Published in final edited form as:
Hear Res. 2013 November ; 305: . doi:10.1016/j.heares.2013.08.009.

Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates

William J. Talkington[1], **Jared P. Taglialatela**[2], and **James W. Lewis**[1]

[1]Department of Neurobiology & Anatomy, Sensory Neuroscience Research Center, and Center for Advanced Imaging, West Virginia University, Morgantown, WV26506, USA

[2]Department of Biology and Physics, Kennesaw State University, Kennesaw, Georgia, USA

Abstract

Humans and several non-human primates possess cortical regions that are most sensitive to vocalizations produced by their own kind (conspecifics). However, the use of speech and other broadly defined categories of behaviorally relevant natural sounds has led to many discrepancies regarding where voice-sensitivity occurs, and more generally the identification of cortical networks, "proto-networks" or protolanguage networks, and pathways that may be sensitive or selective for certain aspects of vocalization processing. In this prospective review we examine different approaches for exploring vocal communication processing, including pathways that may be, or become, specialized for conspecific utterances. In particular, we address the use of naturally produced non-stereotypical vocalizations (mimicry of other animal calls) as another category of vocalization for use with human and non-human primate auditory systems. We focus this review on two main themes, including progress and future ideas for studying vocalization processing in great apes (chimpanzees) and in very early stages of human development, including infants and fetuses. Advancing our understanding of the fundamental principles that govern the evolution and early development of cortical pathways for processing non-verbal communication utterances is expected to lead to better diagnoses and early intervention strategies in children with communication disorders, improve rehabilitation of communication disorders resulting from brain injury, and develop new strategies for intelligent hearing aid and implant design that can better enhance speech signals in noisy environments.

Keywords

hearing perception; spoken language; protolanguage; evolution; proto-networks; fMRI

1. Introduction

Vocalizations represent some of the most complex sounds of the natural world. The acoustic signals of even very short utterances can be rapidly processed to extract distinct meaning. This can include alerting the listener to danger, a mate, or food. More specific socially

Contact author: James W. Lewis, PhD., Center for Advanced Imaging, Sensory Neuroscience Research Center, Department of Neurobiology & Anatomy, P.O. Box 9128, West Virginia University, Morgantown, WV 26506-9128, Phone: 304 293-1517, Fax: 304-293-3850, jwlewis@hsc.wvu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

relevant information such as the identity of the source (e.g. species, gender, or specific individual), its intent, health status, or emotional state in some instances can also quickly be surmised. The auditory systems of vocalizing mammals, notably including humans and non-human primates, develop to rapidly decompose incoming vocal communication signals (on second or sub-second timescales), utilizing multiple hierarchical processing stages from the brainstem to higher order auditory cortices.

The information gleaned from these acoustic analyses leads to recruitment of other cortical regions and subsequently engenders responses, ranging from conspecific recognition, to attention modification, to evasive motor responses. Early auditory circuits must rapidly filter incoming signals for the most behaviorally-relevant content while simultaneously suppressing more irrelevant information (background “noise”, contextually unimportant vocalizations, etc.). Stimuli with very subtle acoustic variations can impart drastically different meaning; human language faculties arguably represent the most salient examples of this property – slight changes in pitch articulation can cause a speech segment to be perceived, for example, as sad, angry, or fearful.

Much of the early work that described mammalian auditory systems effectively incorporated the use of acoustically “simple” stimuli, including pure tones, band-pass noise, amplitude-modulated tones, and harmonic complexes. The major benefits of using simpler stimuli to elucidate the signal processing architecture and function of these networks are obvious; simple sound stimuli permit the design of very exact and controlled experimental manipulations that produce physiological responses with greater interpretational power. While these experiments have garnered crucial information about the function of hierarchically organized auditory networks, they generally do not reflect the nature of sounds that are experienced in real-world situations. Naturalistic sounds like vocalizations can be composed of extremely nuanced combinations of acoustic phenomenon. This quality of vocalization signals becomes especially problematic when attempting to design precise experiments that can produce generalizable results. Probing the auditory system with natural (often behaviorally relevant) stimuli, which it has arguably become optimized to process, not only gives insight into its respective operation, but is likely to reveal more cross-modal “whole-brain” physiological and behavioral responses.

While canonical auditory regions (i.e. primary auditory cortices (PAC), belt, and parabelt regions) obviously play fundamental roles in vocalization processing, the influence of other “non-auditory” cortical and cognitive systems are increasingly becoming necessary to consider and model when examining how the auditory system extracts or derives meaningful representations of naturalistic vocalization stimuli. For instance, vocalizations often lead to assessments of the intentional and emotional states of other animals (conspecific or otherwise). This implies that the processing or perception of vocal communication sounds can readily tap into rather widespread sets of cortical networks and pathways, including motor-related systems such as mirror neuron systems (Rizzolatti et al., 1996), networks associated with tool use (Lewis, 2006), and into interoceptive systems, including posterior and anterior insular regions (Craig, 2009) that are thought to be involved in representing internal states of self, “feelings of knowing”, and meta-representations of self (Herbert and Pollatos, 2012).

This prospective review considers two major themes regarding how neuronal networks that mediate hearing perception may become established in the brain. First, we presume that the communication abilities and processing subtleties that humans have with non-speech utterances are present in some capacity in non-human primates (Goodall, 1986, Fitch, 2011). Thus, one theme will be to examine various studies of the whole-brain cortical processing of natural vocalizations across primate species, especially great apes, using either functional magnetic resonance imaging (fMRI) or positron emission tomography (PET). One hypothesis we pursue is that right hemisphere pathways may have developed to process the spectro-temporal elements that are characteristic of various categories of non-locutionary vocal utterances as well as the prosodic components of “motherese”, and that these processing pathways may represent an ideal link to further explore protolanguage networks across primate species. Furthermore, a listener’s ability to extract “meaning” from both linguistic and non-linguistic communicative utterances ostensibly requires extensive periods (years) of learning. Since much of the foundation of processing pathways and networks that are ultimately recruited for sound perception may develop in early stages of life, a second theme of this review will be an analysis of fetal to early childhood human neurodevelopment in vocal signal processing. Based on commonalities between these two themes, we address future research directions (for both human and non-human primates) regarding how the brain establishes cortical networks and potential intermediate stage “proto-networks” that ultimately subserve hearing perception, with a focus on communicative utterance processing. We propose that human infants and young or adult non-human primates (especially great apes) may be similar in some aspects related to their use and processing of non-locutionary acoustic signals. Furthermore, there may be differences in the cortical stages used for processing these “earlier” forms of communication utterances when compared to those that are used in the far more extensively studied human adults, who have had many years of experience utilizing full-blown human language and benefit from more advanced cognitive and neurolinguistic development.

We feel that together the above two themes of basic scientific research will facilitate significant advances in our understanding of fundamental mechanisms that mediate hearing perception. This in turn will contribute to a better understanding of vocal communication and speech perception development, leading to more targeted therapeutic interventions or treatments for children with impairments, rehabilitation of patients with aphasia following brain injury, and contribute to methods for developing more intelligent biologically-inspired hearing prosthetics and acoustic signal pre-processing algorithms. In the following sections we provide a brief background regarding human cortical regions that are sensitive to human (conspecific) voice (Section 2) and low-level acoustic signal processing of vocalization features (Section 3), and subsequently provide a rationale for embarking on future studies using various “non-locutionary utterances” as representing potential gross-level categories of vocalizations in both human and non-human primates (Section 4). This is followed by a prospective review of non-human primate vocalization processing studies (Section 5) and human studies involving the development of the auditory system at very early stages of life (Section 6). Together, the latter two sections (two themes) strongly advocate the continuation of scientific exploration of both the evolution and early neurodevelopment of putative proto-networks from the perspective of hearing perception, in that proto-networks (or protolanguage networks) may develop similarly (up to a point) to mediate symbolic representations of knowledge and communicative content conveyed through non-locutionary utterances.

2. Human brain regions sensitive to human voice Editor to add Poremba Reference in this section

Cortical regions sensitive to human (conspecific) voice have been traditionally identified bilaterally within the superior temporal sulci (STS; e.g. see Fig. 1), based largely on measuring cortical responses (using fMRI or PET) to stereotypical sounds, including speech, animal vocalizations, environmental sounds, and non-verbal sounds such as coughs, sighs, and moans (Belin et al., 2000, Fecteau et al., 2004b, Lewis et al., 2009). However, more recent studies consider these STS regions to represent “higher-order” auditory cortices that function more generally as substrates for auditory experience, showing activity to artificially constructed non-vocal sounds after perceptual training (Leech et al., 2009, Liebenthal et al., 2010). The STS regions may not function in a domain-specific manner solely for vocalization processing. Rather, humans may typically develop to become “experts” at processing voice and speech signals, which consequently leads to recruitment and development of circuits (metamodal operators) in the STS that compete to process those signals in a domain-general manner (Pascual-Leone and Hamilton, 2001). Additionally, some of the vocalizations and natural sounds used as control conditions for revealing voice sensitivity in previous studies have included broadly defined acoustically meaningful “categories” that may be confounded by more subtle sub-categories upon which the auditory system is more fundamentally organized or optimized to process.

A recent study used a novel class of *non-stereotypical* vocalization sounds, human-mimicked animal vocalizations relative to the corresponding animal vocalizations themselves, to probe human brain regions sensitive to sound utterances produced by the human vocal tract (Talkington et al., 2012). This non-locutionary category (see Section 4 for utterance types) of human (conspecific) vocalization allowed researchers to probe intermediate cortical networks that showed fine-grained sensitivities to the acoustic subtleties of human voice. Using fMRI, they reported a left hemisphere dominant activation to naturally produced sounds unique to the human *conspecific* vocal tract (Figure 1, yellow). This result was contrary to previous findings that described human-voice or species-specific sensitivity as a right hemisphere dominant or bilateral function (Belin et al., 2000, Fecteau et al., 2004a). Moreover, their results suggest that the cortical pathways supporting vocalization perception are initially organized by sensitivity to the human vocal tract in regions prior to the left STS (in terms of the spatial processing hierarchy radiating out laterally from primary auditory cortices along the cortical ribbon). No regions showed greater responses to the reverse contrast (animal vocalizations > human mimics of those sounds). This and other studies have examined different categories of calls, either across or within primate species, and have revealed cortices showing sensitivity to specific bottom-up “lower-level” acoustic signal attributes that may be inherent to communicative utterances or subclasses therein. Such signal attributes may thus serve as computational primitives (spectral or spectro-temporal feature combinations) that early and intermediate stages of the auditory system may use to rapidly process behaviorally relevant sounds. Thus, we next address some relatively simple acoustic features that are characteristic of both human and non-human vocalizations.

3. Low-level acoustic signal processing of vocalizations

Where in the auditory processing hierarchies does vocalization-sensitivity begin to emerge? Numerous “simple” acoustic signal attributes are known, or thought, to be represented in early cortical processing stages, including the filtering or extraction of signal features such as bandwidths, spectral shapes, onsets, and harmonic relationships, which together have a critical role in auditory stream segregation and formation, clustering operations, and sound organization (Medvedev et al., 2002, Nelken, 2004, Kumar et al., 2007, Elhilali and Shamma, 2008, Woods et al., 2010). Later stages are thought to represent processing that segregates spectro-temporal patterns associated with complex sounds, including the processing of acoustic textures, location cues, prelinguistic analysis of speech sounds (Griffiths and Warren, 2002, Obleser et al., 2007, Overath et al., 2010), and representations of auditory objects defined by their entropy and spectral structure variation (Reddy et al., 2009, Lewis et al., 2012). Subsequent cortical processing pathways, such as projections between posterior portions of the superior temporal gyri (STG) and STS, may integrate corresponding acoustic streams over longer time frames (Maeder et al., 2001, Zatorre et al., 2004, Griffiths et al., 2007, Leech et al., 2009, Goll et al., 2011, Teki et al., 2011), involving or leading to processing that may provide a greater sense of semantic meaning to the listener.

Sounds containing strong harmonic content, notably including human and animal vocalizations, evoke bilateral activity along various portions of the superior temporal plane and STG (cf. Figures 1 and 2, color progressions), which subsequently feed into regions that are relatively specialized for processing speech and/or prosodic information (Zatorre et al., 1992, Obleser et al., 2008, Lewis et al., 2009, Rauschecker and Scott, 2009, Leaver and Rauschecker, 2010, Talkington et al., 2012). Studies of vocal and song call processing in birds and lower mammals have further added to our understanding of low-level build-up of receptive fields that represent species-specific information (Fitzpatrick et al., 1993, Lewicki and Konishi, 1995, Medvedev et al., 2002, Kumar et al., 2007). This includes the development of spectro-temporal template models of auditory function, which posit that there exists an increasingly complex hierarchy of “templates” for processing specific sounds or classes of sounds. Each subsequent stage represents another level of processing that likely combines numerous inputs from earlier and parallel stages.

In humans, sensitivity to harmonic content (defined quantitatively by a harmonics-to-noise ratio; HNR) has been interpreted in the context of such models. For instance, Lewis et al. reported cortical regions showing parametric sensitivity to the HNR values of artificially constructed iterated rippled noises (IRNs; Figure 2, green) and of animal vocalizations (blue) (Lewis et al., 2009). These stages of HNR-sensitivity in humans were juxtaposed between tonotopically-defined regions (yellow; reflecting an organization derived initially at the cochlea) and STS regions sensitive to comprehended locutionary speech (purple). Conceivably, specific combinations of tonotopic outputs could converge to form cortical networks that are sensitive to harmonic qualities (i.e. computational primitives), representing intermediate stages of vocalization processing. Interestingly, different conceptual categories of animal calls and utterances could in part be grouped based on harmonicity (HNR values) of different types of vocalizations (Figure 3; ovals and rectangles). Collectively, the results from studies identifying how bottom-up acoustic signals are processed will likely continue to impact the design of intelligent hearing aids and implants, which may enhance or retain relevant signal features or computational primitives relative to background acoustic noise (Coath et al., 2005, Coath and Denham, 2005, Coath et al., 2008, Rumberg et al., 2008, Shannon, 2012). However, future searches for voice-sensitive regions or regions “selective” for processing conspecific vocalizations (e.g. Fig. 1) may critically depend on the specific category or sub-category of vocalization sound(s) under consideration, as well as specific task factors reflecting how the vocal information is to be used by the listener. This brings us to the following section regarding a rationale for future studies of vocalization processing.

4. Rationale for studying utterances, paralinguistic signals and non-speech sounds

During the last 50–150 thousand years of human evolution, spoken language perception arguably became the most important function for the human auditory system. According to one set of linguistic theories, speech acts and utterances can be divided into four main categories that may represent rudimentary elements of spoken communication (Austin, 1975). This includes *locutionary* expressions, which convey semantic information through the use of many different acoustic signal forms ranging from simple phonemes, single words, and to grammatically complex word combinations and phrases used in the 6000 or so language systems currently spoken on our planet (McWhorter, 2004). However, humans commonly produce and hear other forms of communicative utterances on a daily basis, involving non-locutionary utterances and paralinguistic signals that transcend language boundaries. This includes *phatic* expressions that convey social information such as emotional status (e.g. a wince revealing pain, a moan of joy, or a sigh of relief), *perlocutionary* expressions that are intended to cause a desired psychological consequence to a listener (e.g. the tone in voice or grunt to persuade "stay out", scare, convince, inspire, surprise, or mislead), and *illocutionary* expressions wherein the fact that an utterance is even being performed itself conveys the communication that the speaker will undertake an obligation (e.g. a promising "uh, huh", a greeting "what's up?", warning "look out", ordering, congratulating, etc.). Given that the auditory system must develop to accommodate processing to ultimately convey a sense of meaningfulness behind the complex vocalizations and utterances of others, social interactions of humans or non-human primates (and the evolutionary ancestry of hominins in general) are likely to be critical for normal development of how acoustic cues become most efficiently processed (Falk, 2004). Germane to this idea is the hypothesis that social interaction of humans is essential for many levels of natural speech learning (Kuhl, 2007), which is a topic resumed in Section 6.

The use of spoken language as sound stimuli to examine acoustic signal processing and general mechanisms mediating hearing and speech perception has encountered a number of hurdles. For instance, congenitally deaf individuals can also readily acquire locutionary communication skills in the form of sign language and written language. Thus, there are likely to be a myriad of processing stages and hierarchies that link the complexities of language processing networks (e.g. rules of grammar) with those more central to sound perception *per se*. In other words, vocal communication and language systems need not be mutually dependent on one another. While humans are the only species known to fully process, extract, and comprehend locutionary acoustic information (grammatically constructed language), other vocalizing social species, such as monkeys and great apes, presumably have evolved to utilize some of these other non-locutionary classes of utterances, which may be essential for effective social communication. Thus, to understand more basic cortical mechanisms for processing communicative vocalizations and possible foundations underlying spoken language systems, researchers have increasingly been investigating the processing of non-verbal vocalizations and paralinguistic signals (e.g. calls, grunts, coughs, sighs, etc.) not only in humans, but also in non-human primates. The use of non-locutionary utterances as behaviorally relevant, "naturalistic" stimuli permits more direct comparisons across species, wherein responses to identical sets of sound stimuli having more similar levels of meaningfulness can be measured. This topic is addressed next (Section 5) as the first major theme of this prospective review.

5. Vocalization processing in non-human primates

The roots of human auditory structures, function, and skill may be investigated best by examining homology that exists in closely related species. This is not to imply that the anatomy and physiology of closely related extant species represent the exact conditions that were necessary to form the bases of human function, as these species have also continued to evolve away from their precursors concomitantly with humans. Nonetheless, given their evolutionary proximity to humans, data concerning the structures and pathways that are involved in the perception and processing of naturalistic vocalizations in non-human primates, notably Old World primates and great apes, are particularly relevant to discussions of human language origins.

Old World primates

Recently, functional neuroimaging techniques such as fMRI and PET have been employed to visualize activity during passive listening to conspecific vocalizations in macaque monkeys (Gil-da-Costa et al., 2004, Poremba et al., 2004, Gil-da-Costa et al., 2006, Petkov et al., 2008). However, a relatively early study with Japanese macaques, *Macaca fuscata*, provided the foundation for subsequent functional neuroimaging work (Heffner and Heffner, 1984). Here, the researchers evaluated the monkeys' ability to discriminate between two variants of a species-specific call type before and after unilateral or bilateral ablation of the STG. Whereas the discriminative performance of monkeys who sustained unilateral ablation of the right STG was unaffected, those with unilateral ablation of the left STG were temporarily unable to complete the auditory discrimination task, though did subsequently regain their discriminative abilities. Importantly, those individuals who received subsequent bilateral ablations never recovered their discriminative ability. The authors concluded that perception of species-typical vocalizations is mediated in the STG with the left hemisphere playing a predominant role (Heffner and Heffner, 1984).

More recently, Poremba, et al., (2004) utilized PET to determine if increased neuronal metabolic activity is observed in the STG of rhesus monkeys during passive listening to a variety of auditory stimuli including conspecific vocalizations. The authors reported that only rhesus monkey vocalizations (and not phase-scrambled conspecific vocalizations, human vocalizations, ambient background noise, or environmental sounds) resulted in significantly greater metabolic activity in the left dorsal temporal pole of the STG. In a second study, Gil-da-Costa and colleagues (2006) similarly utilized PET to visualize cerebral metabolic activity in the rhesus monkey brain during the presentation of conspecific vocalizations. In contrast to the results reported by Poremba et al., (2004), the authors report significant activation in both the right and left (bilateral) posterior regions of the temporal lobe (their 'temporo-parietal' (Tpt) area) in response to passive listening to conspecific vocalizations when compared to non-biological sounds (e.g. environmental sounds from non-biological sources that were matched to the mean frequency and duration of the monkey vocalizations). (Significant clusters were identified bilaterally in one subject, only in the left hemisphere Tpt in a second subject, and only in the right hemisphere Tpt in a third subject). Significant activation was also observed in the ventral premotor cortex (in the right hemisphere for one subject, in the left hemisphere for the other two). No differences between the two species-specific call types (coos vs. screams) were observed with this paradigm, but, as compared to non-biological sounds, both vocalization call types evoked greater activity across a bilateral cortical network involving the temporo-parietal areas, ventral premotor cortices, and posterior parietal cortices. In another study, Petkov and colleagues (2008) identified a right anterior temporal lobe region of auditory cortex in the macaque brain that was selectively active during the perception of species-specific vocalizations (Petkov et al., 2008). In addition, they reported that this anterior monkey "voice" region was specifically sensitive to the vocalizations produced by familiar conspecifics.

Recent studies have increasingly begun incorporating the *direct* examination across primate species in a single experimental paradigm. For instance, Joly et al. performed near identical experiments with vocalization processing in humans and rhesus monkeys (Joly et al., 2012). During their respective experiments, members of both species heard speech sounds (French, comprehended by the human subjects, and Arabic, unintelligible to the subjects), non-verbal emotional human vocalizations, monkey vocalizations of different emotional valence, and spectrally "scrambled" versions of all stimuli. While humans produced preferential BOLD signals to human vocalizations (versus monkey vocalizations) primarily in the bilateral STG and STS, monkeys produced similar activation patterns and amplitudes to both monkey and human vocalizations along the STG and within the lateral sulci. No significant cortical hemisphere lateralizations were observed with the contrasts they used.

Frontal cortices are also known to play a critical role in vocalization processing. For instance, Romanski and colleagues (2005) examined the response properties and selectivity of neurons in the rhesus macaque ventrolateral prefrontal cortex (vlPFC) to the presentation of species-specific vocalizations (Romanski et al., 2005). The authors were interested in, among other things, whether or not certain neurons in the vlPFC would respond similarly to morphologically distinct calls that have similar functional referents. They reported that of the cells they recorded from, most were selective for two or three vocalization types from ten different exemplar call types. However, they found that the neurons were likely responding to calls with similar acoustic characteristics and signal features, as opposed to similar functional referents. No lateralization biases were noted, though such effects were not systematically explored. Nonetheless, such studies highlight the importance of considering whole-brain networks when addressing vocal signal processing. This is not too surprising given that the communicative behaviors of most primate species–including human language–typically span more than one sensory modality, and therefore include signals that necessarily go beyond the auditory stream.

Briefly, in regard to multisensory interactions and communication processing, a number of researchers have examined audio-visual processing by non-human primates in response to the presentation of conspecific vocalizations and their concomitant facial expressions. The results of these studies primarily indicate that multimodal communicative information (i.e. monkey vocalizations and the corresponding facial expressions) appear to be integrated in the rhesus monkey vlPFC as well as in auditory cortex (Sugihara et al., 2006; Ghazanfar et al. 2005). For example, Sugihara et al., (2006) presented conspecific vocalizations with or without accompanying video/still images of the face of a vocalizing rhesus macaque (Sugihara et al., 2006). They found multisensory neurons in the rhesus macaque vlPFC that exhibited enhancement or suppression in response to the presentation of face/vocalization stimuli. Romanski (2012) has proposed that the integration of vocalizations and faces that occurs in the macaque prefrontal cortex may represent an evolutionary precursor to the processing of multisensory linguistic input in the frontal lobe of the human brain (Romanski, 2012).

In sum, hemisphere lateralizations for processing conspecific vocalizations exist in some cases or paradigms, but this seems to be dependent on the control or contrast conditions used. To our knowledge, no systematic distinctions in functional activation foci have been made with regard to potentially distinct categories of call types or utterances that belong to theoretically distinct vocal communication categories (such as those addressed in Section 4), including categories that may share comparable meaning across species. Thus, some lateralization biases may have been effectively masked in earlier experimental paradigms due to the vocalization types selected as stimuli as well as the specific control/contrast condition(s) used. In future studies, researchers will be challenged to more closely examine the behavioral relevance of the vocal communications of the species under study and examine different putative categories or classes of meaningful and ecologically-valid calls and utterances.

Great Apes

While a considerable number of studies have utilized macaque monkeys to examine processing pathways for conspecific vocalizations, surprisingly little work has been done in other primate species, notably great apes. However, Taglialatela and colleagues recently used PET to visualize cerebral metabolic activity in chimpanzees in response to passive listening to two broad categories of conspecific vocalizations, proximal vocalizations (PRV) and broadcast vocalizations (BCV) (Taglialatela et al., 2009). PRVs are relatively low intensity vocalizations typically produced by individuals in direct proximity of one another, and are seemingly directed towards these individuals. Though speculative, such calls may potentially be considered as representing phatic and/or illocutionary utterances (see Section 4). Broadcast vocalizations are much louder (higher amplitude) calls as compared to the proximal vocalizations. They too are produced by individuals in the presence of conspecifics, but appear to be directed to distant individuals. Such calls may correspond more with perlocutionary expressions with the intent of causing specific psychological consequences to conspecific listeners (e.g. chimpanzee equivalent to “stay out” to non-troop individuals). If and precisely how the proximal versus broadcast call distinctions truly relate to the classifications of human utterances remains to be further established (Austin, 1975, Goodall, 1986), but future study will undoubtedly be critical for evaluating this hypothesis and other hypotheses related to protolanguage networks. Nonetheless, two important findings emerged from this chimpanzee study.

First, right-lateralized activity was observed in the posterior temporal lobe, including the planum temporale, when chimpanzees were presented with proximal PRV calls (but not time-reversed conspecific calls). However, similar lateralized activity was not observed during passive listening to the broadcast BCV calls. These results suggested that a functional distinction may exist between calls classified broadly as BCV versus PRV that correspond to differences in their processing in the chimpanzee brain. With regard to BCV versus PRV differences, previous behavioral work has found evidence of group-level structural variation in “pant hoot” vocalizations (Taglialatela’s broadcast (BCV) category) produced by both wild and captive chimpanzees (Arcadi, 1996, Marshall et al., 1999, Crockford et al., 2004). For example, Crockford and colleagues (2004) reported structural differences in the spectrograms of pant hoots of male chimpanzees living in neighboring communities, but not between groups from a distant community. These results could not be accounted for by genetic or habitat differences suggesting that the male chimpanzees may be actively modifying the structure of their calls to facilitate group identification (Crockford et al., 2004). Therefore, chimpanzees may effectively be using pant hoots as a means for discriminating among familiar and unfamiliar individuals. Overall, the above findings indicating differential and lateralized processing of different categories of call types (broadcast versus proximal) are at least roughly consistent with the human literature discussed earlier, suggesting that the primate auditory system may develop distinct hemisphere biases in processing pathways that are preferential for different categories or types of socially relevant vocalization information.

A second major finding from the single chimpanzee study (Taglialatela et al, 2009) regards some important hemisphere lateralization similarities and differences relative to findings from both Old World monkey species (Gil-da-Costa et al. 2006; Poremba et al. 2004, Petkov et al., 2008) and humans (e.g. Fig. 1). Specifically, Taglialatela and colleagues found right-lateralized activity along the STG for both proximal and broadcast conspecific calls when compared to time-reversed vocalizations. Thus, some consistencies have emerged between great apes and Old World monkeys (addressed earlier) with regard to observations of lateralizations for STG functions. However, more study with a given species call types, their behavioral relevance, and more consistent contrast conditions across primate species will be required to elucidate and further interpret the reported differences to date.

Comparing great apes with humans, inspection of the illustrated chimpanzee PET data and human fMRI data (cf. Fig. 1, cyan, and Fig. 4) show in both species significant right hemisphere STG activations when examining non-locutionary utterances, though in response to rather different control sound stimuli across the two studies. For the human study, the right STG was preferential for the sounds of humans mimicking animal vocalizations in contrast to intelligible locutionary speech with neutral emotional valence (spoken English phrases). In the chimpanzee study, the right STG was preferential for chimpanzee calls (PRV or BRV) in contrast to time-reverse versions of those calls. One possibility is that in both studies the right STG activation reflected processing of prosodic information of the non-locutionary (paralinguistic) utterances, which was less pronounced in neutral speech (human study) and in temporally reversed sounds (chimpanzee study).

Another commonality between chimpanzee and human cortical responses to vocalization utterances was in the right IFG region (cf. Fig. 1, cyan vs. Fig 4C, upper left panel). Though speculative, one hypothesis for future research is that the right IFG region may have a significant role in processing non-locutionary communicative utterances: They may prove to show a propensity toward representing sensitivity to conventionalization (learned meaning) of different types, categories or even sub-categories of utterances across homonids and other primate species. This is an intriguing hypothesis considering recent data indicating that both spoken language and symbolic gestures are processed by a common network in humans, which includes the inferior frontal gyrus and posterior temporal lobe (Xu et al, 2009). Thus, a picture that is emerging is that inferior frontal regions as well as temporal cortex in non-human primates are involved in constructing meaning from incoming signals in multiple modalities, including vocal communication sounds and, from monkey literature, corresponding facial expressions and communication gestures. The exciting observation of human-chimpanzee commonalities such as those illustrated in the present review underscore the rationale for future exploration of specific pathways and networks for processing different categories or sub-categories of communicative utterances. Future cross-species investigations, notably including chimpanzees, will be critical for developing a more robust theoretical framework regarding the phylogenetic changes associated with the evolution of spoken language processing in the human brain. Of course, much if not all of the meaningfulness behind communicative vocalizations and corresponding visual, gestural, or tactile expressions must be learned (or at least refined) through years of experience and neurodevelopment by a listener’s auditory system, which is our second theme, addressed in the final section below.

6. Vocalization processing in human infant auditory circuits

Understanding the human auditory system and its incredible vocalization processing abilities will be greatly benefited by studying it during its most dynamic developmental periods. Most auditory cortical mapping and other physiological studies in both human and non-human primates thus far have largely been performed with adult subjects. However, the operation of very efficient and streamlined mature systems may act to “conceal” or mask critical intermediate auditory processes or stages of proto-network development that may ultimately lead to protolanguage network processing that provides communication sounds with a sense of meaning to the adult listener. Given technical advances in human neuroimaging, testing the immature auditory systems of developing humans is progressively becoming an easier task.

Humans can generate preferential responses to human vocalization sounds at very early developmental time points, several of which are summarized below. Numerous studies of human fetuses and infants have described the greatest physiological responses (e.g. fetal heart rate changes, greater amplitude evoked potential using electroencephalography (EEG)) and behavioral responses (e.g. increased sucking behavior) to maternal vocalizations (versus other vocalizing individuals) or in response to speech produced in the child’s mother tongue (native versus other languages) (DeCasper and Fifer, 1980, Moon et al., 1993, DeCasper et al., 1994, Kisilevsky et al., 2003, Kisilevsky et al., 2009, Beauchemin et al., 2011, Sato et al., 2012). These preferential responses to maternal voices (and speech using the mother’s language) argue for the presence of cortical networks that are intrinsically optimized, or that become quickly optimized, for processing acoustic signals characteristic of those sounds, potentially influencing the development of hearing perception proto-networks *in utero*. Whether these early preferences in auditory circuits are genetically or epigenetically predetermined (Werker and Tees, 1999) (domain-specificity) or whether they are more experience-dependent remains largely unknown, and represents an exciting topic of future study. Parsimony, however, argues for a combination of both. Auditory (and other sensory systems) may begin with “experience-expectant” network structures (proto-networks) and processes that eventually give way to more “experience-dependent” organizational activity. Regardless, neuroimaging methods have begun to reveal the structure and function of early auditory communication processing networks during infant development. Similar to adult-based studies, vocalization processing studies in fetuses and infants have produced a confluence of ideas and findings; emerging themes include the basic acoustic processing of vocalization sounds in the left versus right cortical hemispheres, the roles of emotional or social cortical networks, and the role of developing neuroanatomy.

Hemisphere lateralizations

Vocalization processing is often viewed as an acoustic signal processing problem, wherein vocalization signals are regarded as spectrotemporally complex, multidimensional acoustic events. Previous findings in fully developed adult subjects generally have favored models that posit left/right hemisphere differences defined by different temporal processing timescales (Zatorre et al., 1992, Zatorre and Belin, 2001, Zatorre et al., 2002, Poeppel, 2003, Friederici and Alter, 2004). In particular, the processing of rapid acoustic signal changes is thought to be a left hemisphere dominant function (e.g. consonant sounds) whereas the right hemisphere shows the greatest sensitivity to more spectrally-stable envelope-level structure (e.g. vowel-like speech sounds or sounds containing prosodic cues). Functional and lesion studies with adults suggest that the right inferior frontal cortex (including the IFG) is involved in producing prosody and the right temporo-parietal regions in comprehending prosody in emotional verbal and non-verbal stimuli (Crosson et al., 2002, Ethofer et al., 2006, Pell, 2006). However, findings in this research field are sometimes conflicting depending on the functional contrasts examined and the nature of the speech or vocalization categories used (Boemio et al., 2005, Hickok and Poeppel, 2007, Obleser et al., 2008, Overath et al., 2008, Zatorre and Gandour, 2008). Despite these caveats, hemisphere lateralization theories have been tested with toddlers and infants.

By four years of age, a *left* hemisphere preference for intelligible speech becomes more clearly evident in humans (Wartenburger et al., 2007) showing patterns of network activation similar to typical right-handed adults (Scott et al., 2000, Szaflarski et al., 2002). Thus, segmental (phonological) and suprasegmental (prosodic) speech information processing appears to be at least partially represented along hemispheric boundaries around this age range, wherein the *right* hemisphere exhibits dominance for processing prosodic envelope-level information in vocalization signals.

In newborn babies, the cortical representations of “speech” sounds, as one broad category, showed left hemisphere processing dominance or advantage (Dehaene-Lambertz et al., 2002, Pena et al., 2003, Sato et al., 2012). One study showed a right hemisphere superior temporal cortex preference for human voice sounds in 7-month old infants; this effect was amplified when considering network modulations caused by different categories of prosodic cues in both speech and non-speech vocalizations (Grossmann et al., 2010). In another study, infants ranging from three days old to three and six months old were tested with temporally modulated noise samples using EEG and functional near-infrared spectroscopy (fNIRS); the artificial stimuli were designed to possess vocalization and speech-like spectrotemporal acoustic features (Telkemeyer et al., 2009, Telkemeyer et al., 2011). The use of fNIRS, which is sensitive to changes in hemoglobin and deoxyhemoglobin concentrations indirectly evoked by local neural activity, is becoming increasingly popular for measuring cortical responses in infants due to its relative non-invasiveness and ability to be used in more natural settings (Quaresima et al., 2012). Responses across all of these early age groups were relatively similar; sounds with rapid modulations produced fairly bilateral response patterns whereas the strongest responses to slowly modulating stimuli were dominant in the right hemisphere.

Further highlighting the early importance of emotional prosodic cues in vocal communication, Cheng et al. examined evoked response potentials (ERPs, a variant of EEG recordings) in newborns (less than five days old), reporting a right-lateralized mismatch response between speech samples of varying emotional valence (Cheng et al., 2012). Responses to negative valence stimuli were especially strong, perhaps reflecting an evolutionary processing bias for threatening stimuli (Vuilleumier, 2005). These finding generally corroborate the results from similar studies performed in adults showing stronger right hemisphere activation for emotionally evocative stimuli versus stimuli with neutral prosodic content (Grandjean et al., 2005). However, many studies investigating these functions often include linguistic content which may preclude stronger lateralization results; experimental design can strongly affect the lateralization of prosodic cue induced activity (Kotz et al., 2006).

Behaviorally, human infants generally show biases for many relatively simplistic harmonic vocalization sounds that contain strong prosodic cues. For instance, an unfamiliar adult may use some form of "motherese speech" in the presence of an infant or toddler, ostensibly for the purpose of pleasing them. These utterances usually have elongated and exaggerated vowels or vowel-like sounds (with correspondingly increased global harmonic content, see Fig. 3 colored rectangles) and often have very minimal or no intelligible semantic content. Laughter, smiling, and other positive responses in the infant provide early non-verbal feedback and reciprocation to adults that likely encourages more bonding interactions (Caron, 2002, Mireault et al., 2012), representing a social phenomenon that may have behavioral and acoustic evolutionary origins (Gamble and Poggio, 1987, Knutson et al., 2002, Vettin and Todt, 2005, Davila Ross et al., 2009). Within this framework, preferential right temporo-parietal responses to the prosodic pitch contours of speech are seen in three month old infants and are thought to represent facilitation of burgeoning left hemisphere networks during the subsequent learning of syntactic speech structures (Homae et al., 2006). This idea forms the basis of prosodic bootstrapping theories of language acquisition (Gleitman and Wanner, 1982, Jusczyk, 1997) and may further explain the importance of why infants have a preference for vocalizations with strong prosodic cues during this stage of neurodevelopment.

Similar to adult studies, it has been suggested that general "voice-sensitivity" emerges in the infant brain between four and seven months of age predominantly in posterior right temporal regions (Grossmann et al., 2010). The voice category in this study included speech and non-speech signals that were contrasted against non-voice stimuli (cars, airplanes, telephones, etc.) commonly used in adult studies (Belin et al., 2000). The findings from Grossman et al. may have represented the infant homolog to adult Temporal Voice Areas (TVA); these areas have traditionally been identified using broadly defined vocalization and non-vocalization categories (*ibid*). Blasi et al. has also demonstrated a right hemisphere bias for processing neutral non-verbal vocalizations versus non-voice environmental sounds that would likely be familiar to infants in the right anterior superior temporal cortex (Blasi et al., 2011). However, another fNIRS study using similar stimuli from Blasi et al. described age-dependent preferential voice responses in bilateral temporal cortices (Lloyd-Fox et al., 2012). The results from these studies, similar to those performed in adult subjects, may, however, be a product of extreme categorical differences between human vocalizations (verbal or non-verbal) and sounds that are not produced by human vocal tracts (man-made mechanical objects, environmental sounds, etc.). Namely, the sound stimuli used in the above mentioned contrasts crossed numerous categorical boundaries both acoustically and conceptually (Engel et al., 2009, Lewis et al., 2011, Lewis et al., 2012). Thus, these earlier vocalization processing paradigms may have concealed sub-threshold activity in more intermediate vocalization-specific networks where more subtle acoustic and communicative information distinctions may be realized. Additionally, within the voice categories, the representative stimuli also crossed many categorical boundaries (e.g. speech vs. non-speech, native vs. foreign speech, emotional vs. neutral non-verbal vocalizations, etc.) which again are likely to lead to very broad and relatively non-specific cortical network activations. Future work in vocalization neurodevelopment would benefit from utilizing distinct yet closely related categories of vocal communication sounds when describing regions and activity profiles that show human voice-sensitivity.

With regard to multisensory processing and perception of vocalizations, Grossman et al. recorded evoked response potentials (ERPs) from 4 and 8-month old infants as they watched dynamic audio-visual pairings of monkey faces and vocalizations as well as human-mimicked versions of the same stimuli (Grossmann et al., 2012, Talkington et al., 2012). As they grew older, infants became more sensitive to human-specific face-voice congruency versus monkey face-voice matching. Similar to Talkington et al., (2012) the authors reasoned that using non-stereotypical stimuli in unfamiliar contexts provided for stronger tests and interpretations of neuronal mechanisms. Extending this or a similar paradigm to non-human primates (see Section 5) would aid future investigations of the presence of similar human versus non-human primate processing phenomena.

While human mothers utilize “motherese calls” with their infants to a great extent, only very modest degrees of infant directed vocalizations have been reported for great apes (chimpanzees and bonobos) (e.g. Goodall, 1986). Nonetheless, paleoanthropological evidence suggests that prosodic features similar to current day motherese and infant-directed vocalizations were likely to have evolved as a prelinguistic vocal substrate in late australopithecines (Falk, 2004). Indeed, understanding the role of a developing infant’s social environment (human or great ape) will be crucial to understanding their vocal perception and the developing mechanisms that support this competency (Kuhl, 2007, 2010). Interactions emanating from prosodic socializing at early ages may be critical for promoting the development of cortical network processing of more complex vocalization utterances, notably including locutionary expressions by humans, which can convey a vastly greater degree of communicative content in terms of semantic knowledge. Encoding and storing greater semantic knowledge presumably entails use of more widespread cortical networks, which leads us to issues regarding anatomical constraints of development.

Anatomical constraints

The functions of prosodic versus verbal auditory networks, or proto-networks, further appear to be organized and refined by developmental constraints of a human infant’s early cortical neuroanatomy and physical architectures. For instance, Leroy et al. performed an extensive cortical maturation study in an infant cohort over the first several months of life (Leroy et al., 2011). Calculating a maturation index (MI) derived from T2-weighted magnetic resonance signals, the authors demonstrated that portions of the STS (especially the ventral banks) are of the more slowly developing perisylvian cortical regions, especially when compared to frontal regions. The right STS gray matter showed earlier maturation when compared to the left STS, consistent with other structural and genetic right-sided asymmetries found in the early developing brain (Sun et al., 2005, Hill et al., 2010). Additionally, Dehaene-Lambertz and colleagues also reported earlier maturation of the left arcuate fasciculus white matter tract (measured by higher fractional anisotropy diffusion tensor imaging (DTI) values) (Dubois et al., 2009). The highest correlations between functional cortical activity and age often occur in cortical regions representing the posterior territory of the left arcuate fasciculus, such as the left posterior STG/STS regions (Grossmann et al., 2010, Blasi et al., 2011), perhaps reflecting the rapid development of language skills. This STS/STG area may correspond to a postero-temporal region of cortex (called Spt), which is instrumental for sensory-motor integration, specifically with regard to speech processing (Hickok et al., 2009). The arcuate fasciculus is also proposed to be the structural foundation for the phonological loop and human language faculties (Aboitiz et al., 2010). Thus, an element of anatomical maturation sequences should be considered when attempting to explain the early stages of non-verbal vocal communication acquisition, both in human and non-human primates.

With regard to evolutionary considerations (Section 5), a comparative neuroanatomical study involving macaques, chimpanzees, and humans highlighted the increased cortical connectivity of “language-supporting” regions in humans that may have spurred the development of extensive language skills (Rilling et al., 2008), and the tight link it ultimately has with auditory systems in individuals with hearing ability. Thus, future studies should continue to address the possibility of fine structural changes that occur during infant and child neurodevelopment and concomitant gains in function with regard to processing different categories of non-locutionary communicative utterances.

Critical neurobehavioral milestones during development can be paired with concomitant changes in anatomy and function as nascent auditory networks and proto-networks form. However, future studies including infants will require taking gestational, postnatal, and overall maturational ages (i.e. from conception) into account when assessing specific developmental time frames (e.g. two “four week old” postnatal infants that may have had different gestational lengths). Determining how age (gestational and postnatal), lengths of respective auditory experiences, and brain maturation may or may not alter auditory responses to speech sounds is a promising topic of interest, perhaps reflecting “critical periods” of auditory development (Caskey et al., 2011, Key et al., 2012, Pena et al., 2012) that may also show parallels with great ape vocalization processing development. Together, future studies with human (and non-human primate) infants will promote the formation of more direct and accurate models for describing the relationships between anatomical structures, physiology, acoustic experiences, perception, and higher-order cognitive functions. These improved models will greatly aid in determining the etiology of auditory-related communication disorders as well as provide critical information for evidence-based therapies that can be implemented during specific developmental periods in early childhood.

7. Conclusion

Although diverse in content, the above fields of research regarding vocal communication and hearing perception point toward a need for further examining the idea of proto-networks and the processing of different types or categories of behaviorally-relevant non-locutionary utterances. In particular, evolutionary considerations can be more thoroughly addressed through continued study with chimpanzees and other non-human primates, which may reveal fundamental organizational principles across species. Additionally, neurodevelopmental considerations can be more thoroughly addressed through future study with human infants and fetuses, with one goal being the clarification of how, when, and where proto-networks for non-locutionary utterances may become established in the brain before ultimately settling into adult-like patterns for processing speech. Understanding intermediate processing stages for re-composing acoustic signals that comprise communicative utterances may prove valuable for the design of more intelligent hearing aids and hearing prosthetics that could help segregate salient as well as subtle speech signals from noisy acoustic environments. Moreover, understanding neurodevelopmental issues will be critical for advances in interactive therapies for children exhibiting difficulty with spoken language acquisition, and potentially for the neurorehabilitation of spoken language reception functions in individuals who acquire communication disorders (aphasias) resulting from stroke or other forms of brain injury.

Acknowledgments

This work was supported by the NCRR/NIH COBRE grant P20 RR15574 (to the Sensory Neuroscience Research Center of West Virginia University) and subproject to JWL.

Bibliography

Aboitiz F, Aboitiz S, Garcia R. The Phonological Loop A Key Innovation in Human Evolution. Curr Anthropol. 2010; 51:S55–S65.

Arcadi A. Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability. American Journal of Primatology. 1996; 39:159–178.

Austin, J. How to Do Things with Words. Cambridge: Harvard University Press; 1975.

Beauchemin M, Gonzalez-Frankenberger B, Tremblay J, Vannasing P, Martinez-Montes E, Belin P, Beland R, Francoeur D, Carceller AM, Wallois F, Lassonde M. Mother and stranger: an electrophysiological study of voice processing in newborns. Cereb Cortex. 2011; 21:1705–1711. [PubMed: 21149849]

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature. 2000; 403:309–312. [PubMed: 10659849]

Blasi A, Mercure E, Lloyd-Fox S, Thomson A, Brammer M, Sauter D, Deeley Q, Barker GJ, Renvall V, Deoni S, Gasston D, Williams SC, Johnson MH, Simmons A, Murphy DG. Early specialization for voice and emotion processing in the infant brain. Curr Biol. 2011; 21:1220–1224. [PubMed: 21723130]

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci. 2005; 8:389–395. [PubMed: 15723061]

Caron J. From ethology to aesthetics: Evolution as a theoretical paradigm for research on laughter, humor, and other comic phenomena. Humor-Int J Humor Res. 2002; 15:245–281.

Caskey M, Stephens B, Tucker R, Vohr B. Importance of Parent Talk on the Development of Preterm Infant Vocalizations. Pediatrics. 2011; 128:910–916. [PubMed: 22007020]

Cheng Y, Lee SY, Chen HY, Wang PY, Decety J. Voice and emotion processing in the human neonatal brain. J Cogn Neurosci. 2012; 24:1411–1419. [PubMed: 22360593]

Coath M, Balaguer-Ballester E, Denham SL, Denham M. The linearity of emergent spectro-temporal receptive fields in a model of auditory cortex. Biosystems. 2008; 94:60–67. [PubMed: 18616976]

Coath M, Brader JM, Fusi S, Denham SL. Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity. Network. 2005; 16:285–300. [PubMed: 16411500]

Coath M, Denham SL. Robust sound classification through the representation of similarity using response fields derived from stimuli during early experience. Biol Cybern. 2005; 93:22–30. [PubMed: 15944856]

Craig AD. How do you feel--now? The anterior insula and human awareness. Nat Rev Neurosci. 2009; 10:59–70. [PubMed: 19096369]

Crockford C, Herbinger I, Vigilant L, Boesch C. Wild chimpanzees produce group-specific calls: a case for vocal learning? Ethology. 2004; 110:221–243.

Crosson B, Cato MA, Sadek JR, Gokcay D, Bauer RM, Fischler IS, Maron L, Gopinath K, Auerbach EJ, Browd SR, Briggs RW. Semantic monitoring of words with emotional connotation during fMRI: contribution of anterior left frontal cortex. J Int Neuropsychol Soc. 2002; 8:607–622. [PubMed: 12164671]

Davila Ross M, Owren MJ, Zimmermann E. Reconstructing the evolution of laughter in great apes and humans. Curr Biol. 2009; 19:1106–1111. [PubMed: 19500987]

DeCasper AJ, Fifer WP. Of human bonding: newborns prefer their mothers’ voices. Science. 1980; 208:1174–1176. [PubMed: 7375928]

DeCasper AJ, Lecanuet J, Busnel M, Granierdeferre C, Maugeais R. Fetal Reactions To Recurrent Maternal Speech. Infant Behav Dev. 1994; 17:159–164.

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. Functional neuroimaging of speech perception in infants. Science. 2002; 298:2013–2015. [PubMed: 12471265]

Dubois J, Hertz-Pannier L, Cachia A, Mangin JF, Le Bihan D, Dehaene-Lambertz G. Structural asymmetries in the infant language and sensori-motor networks. Cereb Cortex. 2009; 19:414–423. [PubMed: 18562332]

Elhilali M, Shamma SA. A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. J Acoust Soc Am. 2008; 124:3751–3771. [PubMed: 19206802]

Engel LR, Frum C, Puce A, Walker NA, Lewis JW. Different categories of living and non-living sound-sources activate distinct cortical networks. Neuroimage. 2009; 47:1778–1791. [PubMed: 19465134]

Ethofer T, Pourtois G, Wildgruber D. Investigating audiovisual integration of emotional signals in the human brain. Prog Brain Res. 2006; 156:345–361. [PubMed: 17015090]

Falk D. Prelinguistic evolution in early hominins: Whence motherese? Behav Brain Sci. 2004; 27:491–503. [PubMed: 15773427]

Fecteau S, Armony JL, Joanette Y, Belin P. Is voice processing species-specific in human auditory cortex? - An fMRI study. Neuroimage. 2004a; 23:840–848. [PubMed: 15528084]

Fecteau S, Armony JL, Joanette Y, Belin P. Is voice processing species-specific in human auditory cortex? An fMRI study. NeuroImage. 2004b; 23:840–848. [PubMed: 15528084]

Fitch WT. Speech perception: a language-trained chimpanzee weighs in. Curr Biol. 2011; 21:R543–546. [PubMed: 21783032]

Fitzpatrick DC, Kanwal JS, Butman JA, Suga N. Combination-sensitive neurons in the primary auditory cortex of the mustached bat. The Journal of neuroscience: the official journal of the Society for Neuroscience. 1993; 13:931–940. [PubMed: 8441017]

Friederici AD, Alter K. Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang. 2004; 89:267–276. [PubMed: 15068909]

Gamble, E.; Poggio, T. Visual integration and detection of discontinuities: The key role of intensity edges. 1987.

Gil-da-Costa R, Braun A, Lopes M, Hauser MD, Carson RE, Herscovitch P, Martin A. Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. Proc Natl Acad Sci U S A. 2004; 101:17516–17521. [PubMed: 15583132]

Gil-da-Costa R, Martin A, Lopes MA, Munoz M, Fritz JB, Braun AR. Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. Nat Neurosci. 2006; 9:1064–1070. [PubMed: 16862150]

Gleitman, L.; Wanner, E. Language Acquisition: The State of the Art. Cambridge, UK: Cambridge University Press; 1982. The state of the state of the art; p. 3-48.

Goll JC, Crutch SJ, Warren JD. Central auditory disorders: toward a neuropsychology of auditory objects. Curr Opin Neurol. 2011; 23:617–627. [PubMed: 20975559]

Goodall, J. The Chimpanzees of Gombe: Patterns of behavior. The Belknap Press of Harvard University Press; 1986.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci. 2005; 8:145–146. [PubMed: 15665880]

Griffiths TD, Kumar S, Warren JD, Stewart L, Stephan KE, Friston KJ. Approaches to the cortical analysis of auditory objects. Hear Res. 2007; 229:46–53. [PubMed: 17321704]

Griffiths TD, Warren JD. The planum temporale as a computational hub. Trends Neurosci. 2002; 25:348–353. [PubMed: 12079762]

Grossmann T, Oberecker R, Koch SP, Friederici AD. The developmental origins of voice processing in the human brain. Neuron. 2010; 65:852–858. [PubMed: 20346760]

Heffner HE, Heffner RS. Temporal lobe lesions and perception of species-specific vocalizations by macaques. Science. 1984; 226:75–76. [PubMed: 6474192]

Herbert BM, Pollatos O. The body in the mind: on the relationship between interoception and embodiment. Topics in cognitive science. 2012; 4:692–704. [PubMed: 22389201]

Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 2009; 101:2725–2732. [PubMed: 19225172]

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8:393–402. [PubMed: 17431404]

Hill J, Dierker D, Neil J, Inder T, Knutsen A, Harwell J, Coalson T, Van Essen D. A surface-based analysis of hemispheric asymmetries and folding of cerebral cortex in term-born human infants. J Neurosci. 2010; 30:2268–2276. [PubMed: 20147553]

Homae F, Watanabe H, Nakano T, Asakawa K, Taga G. The right hemisphere of sleeping infant perceives sentential prosody. Neurosci Res. 2006; 54:276–280. [PubMed: 16427714]

Joly O, Pallier C, Ramus F, Pressnitzer D, Vanduffel W, Orban GA. Processing of vocalizations in humans and monkeys: a comparative fMRI study. Neuroimage. 2012; 62:1376–1389. [PubMed: 22659478]

Jusczyk, P. The Discovery of Spoken Language. Cambridge, MA: MIT Press; 1997.

Key AP, Lambert EW, Aschner JL, Maitre NL. Influence of gestational age and postnatal age on speech sound processing in NICU infants. Psychophysiology. 2012; 49:720–731. [PubMed: 22332725]

Kisilevsky BS, Hains SM, Brown CA, Lee CT, Cowperthwaite B, Stutzman SS, Swansburg ML, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z. Fetal sensitivity to properties of maternal speech and language. Infant Behav Dev. 2009; 32:59–71. [PubMed: 19058856]

Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z. Effects of experience on fetal voice recognition. Psychol Sci. 2003; 14:220–224. [PubMed: 12741744]

Knutson B, Burgdorf J, Panksepp J. Ultrasonic vocalizations as indices of affective states in rats. Psychol Bull. 2002; 128:961–977. [PubMed: 12405139]

Kotz SA, Meyer M, Paulmann S. Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. Prog Brain Res. 2006; 156:285–294. [PubMed: 17015086]

Kuhl PK. Is speech learning 'gated' by the social brain? Dev Sci. 2007; 10:110–120. [PubMed: 17181708]

Kuhl PK. Brain mechanisms in early language acquisition. Neuron. 2010; 67:713–727. [PubMed: 20826304]

Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. Hierarchical processing of auditory objects in humans. PLoS Comput Biol. 2007; 3:e100. [PubMed: 17542641]

Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30:7604–7612. [PubMed: 20519535]

Leech R, Holt LL, Devlin JT, Dick F. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2009; 29:5234–5239. [PubMed: 19386919]

Leroy F, Glasel H, Dubois J, Hertz-Pannier L, Thirion B, Mangin JF, Dehaene-Lambertz G. Early maturation of the linguistic dorsal pathway in human infants. J Neurosci. 2011; 31:1500–1506. [PubMed: 21273434]

Lewicki MS, Konishi M. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. Proceedings of the National Academy of Sciences of the United States of America. 1995; 92:5582–5586. [PubMed: 7777552]

Lewis JW. Cortical networks related to human use of tools. The Neuroscientist. 2006; 12:211–231. [PubMed: 16684967]

Lewis JW, Talkington WJ, Puce A, Engel LR, Frum C. Cortical Networks Representing Object Categories and High-level Attributes of Familiar Real-world Action Sounds. Journal of Cognitive Neuroscience. 2011; 23:2079–2101. [PubMed: 20812786]

Lewis JW, Talkington WJ, Tallaksen KC, Frum CA. Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. Front Syst Neurosci. 2012; 6(27):1–16. [PubMed: 22291622]

Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2009; 29:2283–2296. [PubMed: 19228981]

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR. Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex. 2010; 20:2958–2970. [PubMed: 20382643]

Lloyd-Fox S, Blasi A, Mercure E, Elwell CE, Johnson MH. The emergence of cerebral specialization for the human voice over the first months of life. Soc Neurosci. 2012

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S. Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage. 2001; 14:802–816. [PubMed: 11554799]

Marshall AJ, Wrangham RW, Arcadi AC. Does learning affect the structure of vocalizations in chimpanzees? Anim Behav. 1999; 58:825–830. [PubMed: 10512656]

McWhorter, J. The story of human language. Virginia, USA: The Teaching Company; 2004.

Medvedev AV, Chiao F, Kanwal JS. Modeling complex tone perception: grouping harmonics with combination-sensitive neurons. Biol Cybern. 2002; 86:497–505. [PubMed: 12111277]

Mireault G, Sparrow J, Poutre M, Perdue B, Macke L. Infant humor perception from 3- to 6-months and attachment at one year. Infant Behav Dev. 2012; 35:797–802. [PubMed: 22982281]

Moon C, Cooper R, Fifer WP. Two-day-olds prefer their native language. Infant Behavior & Development. 1993; 16:495–500.

Nelken I. Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol. 2004; 14:474–480. [PubMed: 15321068]

Obleser J, Eisner F, Kotz SA. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci. 2008; 28:8116–8123. [PubMed: 18685036]

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. Multiple stages of auditory speech perception reflected in event-related FMRI. Cereb Cortex. 2007; 17:2251–2257. [PubMed: 17150986]

Overath T, Kumar S, Stewart L, von Kriegstein K, Cusack R, Rees A, Griffiths TD. Cortical mechanisms for the segregation and representation of acoustic textures. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30:2070–2076. [PubMed: 20147535]

Overath T, Kumar S, von Kriegstein K, Griffiths TD. Encoding of spectral correlation over time in auditory cortex. J Neurosci. 2008; 28:13268–13273. [PubMed: 19052218]

Pascual-Leone A, Hamilton R. The metamodal organization of the brain. Prog Brain Res. 2001; 134:427–445. [PubMed: 11702559]

Pell MD. Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang. 2006; 96:221–234. [PubMed: 15913754]

Pena M, Maki A, Kovacic D, Dehaene-Lambertz G, Koizumi H, Bouquet F, Mehler J. Sounds and silence: an optical topography study of language recognition at birth. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:11702–11705. [PubMed: 14500906]

Pena M, Werker JF, Dehaene-Lambertz G. Earlier speech exposure does not accelerate speech acquisition. J Neurosci. 2012; 32:11159–11163. [PubMed: 22895701]

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. Nat Neurosci. 2008; 11:367–374. [PubMed: 18264095]

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. Speech communication. 2003; 41:245–255.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M. Species-specific calls evoke asymmetric activity in the monkey’s temporal poles. Nature. 2004; 427:448–451. [PubMed: 14749833]

Quaresima V, Bisconti S, Ferrari M. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. Brain Lang. 2012; 121:79–89. [PubMed: 21507474]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12:718–724. [PubMed: 19471271]

Reddy RK, Ramachandra V, Kumar N, Singh NC. Categorization of environmental sounds. Biol Cybern. 2009; 100:299–306. [PubMed: 19259694]

Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TE. The evolution of the arcuate fasciculus revealed with comparative DTI. Nat Neurosci. 2008; 11:426–428. [PubMed: 18344993]

Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res. 1996; 3:131–141. [PubMed: 8713554]

Romanski LM. Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. Proc Natl Acad Sci U S A. 2012; 109(Suppl 1):10717–10724. [PubMed: 22723356]

Romanski LM, Averbeck BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol. 2005; 93:734–747. [PubMed: 15371495]

Rumberg B, McMillan K, Rea C, Graham DW. Lateral Coupling in Silicon Cochlear Models. Midwest Symp Circuit. 2008:25–28.

Sato H, Hirabayashi Y, Tsubokura H, Kanai M, Ashida T, Konishi I, Uchida-Ota M, Konishi Y, Maki A. Cerebral hemodynamics in newborn infants exposed to speech sounds: a whole-head optical topography study. Hum Brain Mapp. 2012; 33:2092–2103. [PubMed: 21714036]

Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 2000; 123(Pt 12):2400–2406. [PubMed: 11099443]

Shannon RV. Advances in auditory prostheses. Curr Opin Neurol. 2012; 25:61–66. [PubMed: 22157109]

Sugihara T, Diltz MD, Averbeck BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci. 2006; 26:11138–11147. [PubMed: 17065454]

Sun T, Patoine C, Abu-Khalil A, Visvader J, Sum E, Cherry TJ, Orkin SH, Geschwind DH, Walsh CA. Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. Science. 2005; 308:1794–1798. [PubMed: 15894532]

Szaflarski JP, Binder JR, Possing ET, McKiernan KA, Ward BD, Hammeke TA. Language lateralization in left-handed and ambidextrous people: fMRI data. Neurology. 2002; 59:238–244. [PubMed: 12136064]

Taglialatela JP, Russell JL, Schaeffer JA, Hopkins WD. Visualizing vocal perception in the chimpanzee brain. Cereb Cortex. 2009; 19:1151–1157. [PubMed: 18787228]

Talkington WJ, Rapuano KM, Hitt L, Frum CA, Lewis JW. Humans mimicking animals: A cortical hierarchy for human vocal communication sounds. Journal of Neuroscience. 2012; 32:8084–8093. [PubMed: 22674283]

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD. Brain bases for auditory stimulus-driven figure-ground segregation. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2011; 31:164–171. [PubMed: 21209201]

Telkemeyer S, Rossi S, Koch SP, Nierhaus T, Steinbrink J, Poeppel D, Obrig H, Wartenburger I. Sensitivity of newborn auditory cortex to the temporal structure of sounds. J Neurosci. 2009; 29:14726–14733. [PubMed: 19940167]

Telkemeyer S, Rossi S, Nierhaus T, Steinbrink J, Obrig H, Wartenburger I. Acoustic processing of temporally modulated sounds in infants: evidence from a combined near-infrared spectroscopy and EEG study. Front Psychol. 2011; 1:62. [PubMed: 21716574]

Vettin J, Todt D. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. Behaviour. 2005; 142:217–240.

Vuilleumier P. How brains beware: neural mechanisms of emotional attention. Trends Cogn Sci. 2005; 9:585–594. [PubMed: 16289871]

Wartenburger I, Steinbrink J, Telkemeyer S, Friedrich M, Friederici AD, Obrig H. The processing of prosody: Evidence of interhemispheric specialization at the age of four. Neuroimage. 2007; 34:416–425. [PubMed: 17056277]

Werker JF, Tees RC. Influences on infant speech processing: toward a new synthesis. Annu Rev Psychol. 1999; 50:509–535. [PubMed: 10074686]

Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X. Functional properties of human auditory cortical fields. Front Syst Neurosci. 2010; 4:155. [PubMed: 21160558]

Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. Cereb Cortex. 2001; 11:946–953. [PubMed: 11549617]

Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends In Cognitive Sciences. 2002; 6:37–46. [PubMed: 11849614]

Zatorre RJ, Bouffard M, Belin P. Sensitivity to auditory object features in human temporal neocortex. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2004; 24:3637–3642. [PubMed: 15071112]

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science. 1992; 256:846–849. [PubMed: 1589767]

Zatorre RJ, Gandour JT. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos Trans R Soc Lond B Biol Sci. 2008; 363:1087–1104. [PubMed: 17890188]

Highlights (tentative)

1. We review different forms of human voice and vocal tract processing sensitivity in human cortex.
2. Different categories of non-speech utterances are considered for future research in human and non-human primates.
3. Chimpanzee auditory systems are compared to those of humans in the context of proto-networks and protolanguage networks.
4. We review human neurodevelopmental issues pertaining to vocal communication.

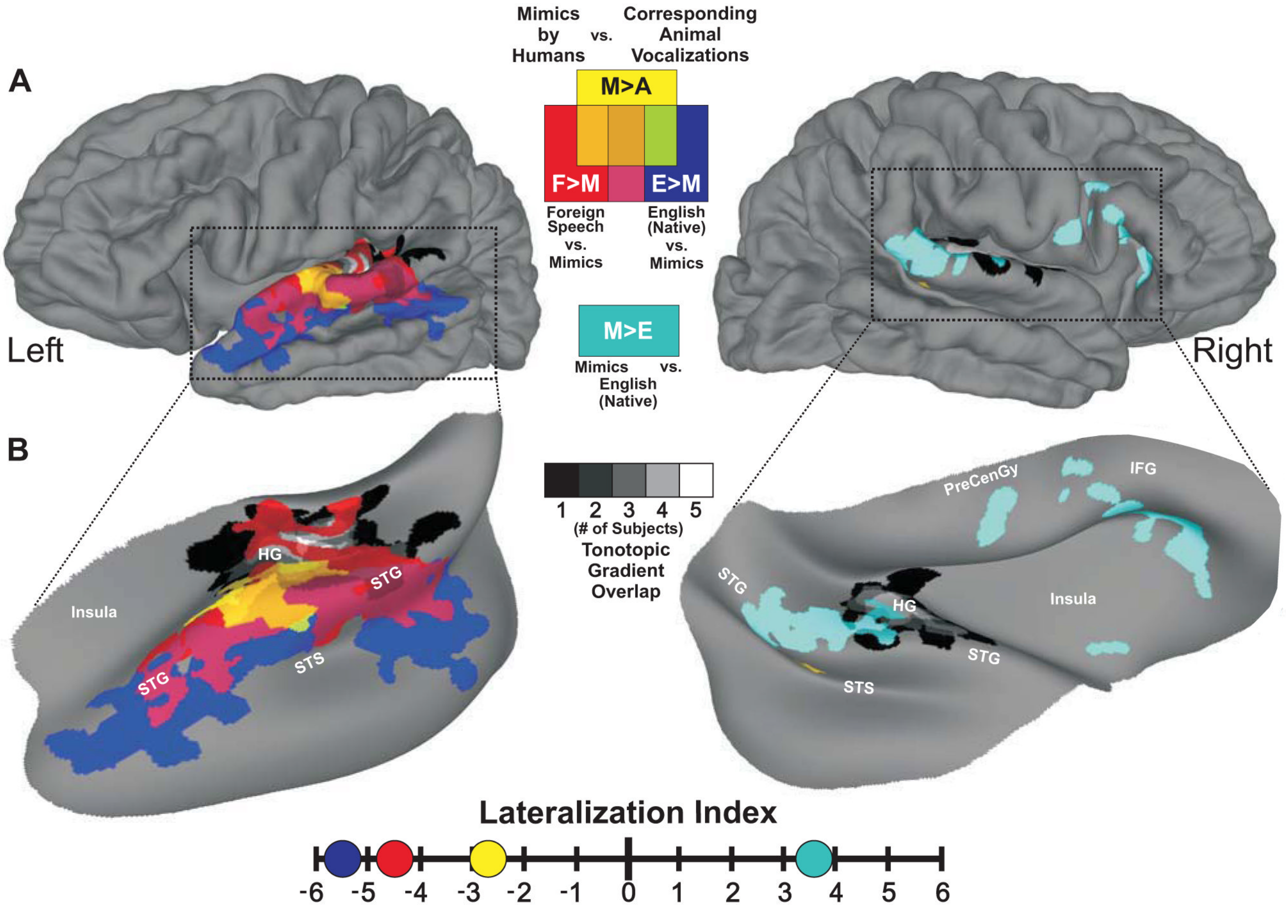
A
B
Left
Right
Mimics by Humans vs. Corresponding Animal Vocalizations
M>A
F>M
E>M
Foreign Speech vs. Mimics
English (Native) vs. Mimics
M>E
Mimics vs. English (Native)
1 2 3 4 5
(# of Subjects)
Tonotopic Gradient Overlap
HG
STG
STS
Insula
PreCenGy
IFG
Lateralization Index
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

Figure 1.

Conspecific vocalization processing hierarchy in human auditory cortex. A. Group-averaged (n=22) functional activation maps displayed on composite hemispheric surface reconstructions derived from the subjects. B. To better visualize the data, we inflated and rotated cortical projections within the dotted-outlines. The spatial locations of tonotopic gradients from five subjects were averaged (black-to-white gradients) and located along Heschl's gyrus (HG). Mimic-sensitive regions (M>A) are depicted by yellow hues, sensitivity to foreign speech samples versus mimic vocalizations (F>M) is depicted by red hues, and sensitivity to native English speech versus mimic vocalizations (E>M) is depicted by dark blue. Regions preferentially responsive to mimic vocalizations versus English speech samples (M>E) are depicted by cyan hues. Corresponding colors indicating functional overlaps are shown in the figure key. All data are corrected for multiple comparisons to p<0.05. Adapted and reprinted with permission from Talkington et al., (2012).

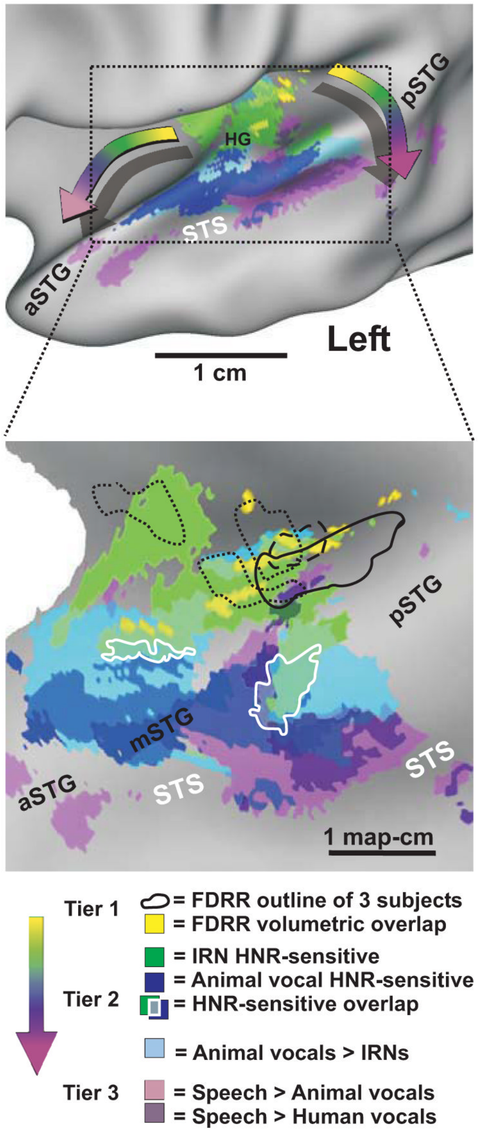

Figure 2.

Location of cortices parametrically sensitive to harmonic content (HNR-sensitive) relative to human vocalization processing pathways and tonotopically-organized regions that estimate the location of primary auditory cortices. Data are illustrated on slightly inflated (upper panel) and "flat map" (lower panel) renderings of averaged human cortical surface models. Data are all at α<0.01, corrected. Refer to key for color codes. FDRR = frequency-dependent response regions (tonotopic maps). Intermediate colors depict regions of overlap. The curved "rainbow" arrows depict two prominent progressions of processing tiers showing increasing specificity for the acoustic signal features present in human vocalizations. Overlap of IRN (green) and animal vocalization (blue) HNR-sensitivity is indicated (white outlines). Adapted and reprinted with permission from Lewis et al., (2009).

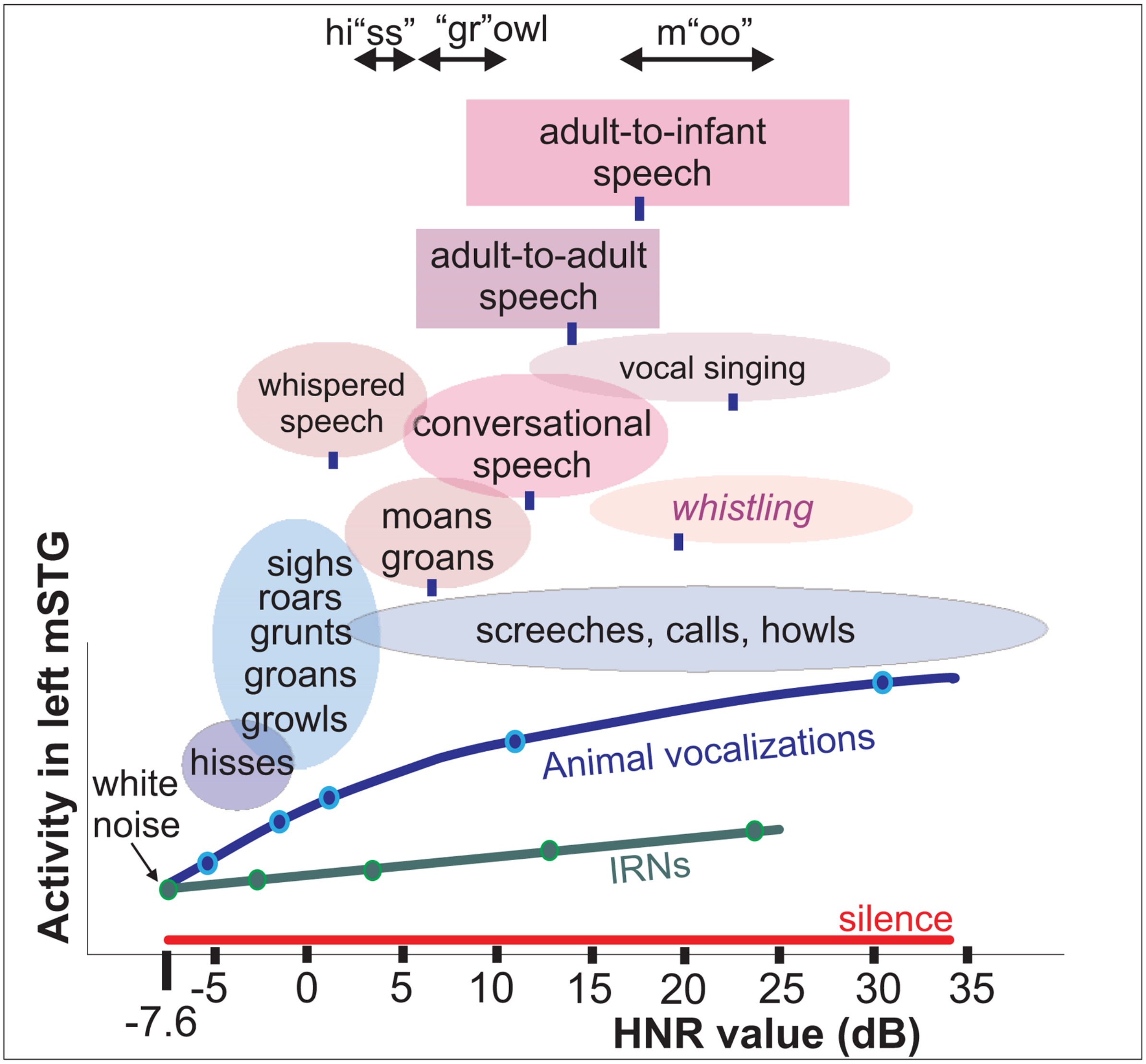

Figure 3.

Typical HNR value ranges for various sub-categories of mammalian vocalizations. Oval and box widths depict the minimum to maximum harmonic content (HNR values) of the sounds sampled, charted relative to the group-averaged HNR-sensitive response profile of the left mSTG (e.g. from Fig. 2). Green and blue dots correspond to IRN and animal vocalization sound stimuli, respectively, from Fig. 2. Blue ovals depict sub-categories of animal vocalizations explicitly tested. Ovals and boxes with violet hues depict sub-categories of human vocalizations (12–18 samples per category), and blue tick marks indicate the mean HNR value. For instance, conversational speech had a mean of +12 dB HNR, within a range from roughly +5 to +20 dB HNR. Adult-to-adult speech (purple box; mean = +14.0 dB HNR) and adult-to-infant speech (violet box; mean = +17.2 dB HNR) produced by the same individual speakers were significantly different (t-test p<10-5). Stress phonemes of three spoken onomatopoetic words depicting different classes of vocalizations are also indicated. Reprinted with permission from Lewis et al., (2009).

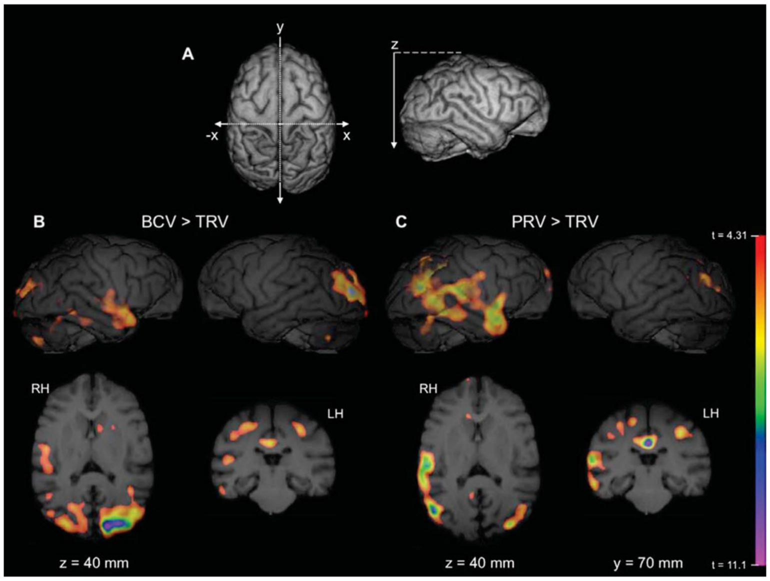

Figure 4.
Significant areas of activation in chimpanzees for (B) broadcast vocalizations (BCV) relative to time-reversed vocalizations (TRV) and for (C) proximal conspecific vocalizations (PRV) relative to TRV. Top images are 3D rendered MR images of chimpanzee right (RH) and left hemispheres (LH) with significant (t>4.31) PET activation overlaid. Adapted and printed with permission from Taglialatela et al., (2009).