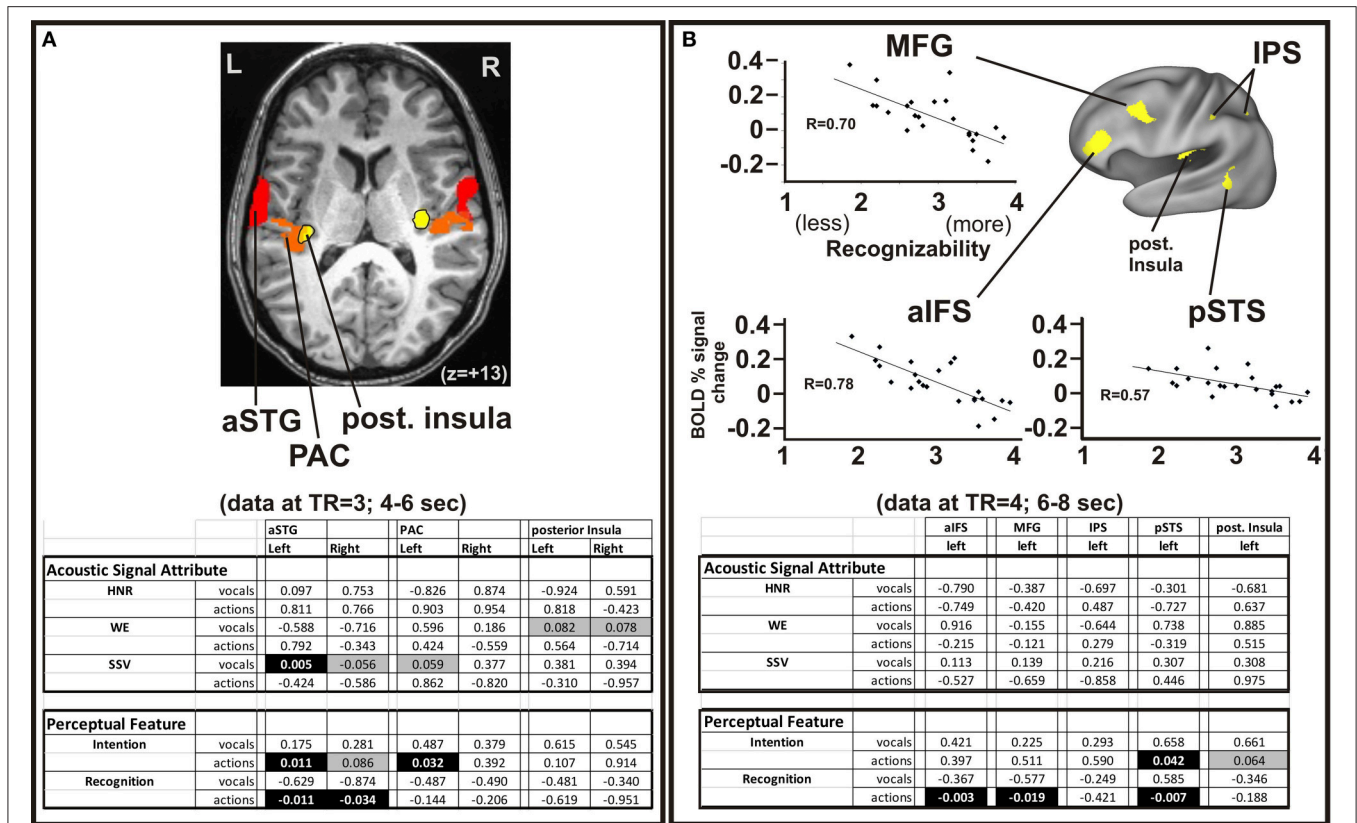
FIGURE 3 | Parametric activation analyses related to the acoustic signals and perceptual attributes of the animal sounds. (A) The group-averaged datasets from **Figure 2** are superimposed on an axial section of one participant (Talairach z= +13) showing the divergent processing regions (yellow vs. red) in both hemispheres. Multiple linear correlation *p*-value fits between regions of interest with the quantifiable high-order acoustic signal attributes (HNR, WE, SSV) and perceptual features (Intention, Recognition) of the animal action sounds and vocalizations (from **Table 1**) at the brain imaging time phase of TR = 3, occurring 4–6 s after sound onset. Negative correlations are indicated by negative value entries. Cells with black highlighting depict *p* < 0.05 (significant), and cells with gray highlighting depict *p* < 0.09 (trending). **(B)** Top panels show linear correlation charts for three of the five left hemisphere ROIs (MFG, aIFS, and pSTS; also evident in **Figure 2**). Charts illustrate BOLD percent signal change activation relative to acoustic signal attributes and perceptual ratings as in **Panel A**, except at the brain imaging time phase of TR = 4, occurring 6–8 s after sound onset. Lower panel shows results from a multiple linear regression analysis (correlation *p*-value fits) for all five ROIs sensitive to animal action sounds. Refer to Methods for further details.

the six concatenated runs. This resulted in the following multiple linear regression model for the fMRI signal:

$$y(t) = \sum_{k=1}^{6} \left[b_{0,k} + b_{1,k}(t - t_{0,k}) + b_{2,k}(t - t_{0,k})^2 \right.$$
$$\left. + b_{3,k}(t - t_{0,k})^3 \right] I_k(t) + \sum_{i=0}^{15} h_{VS}(i)\ VocalSounds\,(t-i)$$
$$+ \sum_{j=0}^{15} h_{AS}(j)\ ActionSounds\,(t-j) + \varepsilon(t)$$

where $y(t)$ = measured fMRI signal (six concatenated runs), k = run index, $t_{0,k}$ = time at beginning of kth run, $I_k(t)$ = indicator function for kth run (i.e., $I_k(t) = 1$ if t occurs during kth run, 0, otherwise), $h_{VS}(i)$ = HRF for vocal sounds, $h_{AS}(j)$ = HRF for action sounds. Least squares fitting of the above model allowed for testing of various hypotheses concerning the fit coefficients, such as testing for significance using the above

signal time series as well as testing at two distinct temporal phases of BOLD acquisition time, including the TR's corresponding to 4–6 s and 6–8 s after sound onset, respectively.

Reaction Time Data Analyses

The average response times to correctly categorized vocalizations across all scanning runs was 2.26 ± 0.25 s and to correctly categorized (and *post-hoc* censored) animal action sounds was 2.26 ± 0.28 s, showing no significant difference [ANOVA $F_{(1, 157)}$ = 0.014, *p* < 0.90]. To characterize potential within-session changes in task performance, we analyzed both reaction time and accuracy response data. On average the reaction times to categorizing the sounds during the first two of six runs vs. the last two runs showed changes across the scanning session: For animal vocalizations was 2.53 s and then 2.35 s respectively (*n* = 13 subjects; two tailed *t-test*, *p* < 0.0013), and for animal action sounds was 2.51 and then 2.35 s (*p* < 0.0011). The first two runs for vocalization sounds yielded 97.6 vs. 97.6% accuracy (no difference) in the last two runs, while animal action

sounds yielded 98.0 vs. 96.0% (two-tailed t-test, $p > 0.15$, not significant) respectively. Thus, while there was some reaction time improvement over the course of the experiment and some mild perceptual deterioration effects over time (Mednick et al., 2005), participants were effectively performing at ceiling levels throughout the scanning procedures, as expected for this simple task of categorizing the easily recognized animal sounds.

Tonotopic Map Construction

As part of an earlier study (Lewis et al., 2009), a different set of participants ($n = 7$) performed a tonotopy mapping paradigm to reveal primary auditory cortices (PACs). This served to create functionally defined estimates of frequency-dependent response regions (FDRRs; or "tonotopic maps"). Regions showing tonotopic progressions (typically abutted by a mirror reversal map) were outlined for each individual and collated as a heat map, providing a probabilistic estimate of the location of PACs. Cortical volumes showing overlap of tonotopic maps by at least two individuals and that intersected with the transverse temporal gyri (Heschl's gyrus) were retained and projected onto the cortical models. The PAC foci from this prior study were largely localized to the medial two thirds of Heschl's gyri, consistent with the location of primary auditory cortices reported in anatomical studies (Rademacher et al., 2001) and functional studies (Formisano et al., 2003; Talavage et al., 2004), and thus adopted here as ROIs (**Figure 2**, orange) for comparative analyses.

RESULTS

Using an event related continuous acquisition fMRI paradigm, we imaged participants while they listened to non-conspecific animal vocalizations and action sounds, plus silent events as a baseline control. They indicated by button press as to whether the sound heard was (1) a vocalization or (2) an action sound by two alternative forced choice button response with the left hand. Both categories of sound stimuli were carefully balanced and matched overall for duration and intensity. Contrasting brain activation preferential for one vs. the other semantic category of sound source revealed a robust double-dissociation of cortical regions showing significant preference for processing animal action sounds medially in the posterior insulae [**Table 2** and **Figure 2**, solid yellow at $p_{(uncorr)} < 0.001$, corrected to $p_{(corr)} < 0.05$ for minimum cluster size, and transparent yellow at $p_{(uncorr)} < 0.01$, $p_{(corr)} < 0.05$] vs. processing of animal (non-conspecific) vocalizations laterally in the middle and anterior superior temporal gyri (STG) [red; $p_{(uncorr)} < 0.001$, $p_{(corr)} < 0.05$]. As functional landmarks, we also charted the estimated location of primary auditory cortices (PACs) based on an overlay of tonotopically organized maps (**Figure 2**, orange, ROIs from our previous work; see Materials and Methods).

Supporting our theoretical model for semantic-level sound categorization (**Figure 1**), the main finding of the present study was the presence of spatially distinct cortical regions in both the left and right hemisphere auditory cortices (i.e., along or immediately surrounding the superior temporal plane) for preferentially processing the two semantically (and

TABLE 2 | Group activation centroids in Talairach coordinate space for activation foci.

Anatomical location	Talairach coordinates			Volume (mm³)
	x	y	z	
VOCALIZATIONS > ACTION SOUNDS ($P < 0.001$)				
Left aSTG	−58	−4	2	20,069
Left temporal pole	−34	10	−34	31
Right aSTG	54	−4	1	16,801
Right precentral g. (M1)	47	0	48	3303
ACTIONS SOUNDS > VOCALIZATIONS ($P < 0.001$)				
Left Posterior Insula	−36	−28	18	1216
Right Posterior Insula	32	−16	15	761
ACTIONS SOUNDS > VOCALIZATIONS ($P < 0.05$)				
Left MFG	−43	9	32	2848
Left Parietal	−42	−36	31	2011
Left pSTS	−52	−50	3	964
Left aIFS/vlPFC	−45	41	8	4908
Right aIFS/vlPFC	30	29	6	2836

*See text and **Figure 2** legend for abbreviations.*

acoustically) distinct categories of non-conspecific biological sounds. Specifically, this included the left and right posterior insulae for animal action sounds, and the left and right aSTG for animal vocalizations. The independently derived functional estimates of PACs showed comparable levels of activation to both categories of sound (**Figure 2A**, orange; see histograms).

Relative to the action sounds, the vocalizations additionally showed preferential activation along the bilateral fronto-parietal operculum (**Figure 2A**, red) that was volumetrically contiguous with portions of the STG foci (evident in many individual datasets, and thus due at least in part to limitations in spatial resolution), plus the right lateral precentral cortex, overlapping functionally identified primary motor cortex ("M1"), which was shown to relate to left-hand button presses (see below). At lower threshold settings, brain responses to the animal action sounds additionally revealed preferential BOLD activation (**Figure 2A**, transparent yellow) in the left middle frontal gyrus (MFG), left parietal cortex (intraparietal sulcus), and left posterior superior temporal sulcus (pSTS), plus the bilateral anterior inferior frontal sulcus (aIFS) regions (left > right; which roughly overlapped ventro-lateral prefrontal cortices, vlPFC, more generally). This network directly overlapped with the classically defined left-lateralized mirror neuron system (Molenberghs et al., 2012). While mapping the full extent of mirror neuron systems (MNS) was beyond the scope of the present study, these results demonstrate at least partial recruitment of a left-lateralized motor-related system for processing and categorizing the animal action sounds relative to networks for processing animal vocalizations.

Button Response and Right Motor Cortex

Participants used their left hand for button box responses with the index finger being used for vocalizations and the middle finger for action sounds. The consistent use of the left index

finger for responses to the vocalization category almost certainly accounted for increased activation located along estimated right primary motor cortex (right "M1"; **Figure 2A**, red) in the vicinity of the left hand homunculus representation (Penfield and Boldrey, 1937). To verify this assertion, we performed three different types of analyses and control paradigms, as addressed in the Supplemental Results section (**Supplemental Figure S2**). In short, the three control paradigms and conditions examined demonstrated that the right precentral gyrus activation was due primarily, if not exclusively, to factors related to button response preparation and execution as opposed to acoustic signal processing related to one vs. the other category of sound.

Sensitivity to Acoustic Signal Attributes and Perceptual Features of Natural Sounds

In addition to category membership, a number of acoustic signal attributes distinguished the animal action sounds from vocalizations. For instance, vocalizations were characterized by greater harmonic content (e.g., **Table 1**), qualitatively higher rates of spectral modulation (**Supplemental Figure S1C**), greater spectral structure variation, and lower signal entropy measures (**Table 1**). Systematic exploration of these attributes, and a potential myriad of other signal attributes, of natural sounds that may lead to the observed processing double-dissociation was beyond the scope of the present study. Nonetheless, to explore *why* the two acoustic-semantic categories of natural sound may have led to differential processing, we examined various regions of interest for group-averaged parametric correlation with some of the signal attributes and perceptual features that could be parameterized along a single dimension (see Methods). We examined both single and multiple linear regression models that tested for significant correlations between BOLD signal responses within the ROIs for all sound stimuli in a given category (**Figure 3**). Note, that this study did not *a priori* attempt to match the range of acoustic signal attributes or perceptual ratings of the action sounds and vocalizations (e.g., clusters by category in **Supplemental Figure S3**). Consequently, ceiling or floor effects could in part account for some of the combinations of category-specific results. Nonetheless, significant correlations and trends persisted that served to facilitate data interpretation. In short, the left and right hemisphere auditory cortex ROIs (**Figure 3A**; aSTG, PAC, and posterior insulae) collectively showed a mix of parametric sensitivity to both acoustic signal attributes and perceptual features, while the left-lateralized ROIs preferential for animal action sound processing (**Figure 3B**; aIFS, MFG, pSTS) showed parametric sensitivity mostly to perceptual features of the sounds (**Figures 3A,B** highlighted cells in tables), as detailed further below.

The aSTG regions (left > right) showed a significant positive correlation of activation strength (at the temporal phase of 4–6 s after sound onset) with increasing spectral structure variation (SSV) measures of the animal vocalizations (**Figure 3A**), and the right aSTG and left PAC both trended toward positive linear correlation of activation strength with SSV. The vocalizations showed greater spectral modulation rates at least qualitatively (**Supplemental Figure S1C**), which may

be correlated with SSV measures—thought detailed analyses along these lines went beyond the scope of the present study. Additionally, SSV measures were positively correlated with perceived emotional valence ratings (**Supplemental Figure S3A**, chart: $R^2 = 0.56$; Steiger's Z = 3.31, two-tailed $p < 0.01$). Thus, there may be a relationship between perception of emotional cues in the animal vocalizations and aSTG activation (SSV-sensitivity), though this determination would require future study. The left PAC and bilateral aSTG regions also showed significant activation strength that demonstrated a positive co-linear correlation (**Figure 3A**, lower table) with the perceived understanding of intention behind the animal action sounds; the left and right aSTG also showed a positive co-linear correlation between activation to animal action sounds and the degree of perceived recognition of the animal action sounds. The posterior insulae showed a trend toward correlation between activation strength and decreasing acoustic signal entropy. Moreover, the entropy measures of the vocalizations were found to be positively correlated with the perception of understanding the intention behind the sound (**Supplemental Figure S3B**; chart $R^2 = 0.31$; Steiger's Z = 2.20, two-tailed $p < 0.05$).

In our earlier studies using animal vocalizations, we revealed auditory cortices and STG regions that were sensitive to harmonic content (Lewis et al., 2005, 2009). In the present study, the left aSTG in this ROI-based analysis only trended toward showing a linear parametric correlation with the harmonics-to-noise ratio (HNR) values of the animal vocalizations (data not shown), but this trend in correlation did not survive with the linear regression analyses that included multiple signal and perceptual parameters (**Figure 3A**). This apparently weak presence of parametric sensitivity to harmonic content was likely due to both the specific location of the aSTG foci of the present study (being relatively more lateral to HNR-sensitive regions) plus the high end range of HNR values of the vocalizations (avg. $+7.34$ dB_{HNR}), which were effectively at or near ceiling levels. The animal action sounds did not exhibit parametric HNR-sensitivity in any of the ROIs using any of the regression model configurations.

The left hemisphere aIFS, MFG, IPS, pSTS, and posterior insula ROIs (**Figure 3B**) were also tested for parametric sensitivity at the 4–6 and 6–8 s TR delays. Interestingly, at the later temporal phase (6–8 s after sound onset) the left aIFS, MFG, and pSTS regions all showed significant negative co-linear correlations between activation strength and degree of perceived recognition of the animal action sounds (**Figure 3B**, charts). The left posterior insula showed a trend for activation strength that correlated with the degree of perceived understanding of the intention behind the animal action sounds, but no correlation with perceived recognition in this or any of the other multiple linear regression analyses tested. The motor-related cortical network, however, was parametrically activated more strongly by animal action sounds that were generally deemed as more difficult to recognize. Thus, there was an apparent hierarchy of activation stages that correlated with different aspects of hearing perception, with the posterior insulae followed by left-lateralized non-auditory regions including the aIFS, MFG, and pSTS.

In sum, the above results revealed two distinct regions of auditory cortex *proper* located medial-posterior vs. lateral-anterior to primary auditory cortices (core) regions that showed preferential activation to processing animal action sounds vs. animal vocalizations, respectively. A number of left-lateralized cortical regions consistent with motor-related functions additionally showed preferential activation to the animal action sounds. An ROI analysis of parametric sensitivity to high order acoustic signal attributes and/or psychophysically assessed perceptual features revealed a gradient of response profiles for the different regions, reflective of their different functional roles in natural sound processing. In particular, these results were consistent with designating the posterior insulae as representing intermediate stages of a cortical processing hierarchy that subserves categorical processing and perception of action events at an acoustic-semantic level.

DISCUSSION

The present study revealed not only the existence of a major divergence of cortical regions in humans for processing action sounds vs. vocalizations, using duration and intensity matched non-conspecific natural sounds (**Figure 2**; and Supplemental Figure S4), but further identified a combination of both bottom-up high-order quantifiable acoustic signal attributes and top-down perceptual features that may mechanistically guide the divergence of processing along these two apparently distinct pathways. Importantly, these results support our neurobiological theoretical framework of cortical organizations mediating the perception of natural sounds at an acoustic-semantic category level (**Figure 1**). In particular, the results identify intermediate cortical processing stages along a hierarchical network that appears to mediate the perception of natural sounds. We selected non-human animal sounds as a sub-category of "living things," which entailed real-world ethologically meaningful sound-producing events that most land mammals arguably need to be able to efficiently process and interpret to survive. Given our use of non-conspecific natural sounds, the cortical organizations revealed herein should apply to most, if not all, land mammalian species with hearing (and perhaps oral communication) ability.

Analogous to hierarchical and parallel processing pathways along core, belt, and parabelt auditory regions in primates (Rauschecker et al., 1995; Rauschecker, 1998; Kaas and Hackett, 2000), the separate pathway findings of the present study were consistent with representing a hierarchical processing system for extracting acoustic attributes (e.g., harmonicity, entropy, spectral structure variation, and high spectral modulation rates) and/or matching sound features to templates, which is in line with aspects of earlier models of spectral and temporal processing pathways in primates (Bendor and Wang, 2008). Presumably, acoustic information processing in primary auditory cortices (**Figure 2B**, orange) was subsequently routed out to the aSTG and the posterior insulae, which are regions that are not activated by lower order acoustic attributes or "simpler artificially produced sounds" such as pure tones, white noises, amplitude or frequency

modulated tones and the like (Kaas and Hackett, 1998, 2000; Rauschecker and Scott, 2009). The aSTG and posterior insula regions showed parametric sensitivity to some of the perceptual features tested, including intention and emotional valence, and thus reflected processing or representation of either bottom-up acoustic signal attributes or some combination of signal attributes and top-down perceptual features. In this manner, the current findings extend "dorsal-ventral" dual-stream models of sound processing (Rauschecker and Tian, 2000; Arnott et al., 2008; Rauschecker and Scott, 2009; Hamzei et al., 2016). In particular, with regard to the ventral sound recognition stream ("what is it"), the two observed regions of the present study (bilateral posterior insulae and aSTG foci) appear to reflect portions of two major auditory pathways for purposes of sound categorization and recognition, distinguishing among different categories of biological ("living") sound-sources at an acoustic and acoustic-semantic level—action sounds vs. vocalizations.

Notwithstanding, auditory streams directed dorsally from primary auditory cortices have often been associated with spatial-related ("where is it") processing (*ibid*). However, the present data suggest that the posterior insulae and the more dorsally located "non-auditory" regions (aIFS, MFG, IPS) in the left hemisphere subserve functions relating specifically to sound recognition. These more dorsal "non-auditory" cortical stages showed sensitivity almost exclusively to high level perceptual features of the animal sounds (e.g., **Figure 3B**), such as the degree of perceived intention and/or perceived recognition. Activation of these latter stages presumably had greater dependence on factors such as personal experience and expertise, which was consistent with earlier animal studies that fail to find neurons in prefrontal regions that are sensitive to acoustic signal attributes *per se* (Cohen et al., 2007). This basic finding of a three tiered hierarchy for the processing of sounds (auditory core, posterior insulae, dorsal left-lateralized cortical network) revealed the bilateral posterior insulae as intermediate processing stages subserving categorization and recognition of real-world natural sounds at an acoustic-semantic level. Consequently, the results of the present study may reflect central processing mechanisms that were critical to the evolution and neurodevelopment of acoustic communication systems in the brain, as addressed below in the context of proposed functional roles of the posterior insulae and aSTG regions.

Functional Roles of the Posterior Insulae

Our earlier studies of hearing perception revealed the left and right posterior insulae as regions preferential for processing non-conspecific animal action sounds relative to other categories of sound-producing events, including human, mechanical and environmental sounds (Engel et al., 2009; Lewis et al., 2011). However, the functional role(s) of the posterior insulae in humans remained elusive. In the macaque monkey, the retroinsular (Ri) area shares at least some topological overlap with the posterior insulae of humans (Hackett et al., 2007) and thus may share some homologous functions. Area Ri has direct neuronal connections with auditory cortices as well as with second somatosensory cortex (S2) and several parietal

areas (*ibid*). Moreover, preliminary electrophysiological data indicate that area Ri in the macaque has strong auditory and somatosensory responses, and robust superadditive multisensory (audio-tactile) interactions (O'Connell et al., 2004). Sound stimuli used in the present study revealed animal action sounds as having greater power in the lower frequency ranges (**Supplemental Figure S1A**), possibly reflective of the temporal dynamics or cadences of locomotion sounds. This finding may prove to relate to greater temporal modulation rates of action sounds more generally (Schönwiesner and Zatorre, 2009; Herdener et al., 2013), and thus is a promising topic for further study using sound stimuli that have acoustically well-controlled higher order signal attributes. Animal action sounds also had higher entropy measures than vocalizations, reflective of action sounds being significantly less "acoustically organized" on average. Together, the above neuroanatomical considerations and perceptual features suggest a possible role for the human posterior insulae in representing action sounds in relation to audio-tactile and audio-motor associations, or perhaps more generally in abstracting representations in the neural code of the temporal dynamics of observable real-world action events.

At lower threshold settings, the animal action sounds additionally evoked preferential activation involving left lateralized frontal, parietal, and pSTS regions (**Figures 2A**, **3B**), which appeared to entail some of the classical dorsal ("where is it") pathways (refer to Introduction). However, this more dorsally-directed system was unlikely to be recruited for purposes of sound localization relative to one's own motor spatial coordinates *per se* (e.g., to engage or avoid the sound-sources), since binaural spatial cues were removed and there was no overt spatial processing task. Nor was this dorsal system recruited for any overt linguistic-related processing (e.g., segmenting amplitude envelopes of human communication sounds) as observed in some studies of spoken language (Burton et al., 2000; LoCasto et al., 2004), since only non-verbal, non-conspecific animal sounds were used. Rather, this left-lateralized fronto-parietal plus pSTS system was more reminiscent of the location and functions ascribed to classically defined left lateralized mirror neuron systems (MNS) revealed during overt action observation (Rizzolatti and Arbib, 1998; Rizzolatti and Craighero, 2004; Molenberghs et al., 2012). MNS-like network activation in hearing perception studies in humans are thought to reflect processing that provides a probabilistic match between heard action sounds and the listener's own repertoire of sound producing actions (e.g., walking, chewing food), thereby providing a sense of meaning or intention behind the action (Buccino et al., 2001; Bidet-Caulet et al., 2005; Calvo-Merino et al., 2005; Gazzola et al., 2006; Lahav et al., 2007; Mutschler et al., 2007; Galati et al., 2008; Engel et al., 2009; Lewis et al., 2011). Ostensibly, recordings from frontal and parietal cortex neurons of the macaque monkey has revealed responsiveness when the animal either performs a goal-directed, sound-producing action or while hearing, without viewing, the same motor act (Kohler et al., 2002; Keysers et al., 2003). Thus, the primate brain appears to map complex acoustic signals and attributes of action sounds to motor- and

sensory-motor-related associations as a general strategy that may subserve categorization and aspects of recognition of natural sounds.

Importantly, in the present study the non-conspecific animal action sounds, relative to vocalizations, led to relatively more robust activation of the bilateral insulae than to the left-lateralized MNS-like system. In contrast, we reported the opposite pattern when processing <u>human</u> action sounds relative to animal action sounds in our earlier studies of action sounds (Engel et al., 2009; Lewis et al., 2011). Thus, the posterior insulae in humans may represent stages prior to MNS-like systems for processing acoustic and perceptual attributes statistically characteristic of heard action events. While the animal action sounds were generally characterized by significantly lower harmonic content, lower spectral structure variation, and less acoustic signal organization (i.e., greater signal entropy) relative to vocalizations (**Supplemental Figure S3B**), they did show greater average spectral power in the 10–50 Hz range (see **Supplemental Figure S1A**), which is a range known to be important for conveying amplitude envelope information, such as for speech processing (Rosen, 1992). The amplitude envelope cadences representative of locomotion sounds (e.g., a horse trotting) presumably contain both low and high-order quantifiable signal attributes, such as specific envelope ranges and combinations of temporal modulation rates, that could be driving bottom-up information processing to and including pathways involving the left and right posterior insulae. This processing may in turn lead to network comparisons of acoustic temporal dynamics with the listener's own repertoire of sound-producing motoric representations in motor- and visuomotor-related networks (including left-lateralized aIFS, MFG, and IPS regions). Processing in these regions with the task of action event categorization and recognition may subsequently match, or fail to adequately match, the listeners' repertoire of sensorimotor sound producing actions to facilitate or mediate the perception of intention, and/or a sense of recognition, behind the action sound. The animal action sounds of the present study were clearly perceived as non-human, as opposed to human or self-like, and thus would arguably be less readily emulated or mimicked through motor imagery-related mechanisms. Thus, animal action sounds lead to relatively less robust recruitment of MNS-like systems for purposes of recognition, consistent with our earlier studies contrasting human and animal action sounds (Engel et al., 2009; Lewis et al., 2011). Moreover, the non-auditory dorsal network for action sounds revealed activation strength that was parametrically greater for animal actions that were generally judged by participants as being more difficult to recognize. This suggests that audio-motor association networks may have been activated to a relatively greater extent by the sounds that were harder to recognize before finally settling on a "best solution" match (Hopfield and Tank, 1985) that effectively deemed the sound as being produced by a living agent, but was not the action of a self-like (i.e., conspecific) sound-source.

Given the current interest and contentions in debates regarding MNS systems in the context of gestural origins behind the evolution of spoken language systems (Gallese and

Goldman, 1998; Corballis, 1999; Arbib, 2005; Cerri et al., 2015), further study of the functional roles of the posterior insulae would be merited to address predictions made by anthropological theories. In particular, one theory hypothesizes that symbolic representations in the brain for specific natural categories, including "vocalizations" and "incidental sounds of locomotion" (in addition to "tool-use sounds"), reflect some of the earliest sound-producing categories of events that would need to have been effectively communicated orally by hominins (Hewes, 1992; Larsson, 2014, 2015). Findings of the present study, suggestive of distinct cortical processing pathways for representing action sounds as a distinct acoustic-semantic category of natural event types, may thus reflect a fundamental cortical organizational principle for mapping complex spectro-temporal attributes of natural sounds to perceptual representations, and ultimately supporting elements of acoustic communication symbolism and iconicity (Lotze et al., 2000; Monaghan et al., 2012; Perniss and Vigliocco, 2014; Lockwood and Dingemanse, 2015). Vocalizations, as another distinct acoustic-semantic category, have perhaps more obvious relationships between symbolism and oral communication, as addressed next.

Functional Roles of the aSTG

Animal vocalizations are arguably produced to convey intentional communication of some sort (e.g., distress, territorial warning, mating calls), which derive from motivational states thought to be similar across many species (Hauser, 2000; Bass et al., 2008; Lingle et al., 2012). The animal vocalizations in the present study had on average a significantly greater degree of perceived negative emotional valence in comparison to the relatively neutral valence of the animal action sounds (**Table 1**). The aSTG regions further showed parametric sensitivity to SSV measures, perceived intention and perceived recognition (**Figure 3A**, table), and also to the perceived emotional valence ratings (data not shown) which showed a correlation with SSV measures (**Supplemental Figure S3A**). Thus, the aSTG activation appeared to reflect both intermediate and higher stage processing of vocalizations with a functional role in representing acoustic signals containing prosodic information. Consistent with this interpretation, earlier studies have similarly reported differential activation to vocalizations conveying emotional content along the aSTG brain regions (Zatorre and Penhune, 2001; Kotz et al., 2003; Friederici and Alter, 2004; Ethofer et al., 2006; Schirmer and Kotz, 2006; Grossmann et al., 2010). Additional other brain regions commonly reported to be sensitive to emotional or prosodic cues in sound include the left anterior insula, for linking action representations to emotions (Carr et al., 2003; Wicker et al., 2003; Craig, 2009), and the left and right amygdala (Fecteau et al., 2007; Belin et al., 2008). Both of these latter regions were also preferentially activated by the vocalizations at lower threshold settings in the present study (data not shown). Therefore, the aSTG, together with a network of other brain regions, may have been extracting acoustic signal attributes that conveyed emotion and thus perhaps other forms of representation of intention behind the sound-sources. Thus, in addition to the roles that the bilateral mSTG/aSTG

have in spoken language processing (Binder et al., 2000), the present results speak to a more fundamental role in processing acoustic information that conveys information regarding the sound-source's intentions or emotional state, which is another rudimentary aspect of auditory communication thought to be critical to the evolution of proto-networks for spoken language systems in hominins (Falk, 2004).

CONCLUSIONS

In sum, the present study provides novel evidence of gross-level divergent, intermediate stage regions of auditory cortex for processing action sounds vs. vocalizations as distinct acoustic and acoustic-semantic categories of natural sounds. These pathways were consistent with neurocognitive models of embodied semantic knowledge representations and with a global theoretical framework for how ethologically valid categories of sounds may become processed as meaningful events by the brain. They were also consistent with established dual-stream (dorsal-ventral) models for auditory perception (Rauschecker and Scott, 2015); however, the "dorsal" pathways, notably including the posterior insulae of the present study, were not recruited for purposes of sound localization nor segmentation of spoken communication (language) sounds, but rather appeared to be recruited more for purposes of comparing the dynamics of spectro-temporal sound patterns to motor-related representations of actions, which may have served to help categorize and provide a sense of meaning (recognition or intentional understanding) behind the animal action sounds (Gazzola et al., 2006; Lahav et al., 2007; Pazzaglia et al., 2008). Further exploring auditory processing stages and neuronal mechanisms for representing natural sounds, including the roles of the posterior insulae relative to audio-motor association systems, will likely provide a promising approach for advancing anthropological models of language evolution. Moreover, such exploration is also likely to advance our understanding of the neurodevelopmental mechanisms underlying auditory communication, and thus impact models of acoustic signal processing in children with spoken language disorders.

ETHICS STATEMENT

All participants were consented by the laboratory principle investigator or his designated assistants, as describe in the West Virginia University approved IRB protocol (on file kc#1311130065). In short this entailed reading and signing a safety screening form for MRI safety, and a consent form, both of which are kept confidential.

AUTHOR CONTRIBUTIONS

JL conceived and designed the study, and JL and PW interpreted the data. PW, LS, CF, HS, and JL acquired the data. All contributed to analyzing the data and revising the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnins. 2016.00579/full#supplementary-material

Supplemental Figure S1 | Animal action sound (yellow) and vocalization (red) acoustic features and emotional valence ratings. (A) Chart illustrating differences in the average power spectra of the two categories of sound (from **Table 1**). **(B)** Chart illustrating the scale rating responses (n = 15) for the perceived emotional valence for each category of sound (red curve = 29 vocalizations; yellow curve = 23 animal action sounds). Button press ratings included very negative (−2), negative (−1), neutral (0), positive (+1), and very positive (+2). **(C)** Modulation power spectra for the retained 23 animal action sounds relative to 23 (of the 29) animal vocalizations. Note that the vocalizations qualitatively reveal relatively greater power at high spectral modulation rates. Refer to text for other details.

Supplemental Figure S2 | Results from a control condition scan wherein an individual performed the main auditory paradigm, using either their (A) right hand or **(B)** left hand for button press responses. Green cortical regions depict activation to both vocalization/index-finger and action sound/middle-finger response trial conditions relative to silence/no-response conditions (p(uncorr)<0.000001). Each trial response condition (for vocalizations and action sounds) resulted in primary motor cortex activation (green with outlines) only in the hemisphere opposite the hand used (axial slice at z = +62). Histograms show the degree of activation for the left and right hemisphere ROIs for both tasks.

Supplemental Figure S3 | Correlations between acoustic attributes and perceptual features. (A) Chart showing the correlation between emotional ratings of animal action sounds (yellow squares) and animal vocalization sounds (red diamonds) relative to spectral structure variation measures. **(B)** Chart showing the correlation between one's sense of perceiving the intention of the action sound or vocalization relative to the measure of entropy in the sound stimuli. Refer to Methods for further details.

Supplemental Figure S4 | Photos depicting non-human animals producing either action sounds (yellow outlines) or vocalizations (red). Colored brain regions (yellow and red) depict functionally distinct regions (using fMRI) that are preferential for processing the corresponding acoustic-semantic category of sound, colored respectively. Orange brain region depicts the functionally defined location of primary auditory cortex. These findings extend concepts from the ventral ("what is it") processing pathways for sound recognition, and support a general theoretical framework for how the mammalian brain may be organized to represent natural sounds as meaningful events to the listener.

REFERENCES

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236. doi: 10.1016/j.heares.2007. 01.011

Alluri, V., Toiviainen, P., Lund, T. E., Wallentin, M., Vuust, P., Nandi, A. K., et al. (2013). From Vivaldi to Beatles and back: predicting lateralized brain responses to music. *Neuroimage* 83, 627–636. doi: 10.1016/j.neuroimage.2013. 06.064

Altmann, C. F., Doehrmann, O., and Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cereb. Cortex* 17, 2601–2608. doi: 10.1093/cercor/bhl167

Andics, A., Gácsi, M., Faragó, T., Kis, A., and Miklósi, A. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Curr. Biol.* 24, 574–578. doi: 10.1016/j.cub.2014.01.058

Arbib, M. A. (2001). Co-evolution of human consciousness and language. *Ann. N. Y. Acad. Sci.* 929, 195–220. doi: 10.1111/j.1749-6632.2001.tb05717.x

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28, 105–124; discussion 125–167. doi: 10.1017/s0140525x05000038

Arbib, M. A. (2010). Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. *Brain Lang.* 112, 12–24. doi: 10.1016/j.bandl.2009.10.001

Arnott, S. R., Cant, J. S., Dutton, G. N., and Goodale, M. A. (2008). Crinkling and crumpling: an auditory fMRI study of material properties. *Neuroimage* 43, 368–378. doi: 10.1016/j.neuroimage.2008.07.033

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91. doi: 10.1016/S1364-6613(02)00029-3

Bass, A. H., Gilland, E. H., and Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. *Science* 321, 417–421. doi: 10.1126/science.1157632

Beer, A. L., Plank, T., Meyer, G., and Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Front. Integr. Neurosci.* 7:5. doi: 10.3389/fnint.2013.00005

Beisteiner, R., Windischberger, C., Lanzenberger, R., Edward, V., Cunnington, R., Erdler, M., et al. (2001). Finger somatotopy in human motor cortex. *Neuroimage* 13(6 Pt 1), 1016–1026. doi: 10.1006/nimg.2000.0737

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., and Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proc. Biol. Sci.* 275, 473–481. doi: 10.1098/rspb.2007.1460

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078

Bendor, D., and Wang, X. (2008). Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *J. Neurophysiol.* 100, 888–906. doi: 10.1152/jn.00884.2007

Bidet-Caulet, A., Voisin, J., Bertrand, O., and Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage* 28, 132–139. doi: 10.1016/j.neuroimage.2005.06.018

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796. doi: 10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., and Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn. Sci.* 19, 142–150. doi: 10.1016/j.tics.2014.12.008

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404. doi: 10.1046/j.1460-9568.2001.01385.x

Burton, M. W., Small, S. L., and Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690. doi: 10.1162/089892900562309

Cai, Q., Van der Haegen, L., and Brysbaert, M. (2013). Complementary hemispheric specialization for language production and visuospatial attention. *Proc. Natl. Acad. Sci. U.S.A.* 110, E322–E330. doi: 10.1073/pnas.1212956110

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249. doi: 10.1093/cercor/bhi007

Canessa, N., Borgo, F., Cappa, S. F., Perani, D., Falini, A., Buccino, G., et al. (2007). The Different neural correlates of action and functional knowledge in semantic memory: an fMRI study. *Cereb. Cortex* 18, 740–751. doi: 10.1093/cercor/bhm110

Capitani, E., Laiacona, M., Mahon, B., and Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn. Neuropsychol.* 20, 213–261. doi: 10.1080/02643290244000266

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502. doi: 10.1073/pnas.0935845100

Cerri, G., Cabinio, M., Blasi, V., Borroni, P., Iadanza, A., Fava, E., et al. (2015). The mirror neuron system and the strange case of Broca's area. *Hum. Brain Mapp.* 36, 1010–1027. doi: 10.1002/hbm.22682

Chouchourelou, A., Golden, A., and Shiffar, M. (2012). *What Does "Biological Motion" Really Mean? Differentiating Visual Percepts of Human, Animal and Non-Biological Motion.* Oxford: Oxford University Press.

Cohen, Y. E., Theunissen, F., Russ, B. E., and Gill, P. (2007). Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol.* 97, 1470–1484. doi: 10.1152/jn.00769.2006

Corballis, M. C. (1989). Laterality and human evolution. *Psychol. Rev.* 96, 492–505. doi: 10.1037/0033-295X.96.3.492

Corballis, M. C. (1999). The gestural origins of language. *Am. Sci.* 87, 138–145. doi: 10.1511/1999.2.138

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Craig, A. D. (2009). How do you feel–now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

Desai, R., Liebenthal, E., Possing, E. T., Waldron, E., and Binder, J. R. (2005). Volumetric vs. surface-based alignment for localization of auditory cortex activation. *Neuroimage* 26, 1019–1029. doi: 10.1016/j.neuroimage.2005.03.024

DeWitt, I., and Rauschecker, J. P. (2013). Wernicke's area revisited: parallel streams and word processing. *Brain Lang.* 127, 181–191. doi: 10.1016/j.bandl.2013.09.014

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., and Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* 46, 2776–2786. doi: 10.1016/j.neuropsychologia.2008.05.011

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7900–7905. doi: 10.1073/pnas.1602413113

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5:e1000302. doi: 10.1371/journal.pcbi.1000302

Engel, L. R., Frum, C., Puce, A., Walker, N. A., and Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage* 47, 1778–1791. doi: 10.1016/j.neuroimage.2009.05.041

Engelien, A., Tüscher, O., Hermans, W., Isenberg, N., Eidelberg, D., Frith, C., et al. (2006). Functional neuroanatomy of non-verbal semantic sound processing in humans. *J. Neural Transm.* 113, 599–608. doi: 10.1007/s00702-005-0342-0

Ethofer, T., Pourtois, G., and Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Prog. Brain Res.* 156, 345–361. doi: 10.1016/S0079-6123(06)56019-4

Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* 27, 491–503. doi: 10.1017/S0140525X04000111

Fecteau, S., Armony, J. L., Joanette, Y., and Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23, 840–848. doi: 10.1016/j.neuroimage.2004.09.019

Fecteau, S., Belin, P., Joanette, Y., and Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage* 36, 480–487. doi: 10.1016/j.neuroimage.2007.02.043

Fiebach, C. J., Friederici, A. D., Müller, K., von Cramon, D. Y., and Hernandez, A. E. (2003). Distinct brain representations for early and late learned words. *Neuroimage* 19, 1627–1637. doi: 10.1016/S1053-8119(03)00227-1

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647. doi: 10.1002/mrm.1910330508

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40, 859–869. doi: 10.1016/S0896-6273(03)00669-X

Friederici, A. D., and Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang.* 89, 267–276. doi: 10.1016/S0093-934X(03)00351-1

Friederici, A. D., Kotz, S. A., Scott, S. K., and Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum. Brain Mapp.* 31, 448–457. doi: 10.1002/hbm.20878

Galati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., et al. (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage* 40, 1274–1286. doi: 10.1016/j.neuroimage.2007.12.044

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/S1364-6613(98)01262-5

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23, 344–357. doi: 10.1016/j.neuroimage.2004.06.004

Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Curr. Biol.* 16, 1824–1829. doi: 10.1016/j.cub.2006.07.072

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., and Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cereb. Cortex* 23, 2025–2037. doi: 10.1093/cercor/bhs162

Giordano, B. L., McDonnell, J., and McAdams, S. (2010). Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds. *Brain Cogn.* 73, 7–19. doi: 10.1016/j.bandc.2010.01.005

Grafton, S. T., Fadiga, L., Arbib, M. A., and Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* 6, 231–236. doi: 10.1006/nimg.1997.0293

Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858. doi: 10.1016/j.neuron.2010.03.001

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.* 502, 924–952. doi: 10.1002/cne.21326

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C., and Büchel, C. (2003). The human action recognition system and its relationship to Broca's area: an fMRI study. *Neuroimage* 19, 637–644. doi: 10.1016/S1053-8119(03)00087-9

Hamzei, F., Vry, M.-S., Saur, D., Glauche, V., Hoeren, M., Mader, I., et al. (2016). The dual-loop model and the human mirror neuron system: an exploratory Combined fMRI and DTI study of the inferior frontal gyrus. *Cereb Cortex* 26, 2215–2224. doi: 10.1093/cercor/bhv066

Han, Z., Bi, Y., Chen, J., Chen, Q., He, Y., and Caramazza, A. (2013). Distinct regions of right temporal cortex are associated with biological and human-agent motion: functional magnetic resonance imaging and neuropsychological evidence. *J. Neurosci.* 33, 15442–15453. doi: 10.1523/JNEUROSCI.5868-12.2013

Hauser, M. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cogn. Sci.* 24, 445–475. doi: 10.1207/s15516709cog2403_5

Herdener, M., Esposito, F., Scheffler, K., Schneider, P., Logothetis, N. K., Uludag, K., et al. (2013). Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex* 49, 2822–2833. doi: 10.1016/j.cortex.2013.04.003

Hewes, G. W. (1992). Primate communication and the gestural origin of language. *Curr. Anthropol.* 33, 65–84. doi: 10.1086/204019

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hillis, A. E., and Caramazza, A. (1991). Category-specific naming and comprehension impairment: a double dissociation. *Brain* 114 (Pt 5), 2081–2094. doi: 10.1093/brain/114.5.2081

Hopfield, J. J., and Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biol. Cybern.* 52, 141–152.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., et al. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE* 10:e0116494. doi: 10.1371/journal.pone.0116494

Indovina, I., and Sanes, J. N. (2001). Combined visual attention and finger movement effects on human brain representations. *Exp. Brain Res.* 140, 265–279.

Kaas, J. H., and Hackett, T. A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiol. Neurootol.* 3, 73–85. doi: 10.1159/000013783

Kaas, J. H., and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799. doi: 10.1073/pnas.97.22.11793

Keysers, C., Kohler, E., Umiltá, A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* 153, 628–636. doi: 10.1007/s00221-003-1603-5

Kohler, E., Keysers, C., Umiltá, A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 297, 846–848. doi: 10.1126/science.1070311

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., and Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang.* 86, 366–376. doi: 10.1016/S0093-934X(02)00532-1

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., and Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Comput. Biol.* 3:e100. doi: 10.1371/journal.pcbi.0030100

Kuśmierek, P., Ortiz, M., and Rauschecker, J. P. (2012). Sound-identity processing in early areas of the auditory ventral stream in the macaque. *J. Neurophysiol.* 107, 1123–1141. doi: 10.1152/jn.00793.2011

Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* 27, 308–314. doi: 10.1523/JNEUROSCI.4822-06.2007

Larsson, M. (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Anim. Cogn.* 17, 1–14. doi: 10.1007/s10071-013-0678-z

Larsson, M. (2015). Tool-use-associated sound in the evolution of language. *Anim. Cogn.* 18, 993–1005. doi: 10.1007/s10071-015-0885-x

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nat. Neurosci.* 5, 356–363. doi: 10.1038/nn831

Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., and DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* 25, 5148–5158. doi: 10.1523/JNEUROSCI.0419-05.2005

Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., and DeYoe, E. A. (2006). Lefties get it "right" when hearing tool sounds. *J. Cogn. Neurosci.* 18, 1314–1330. doi: 10.1162/jocn.2006.18.8.1314

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., and Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *J. Cogn. Neurosci.* 23, 2079–2101. doi: 10.1162/jocn.2010.21570

Lewis, J. W., Talkington, W. J., Tallaksen, K. C., and Frum, C. A. (2012). Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. *Front. Syst. Neurosci.* 6:27. doi: 10.3389/fnsys.2012.00027

Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., et al. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J. Neurosci.* 29, 2283–2296. doi: 10.1523/JNEUROSCI.4145-08.2009

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., and Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Curr. Zool.* 58, 698–726. doi: 10.1093/czoolo/58.5.698

LoCasto, P. C., Krebs-Noble, D., Gullapalli, R. P., and Burton, M. W. (2004). An fMRI investigation of speech and tone segmentation. *J. Cogn. Neurosci.* 16, 1612–1624. doi: 10.1162/0898929042568433

Lockwood, G., and Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Front. Psychol.* 6:1246. doi: 10.3389/fpsyg.2015.01246

Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., and Grodd, W. (2000). fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage* 11, 473–481. doi: 10.1006/nimg.2000.0556

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–511; discussion 511–446. doi: 10.1017/s0140525x98001265

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–816. doi: 10.1006/nimg.2001.0888

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45. doi: 10.1146/annurev.psych.57.102904.190143

McNamara, A., Buccino, G., Menz, M. M., Gläscher, J., Wolbers, T., Baumgärtner, A., et al. (2008). Neural dynamics of learning sound-action associations. *PLoS ONE* 3:e3845. doi: 10.1371/journal.pone.0003845

Mednick, S. C., Arman, A. C., and Boynton, G. M. (2005). The time course and specificity of perceptual deterioration. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3881–3885. doi: 10.1073/pnas.0407866102

Medvedev, A. V., and Kanwal, J. S. (2004). Local field potentials and spiking activity in the primary auditory cortex in response to social calls. *J. Neurophysiol.* 92, 52–65. doi: 10.1152/jn.01253.2003

Meyer, G. F., Greenlee, M., and Wuerger, S. (2011). Interactions between auditory and visual semantic stimulus classes: evidence for common processing networks for speech and body actions. *J. Cogn. Neurosci.* 23, 2291–2308. doi: 10.1162/jocn.2010.21593

Mikl, M., Marecek, R., Hlustík, P., Pavlicová, M., Drastich, A., Chlebus, P., et al. (2008). Effects of spatial smoothing on fMRI group inferences. *Magn. Reson. Imaging* 26, 490–503. doi: 10.1016/j.mri.2007.08.006

Miller, E. K., Nieder, A., Freedman, D. J., and Wallis, J. D. (2003). Neural correlates of categories and concepts. *Curr. Opin. Neurobiol.* 13, 198–203. doi: 10.1016/S0959-4388(03)00037-0

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neurosci. Biobehav. Rev.* 36, 341–349. doi: 10.1016/j.neubiorev.2011.07.004

Monaghan, P., Mattock, K., and Walker, P. (2012). The role of sound symbolism in language learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1152–1164. doi: 10.1037/a0027747

Moore, C. J., and Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 122, 943–962. doi: 10.1093/brain/122.5.943

Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., and Ball, T. (2007). A rapid sound-action association effect in human insular cortex. *PLoS ONE* 2:e259. doi: 10.1371/journal.pone.0000259

O'Connell, N., McGinnis, T., Mills, A., Lakatos, P., Lipton, M., and Branch, C., C. E. S. (2004). *Spatiotemporal Dynamics of Unisensory and Multisensory Processes in Neocortex.* Wasington, DC: Society for Neuroscience..

Ohnaka, K., Sakurai, Y., Fuse, S., Shimpo, T., and Kaga, K. (1995). [Pure word deafness after cerebral hemorrhage in the left temporal lobe: a case report]. *Rinsho Shinkeigaku* 35, 290–295.

Pazzaglia, M., Pizzamiglio, L., Pes, E., and Agliota, S. M. (2008). The sound of actions in apraxia. *Curr. Biol.* 18, 1766–1772. doi: 10.1016/j.cub.2008.09.061

Peirce, J. W. (2007). PsychoPy–Psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* lx, 389–443. doi: 10.1093/brain/60.4.389

Peretz, I., and Zatorre, R. J. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114. doi: 10.1146/annurev.psych.56.091103. 070225

Perlovsky, L. (2011). Language and cognition interaction neural mechanisms. *Comput. Intell. Neurosci.* 2011:454587. doi: 10.1155/2011/454587

Perniss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130300. doi: 10.1098/rstb.2013.0300

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., and Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374. doi: 10.1038/nn2043

Pollak, G. D., and Bodenhamer, R. D. (1981). Specialized characteristics of single units in inferior colliculus of mustache bat: frequency representation, tuning, and discharge patterns. *J. Neurophysiol.* 46, 605–620.

Preuss, T. M. (2011). The human brain: rewired and running hot. *Ann. N. Y. Acad. Sci.* 1225(Suppl. 1), E182–E191. doi: 10.1111/j.1749-6632.2011.06001.x

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 13, 669–683. doi: 10.1006/nimg.2000.0714

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiol. Neurootol.* 3, 86–103. doi: 10.1159/000013784

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Rauschecker, J. P., and Scott, S. K. (2015). "Pathways and streams in the auditory cortex," in *Neurobiology of Language,* eds G. Hickok and S. L. Small (Elsevier), 287–298. Available online at: https://books.google.com/books?id=9c2cBAAAQBAJ&dq=Pathways+and+streams+in+the+auditory+cortex+Rauschecker&source=gbs_navlinks_s

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Rauschecker, J. P., Tian, B., and Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268, 111–114. doi: 10.1126/science.7701330

Reddy, R. K., Ramachandra, V., Kumar, N., and Singh, N. C. (2009). Categorization of environmental sounds. *Biol. Cybern.* 100, 299–306. doi: 10.1007/s00422-009-0299-4

Rehme, A. K., Eickhoff, S. B., and Grefkes, C. (2013). State-dependent differences between functional and effective connectivity of the human cortical motor system. *Neuroimage* 67, 237–246. doi: 10.1016/j.neuroimage.2012. 11.027

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci.* 21, 188–194. doi: 10.1016/S0166-2236(98)01260-0

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203. 144230

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* 2, 1131–1136. doi: 10.1038/16056

Rose, G., and Capranica, R. R. (1983). Temporal selectivity in the central auditory system of the leopard frog. *Science* 219, 1087–1089. doi: 10.1126/science.6600522

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., et al. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* 10:e1003412. doi: 10.1371/journal.pcbi.1003412

Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009

Schönwiesner, M., and Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14611–14616. doi: 10.1073/pnas.0907682106

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Pt 12), 2400–2406. doi: 10.1093/brain/123.12.2400

Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech–a positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083. doi: 10.1121/1.2216725

Shallice, T. (1988). *From Neuropsychology to Mental Structure.* Cambridge: Cambridge University Press.

Suga, N. (1965). Functional properties of auditory neurones in the cortex of echo-locating bats. *J. Physiol.* 181, 671–700. doi: 10.1113/jphysiol.1965.sp007791

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme Medical Publishers.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., and Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J. Neurophysiol.* 91, 1282–1296. doi: 10.1152/jn.01125.2002

Talkington, W. J., Rapuano, K. M., Hitt, L., Frum, C. A., and Lewis, J. W. (2012). Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. *Jo. Neurosci.* 32, 8084–8093. doi: 10.1523/JNEUROSCI.1118-12.2012

Talkington, W. J., Taglialatela, J. P., and Lewis, J. W. (2013). Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. *Hear. Res.* 305, 74–85. doi: 10.1016/j.heares.2013.08.009

Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291, 2564–2569. doi: 10.1126/science.1058522

Thierry, G., Giraud, A. L., and Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron* 38, 499–506. doi: 10.1016/S0896-6273(03)00199-5

Turkeltaub, P. E., and Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain Lang.* 114, 1–15. doi: 10.1016/j.bandl. 2010.03.008

Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28, 635–662. doi: 10.1016/j.neuroimage.2005.06.058

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* 8, 443–459. doi: 10.1136/jamia.2001.0080443

Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairments. *Brain* 107 (Pt 3), 829–854. doi: 10.1093/brain/107.3.829

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Woods, D. L., Herron, T. J., Cate, A. D., Kang, X., and Yund, E. W. (2011). Phonological processing in human auditory cortical fields. *Front. Hum. Neurosci.* 5:42. doi: 10.3389/fnhum.2011.00042

Woolley, S. M., Fremouw, T. E., Hsu, A., and Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* 8, 1371–1379. doi: 10.1038/nn1536

Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849. doi: 10.1126/science.1589767

Zatorre, R. J., and Penhune, V. B. (2001). Spatial localization after excision of human auditory cortex. *J. Neurosci.* 21, 6321–6328.