10-1-2017

Auditory object perception: A neurobiological model and prospective review

Julie A. Brefczynski-Lewis
West Virginia University

James W. Lewis
West Virginia University

Digital Commons Citation
Brefczynski-Lewis, Julie A. and Lewis, James W., "Auditory object perception: A neurobiological model and prospective review" (2017). *Clinical and Translational Science Institute*. 700.
https://researchrepository.wvu.edu/ctsi/700

Published in final edited form as:
Neuropsychologia. 2017 October ; 105: 223–242. doi:10.1016/j.neuropsychologia.2017.04.034.

Auditory object perception: A neurobiological model and prospective review[☆]

Julie A. Brefczynski-Lewis[a,b] and **James W. Lewis**[a,b,*]

[a]Blanchette Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV 26506, USA

[b]Department of Physiology, Pharmacology, & Neuroscience, West Virginia University, PO Box 9229, Morgantown, WV 26506, USA

Abstract

Interaction with the world is a multisensory experience, but most of what is known about the neural correlates of perception comes from studying vision. Auditory inputs enter cortex with its own set of unique qualities, and leads to use in oral communication, speech, music, and the understanding of emotional and intentional states of others, all of which are central to the human experience. To better understand how the auditory system develops, recovers after injury, and how it may have transitioned in its functions over the course of hominin evolution, advances are needed in models of how the human brain is organized to process real-world natural sounds and "auditory objects". This review presents a simple fundamental neurobiological model of hearing perception at a category level that incorporates principles of bottom-up signal processing together with top-down constraints of grounded cognition theories of knowledge representation. Though mostly derived from human neuroimaging literature, this theoretical framework highlights rudimentary principles of real-world sound processing that may apply to most if not all mammalian species with hearing and acoustic communication abilities. The model encompasses three basic categories of sound-source: (1) action sounds (non-vocalizations) produced by 'living things', with human (conspecific) and non-human animal sources representing two subcategories; (2) action sounds produced by 'non-living things', including environmental sources and human-made machinery; and (3) vocalizations ('living things'), with human versus non-human animals as two subcategories therein. The model is presented in the context of cognitive architectures relating to multisensory, sensory-motor, and spoken language organizations. The models' predictive values are further discussed in the context of anthropological theories of oral communication evolution and the neurodevelopment of spoken language proto-networks in infants/toddlers. These phylogenetic and ontogenetic frameworks both entail cortical network maturations that are proposed to at least in part be organized around a number of universal acoustic-semantic signal attributes of natural sounds, which are addressed herein.

[☆]Guest Editors: Alfonso Caramazza & Marius Peelen.

[*]Correspondence to: Blanchette Rockefellar Neuroscience Institute, West Virginia University, Morgantown, WV, USA, jwlewis@hsc.wvu.edu (J.W. Lewis).

Keywords

Acoustic communication; Grounded cognition; Language evolution; Language neurodevelopment; Neuroimaging; Meta-analysis; Acoustic signal processing

1. Introduction: A new model for auditory perception, and why we need one

Categorization and recognition of objects are crucial to survival, as it allows us to interact with the world as well as predict what might happen and what actions we may immediately need to take. What defines an ‘object’ (visual, auditory, or haptic; see glossary) is not static, and changes due to experience, circumstances, and faculties of the perceiver. Nonetheless, we can gain understanding of the mechanistic underpinnings of cognition in humans in part by studying how we process and interact with the objects that we perceive (Varela et al., 1991). Because human and non-human primates are highly visiondominated species, and visual object processing in humans has been relatively easy to study in neuroimaging environments (such as with functional magnetic resonance imaging; fMRI), neurobiological models of human perception have been more thoroughly developed by studies in the realm of vision (Bar et al., 2001; Kaas and Collins, 2001; GrillSpector, 2003). Moreover, the visual acuity and ability to discriminate visual objects at a basic level appears to be reasonably similar in monkeys, great apes, and humans (Schmitt et al., 2013). Thus, models of visual object processing in the brain have been well formulated and refined through cross-species comparisons, with the presumption that the basic neuronal architectures for parallel, hierarchical processing in visual cortices have been relatively stable over the time course of primate evolution. However, hearing is different, both in terms of the physical signal attribute differences that need to be reconstructed by the brain, but more so the level of spectro-temporal discrimination that evolved in *hominins* (see glossary). Out of survival necessity, and now unique to humans, the auditory-vocal system had evolved the ability to convey and interpret increasingly subtle socially relevant emotional states of others (Donald, 1991; Munoz-Lopez et al., 2015; Fritz et al., 2016). Moreover, the formalization and unification of thought and knowledge in cortex is thought to have set the stage for spoken language systems to develop together with vastly greater degrees of auditory working memory abilities. In other words, humans may evoke a number of different sound processing strategies relative to other animals with dependence on listening task demands, intentional or unintentional influences by language systems, and/or our ability to decode a variety of high order signal attributes that may not be behaviorally relevant to other species.

Notwithstanding, many models of sensory perception acknowledge that object representations are quintessentially multisensory in global organization (Calvert and Lewis, 2004; Lewis, 2010; Murray et al., 2016) and we further assert that the auditory system in humans is also shaped by sensory-motor organizations in the brain that relate to acoustic communication. In the current review, we present a fundamental neurobiological model of hearing perception for natural sounds that addresses the question of what an “auditory object” is and the degree to which this is a useful concept for thinking about the cortical mechanisms that mediate hearing perception.

While many signal processing models have been developed for early stages of visual and auditory systems in human and non-human primates, such bottom-up models often fail to fully address the questions of “why” or to what end are the stimuli being processed. Many top-down cognitive models of sensory perception, even if heavily vision-dominated, do capture principles of knowledge representation that apply across sensory modalities (Tranel et al., 1997; Caramazza and Shelton, 1998; Caramazza and Mahon, 2003; Damasio et al., 2004). However, tests of such models often use language as either a stimulus, part of the task, or in other ways may inadvertently incorporate language system recruitment, which, from a hominid evolution perspective, may mask more fundamental processing principles or organizations of the auditory system. Thus, there remain significant gaps in our understanding of non-linguistic natural sound recognition mechanisms by the human brain at rudimentary perceptual levels.

Advances in models of hearing perception for everyday real-world natural sounds should be able to account for (i) previously established models of how “bottom-up” acoustic signal attributes at the cochlea become reconstructed along subcortical pathways and feed into appropriately specialized cortical processing pathways; (ii) apply to basic models in other mammals with hearing and oral communication ability; (iii) help account for the neuropsychology of auditory agnosias observed in some individuals after specific brain lesions, and thus incorporate “top-down” cognitive models of knowledge representations; and (iv) address how uniquely human qualities for comprehending spoken language, music forms, and interpreting subtle emotional cues may be interrelated with more rudimentary acoustic signal processing systems. Adhering to these considerations, the purpose of this review is to present a simple, formalized neurobiological model of hearing perception at an acoustic-semantic level, which has largely been refined based on fMRI neuroimaging findings over the past decade.

First, the model is introduced briefly below. Then, after addressing considerations of both bottom-up visual and auditory system models (Section 2) and top-down cognitive models (Section 3), we further elaborate on the general tenants of the hearing perception model at a more theoretical level (Section 4). This is followed by a discussion of testable predictions based on the model both in human and non-human primates (Section 5), and of limitations and future directions (Section 6). We believe that cortical models of auditory perception, in concert with visual perception models (and haptic perception, though this falls outside the scope of this review), will make unique contributions to understanding the nature of the human brain/mind and perceptual awareness on a number of fronts. This includes (1) fundamental advances in understanding how cognition and perception may function more globally, (2) understanding spoken language development in children, (3) understanding central hearing deficits and recovery after brain injury, (4) advances in biomimetic hearing aid algorithm designs, and (5) advances in anthropological models of oral communication during hominin evolution, by revealing potential vestiges in cortical networks that antedated modern spoken language systems.

In its simplest form, the proposed neurobiological model of hearing perception asserts that the cortical processing of meaningfulness associated with all natural sounds can be characterized by one of three major acoustic-semantic categories (Fig. 1, bold text) depicted effectively as a 2×2 design with subcategories therein. The first major category distinction is living (animate, biological) versus non-living. This boundary was derived from nearly a century of neuropsychology literature (Martin et al., 1996; Moore and Price, 1999; Grossman et al., 2002; Damasio et al., 2004), and further incorporates theories of embodiment mechanisms that may function to convey a sense of meaning or intent behind biological sounds to the listener (Barsalou et al., 2003; Barsalou, 2008). A second major category distinction is vocalizations versus non-vocalizations, wherein harmonic content (specific frequency combinations) reflects a primary acoustic signal attribute that is probabilistically more characteristic of vocalizations relative to most other behaviorally relevant action sounds (defined here as being natural sounds devoid of vocal content). There are no non-living vocalizations from an ethological perspective of this otherwise 2×2 model, though an interesting possibility is that this conceptual niche may in part be filled by sounds of musical instruments at some higher cognitive levels (addressed in Section 6).

The three major acoustic-semantic categories encompass other subcategories of natural sounds (Fig. 1, plain text). For the subdivision of “living things”, this includes actions sounds and vocalizations made by conspecific versus non-conspecific animals (i.e. human versus non-human in this review). Cortical representations of conspecific sounds are generally presumed to develop through experience, mediated by attentional systems that relegate greater processing to socially and behaviorally relevant nuances contained in conspecific action events and their resulting sounds (e.g. skilled tool use sounds) and in conspecific vocalizations (e.g. native speech sounds and singing talent). These subcategories are also likely to develop richer cortical representations due a listener’s physical experiences with producing the sounds themselves, and thus encoding a sense of a sound’s meaningfulness through audio-motor associations and ‘embodiment’ mechanisms. The generalizability of this model to all mammals would, of course, need to be adapted to account for a given species neurobiological constraints for sound transduction at the cochlea and capabilities for sound production and acoustic communication.

The category of non-living environmental action sounds (e.g. rain, water flow, fire, wind) includes the evolutionarily more recent subcategory of sounds produced by human-made automated machinery (mechanical action sounds devoid of an agent immediately instigating the action). Both of these non-living subcategories contain a number of distinguishing acoustic-semantic attributes that allow them to be characterized along an object-vs-scene signal dimension (addressed in Section 2), and thus perhaps best associated with the concept of an “auditory object”.

Because the three major category boundaries of this model have foundations in acoustic signal attribute processing that are probabilistically related to *semantic* features of their sound sources, we argue that these natural sound categories (and subcategories) are founded upon *acoustic-semantic universals*, for which the brain evolved to efficiently process, and serve as a foundation for other higher cognitive architectures, being processed by cortical networks that ultimately mediate hearing perception in non-deaf individuals.

2. Bottom-up perspectives of vision and hearing models

Over the past few decades, neuroimaging studies in human and non-human primates have revealed a panoply of parallel, hierarchical pathways for the processing of visual information (Farah and Aguirre, 1999; Bar, 2004; Palmeri and Gauthier, 2004; Bi et al., 2016), which has been compared and contrasted with audition (Kaas and Hackett, 2000b; Adams and Janata, 2002; Calvert and Lewis, 2004; Poremba and Mishkin, 2007; Lewis, 2010). This section aims to give the reader basic knowledge of these hierarchical pathways and networks of the mammalian brain that underlie perception of “what” the sound is, namely hearing for perception, recognition, and identification, which will be compared with the hierarchical pathways for visual object processing. Below we briefly summarize visual and auditory system models from a bottom-up unisensory signal processing perspective.

Visual inputs that define objects arrive as light rays along the two-dimensional array of the retina in the form of changes in luminance and frequency over time, and remains organized in a retinotopic (visuotopic) manner that is propagated in the neural code along various parallel, hierarchical pathways in visual cortices – a set of features that allows very high precision in spatial representation and the ability to discriminate both moving and static stimuli (Box 1). Visual objects can be segmented from the background at a fine spatial resolution, within roughly 100–250 ms (Thorpe et al., 1996). Object processing primarily, though not exclusively, involves ventrally directed streams from occipital to inferior temporal cortices (Ungerleider et al., 1982; Goodale and Milner, 1992; Freud et al., 2016). Visual object representations are built up in the brain through collections of neurons with lower-level visual receptive field properties to higher-level stages with receptive fields having greater specificity for different combinations of lower level visual features (Felleman and Van Essen, 1991; DeYoe et al., 1994). This includes increases in size and complexity of receptive fields, ranging from center-surround cells in the lateral geniculate nucleus, to representations of edges, lines/curves in primary visual cortices, to “object-like” features in the lateral occipital cortices (LOC). This information reaches different portions of the inferior temporal cortices that show selective, or at least preferential, processing for specific categories of visual entities such as for faces, distinct categories of objects, places, and scenes (Sergent et al., 1992; Kanwisher et al., 1997; Epstein et al., 1999; Pietrini et al., 2004). Lesions to various brain regions can lead to specific visual agnosias, including deficits in face recognition (prosopagnosia), an inability to read written words (pure alexia), or for general misperception of objects (visual object agnosias), among others (Geschwind, 1965; Damasio et al., 1982; Liepmann, 2001; Karnath et al., 2009).

Visual motion processing is another attribute of visual perception that impacts object recognition, ranging from structure-from-motion to gaining social cues from complex biological motion. Since sound production (by ‘auditory objects’) necessarily implies physical motion of some form, a consideration of visual motion processing systems (i.e. beyond static visual image processing) are imperative when making comparisons with auditory object processing systems. Ostensibly, a listener’s ability to recognize real-world auditory objects and acoustic events often requires processing of sound wave signal changes over considerably longer periods of time relative to vision, in some cases on the order of several seconds to recognize temporal patterns, and assert a sense of accurate recognition or identification.

The auditory system also has a parallel, hierarchical organization but there are key differences compared to vision. The everyday sounds we hear are initially represented at the cochlea, which is organized around tonotopy—a map of high to low frequencies of sound waves (organized roughly like a piano keyboard). Neural representations of any sound can be precisely defined physically by three attributes: the specific frequencies present, their relative intensities, and duration. The auditory system is far more sensitive than the visual system to timing and fast changes in signal energy, with temporal resolution of roughly 10's msec (Phillips, 1999). Information is sent up from the cochlea to the cortex via roughly half a dozen brainstem areas (Kandel et al., 2000), incorporating receptive fields with monaural or binaural representations conveying spatial location and increasing selectivity for various combinations of spectral and temporal signal information. The receptive field properties of auditory neurons similarly build-up in complexity and specificity along the cortical mantle to represent objects and action events (Rauschecker et al., 1995, 1997; Kaas and Hackett, 1998). Representations of sound at the level of primary auditory cortex *proper* maintains at least a roughly tonotopic organization in humans (Wessinger et al., 1997; Formisano et al., 2003), and neuropsychological studies indicate that information must reach this stage for normal bottom-up conscious awareness of sound and auditory object perception (Engelien et al., 2000). Auditory "core" regions of cortex receive strong thalamic inputs, are tonotopically organized, and thus generally regarded as primary auditory cortices (Kaas and Hackett, 1998; Sweet et al., 2005). These primary auditory regions send information on to higher hierarchical stages, which in macaques are termed "belt" and "parabelt" regions that surround the core (Rauschecker et al., 1995; Kaas and Hackett, 1998, 2000a; Rauschecker, 1998). In humans, primary auditory cortices (two or three auditory fields) are located along the medial two-thirds of the transverse temporal gyrus, or Heschl's gyrus (Rademacher et al., 2001; Formisano et al., 2003; Talavage et al., 2004; Lewis et al., 2009), which are depicted as functional landmarks in Figs. 2 and 3.

The tonotopic organizations of primary auditory cortices (PAC) give way to other organizations, including broadly defined "what" versus "where" divisions, which respectively include ventral versus dorsal pathways as major divisions of labor (Rauschecker, 1998; Kaas and Hackett, 1999; Rauschecker and Tian, 2000). Akin to the visual system, a dorsally directed processing stream (relative to the location of PAC) is critical for processing the location of sound ("where is it") relative to the listener's body for purposes of potentially interacting with the sound-source, while a ventral stream is thought to be involved more in processing for perception ("what is it"). This basic division reported in macaque monkeys is also supported by human neuroimaging studies using fMRI (Belin and Zatorre, 2000; Clarke et al., 2002; Arnott et al., 2004, 2008; Barrett and Hall, 2006; Wang et al., 2008; Recanzone and Cohen, 2009), neuropsychological brain lesion studies (Goodale and Milner, 1992), and electroencephalography (EEG) imaging (Ahveninen et al., 2006; Murray et al., 2006), and is proposed to further include a possible stream for defining "how" a sound is produced (Belin and Zatorre, 2000; Johnson-Frey, 2003; Kellenbach et al., 2003; Lewis et al., 2005). An extension of this dual-stream theoretical framework has also been applied to models of spoken language perception (Obleser et al., 2006; Dick et al., 2007; Rauschecker and Scott, 2009), including the idea that dorsal systems may be involved in the framing and temporal dynamics of word processing while ventral systems are more involved in representing content (MacNeilage, 1998). We further address spoken language processing from the perspective of top-down influences in Section 3, while in this section we address more fundamental natural sound processing that is more likely to be relevant to mammalian auditory systems more generally.

The comparisons thus far of bottom-up signal processing in visual and auditory systems for object representation have led us to a stage for addressing details of neuroimaging evidence that helped establish the model as a parsimonious account of the neurobiological data to date. This includes (2.1) vocalizations (and speech signals), which entail the processing of quantifiably differing degrees of harmonic content, (2.2) action sounds produced by living things, which can often be ‘embodied’ by sensory-motor systems, and (2.3) action sounds produced by non-living things, which are less likely to be embodied by sensorimotor systems but can be associated with high level representations of visual objects and scenes (in sighted individuals). The acoustic signal attributes that distinguish these three major categories of natural sounds may seem complex mathematically from a signal processing perspective, but we propose that they can at least be probabilistically modeled and thus represent acoustic-semantic universals that serve to bridge bottom-up processing with top-down multisensory, sensory-motor, and cognitive representations (Section 3) that help address the end(s) to which natural sounds are being processed in a listener’s brain.

2.1. Vocalization processing pathways and harmonic content

Unique to visual object processing are attributes such as color, relative brightness, and figure-ground segregation of static non-moving images or scenes. Conversely, unique to auditory processing are attributes such as pitch and harmonicity, which are prevalent in vocal sounds that convey communicative intent and/or emotional states in many species. We argue that harmonic content represents a universal acoustic-semantic signal attribute that effectively shapes (or helps shape) the organization of cortical networks for vocalization perception, and thus ultimately for spoken language perception. Harmonic content, or harmonicity, can be quantified by a harmonics-to-noise ratio (HNR) value over discrete time periods of a sound (Boersma, 1993; Riede et al., 2001; Ferrand, 2002). For instance, a wolf howl has much stronger harmonic content than a snake hiss, which is readily evident by strong frequency bands (“stacks”) in their spectrograms (cf. Figs. 2a and c). Moreover, various subcategories of behaviorally relevant vocalizations (Fig. 2e, colored ovals and boxes) can at least roughly be organized along a continuum of harmonic structure (Lewis et al., 2009). The lower end of the harmonic content scale includes hisses, growls, grunts, and groans, which are mostly associated with threat warnings and phatic utterances expressing negative emotional valence (Austin, 1975; Lewis et al., 2009; Talkington et al., 2013). At the other extreme, whistling, howls, and vocal singing are characterized by relatively higher harmonic content. Adult-to-adult conversational speech lies in a range between these extremes, and interestingly, adult-to-infant speech by the same individual speakers shows higher degrees of harmonicity in addition to the higher pitch ranges typically associated with “motherese” (Cooper and Aslin, 1990; Mastropieri and Turkewitz, 1999; Falk, 2004a). With regard to speech signals, the stress phonemes of onomatopoetic descriptors of several

English words symbolically representing the different subcategories of vocalizations (such as the ‘ss’ in hissing, the ‘gr’ in growling, and the ‘oo’ in mooing), also correlated with the relative harmonic content ranges of their respective vocal call subcategories.

To highlight the essence of the cortical pathways for vocal sound processing of the model (Fig. 1), Fig. 2 shows results from an fMRI paradigm that used the above-mentioned subcategories of vocalizations as sound stimuli (Lewis et al., 2009). This study revealed a sound processing hierarchy emanating out from PAC along the left and right hemisphere auditory cortices. The progression was based along dimensions of parametric sensitivity to harmonic content as well as increasing levels of information content or meaningfulness, and included three processing tiers (Fig. 2f, rainbow arrows). Outside of tonotopically organized core regions (Fig. 2f, yellow cortices), portions of the superior temporal plane showed parametric sensitivity to the harmonic content of artificially constructed sounds (green hues). Further lateral, the superior temporal gyri (STG) were regions parametrically sensitive to the harmonic content of animal vocalizations (dark blue hues). The processing of human speech sounds, which contained a greater degree of specific frequency combinations characteristic of human (conspecific) voice, resulted in activation located further antero-lateral and postero-lateral along the STG and superior temporal sulcus (STS) in the left hemisphere (purple hues). Similarly, human non-verbal vocalizations (e.g. sighs, moans, crying, laughter), which are conspecific signals that also contain specific combinations of co-modulating frequencies and presumably other higher order acoustic signal attributes that are learned to convey subtleties in meaning, recruited cortices along the STS, but mostly in the right hemisphere (pink hues).

Though the specific receptive field properties and processing mechanisms of vocalization signal preferring pathways remain to be more fully characterized, elements of harmonically structured signals inherent to vocalizations are clearly being utilized during vocal sound reconstruction, and are a feature unique to the hearing modality. Thus, in the model, harmonic content is proposed to represent an *acoustic-semantic universal* that the auditory system uses at a fundamental level as part of an organization to efficiently process natural sounds for determining meaningfulness.

This pathway for processing vocalizations as one of the major categories of natural sound is also illustrated as part of a meta-analysis (Fig. 2f data re-color coded in Fig. 3, contributing to red hues). This meta-analysis figure includes select data from several fMRI studies by our group to highlight the fundamental processing pathways and networks adhering to the acoustic-semantic categories defined in the proposed model. Neuroimaging results that further contributed to this meta-analysis are described briefly below and further in Section 3 (also see figure legend), followed by a formalization of the general tenants of the model.

Continuing with vocalization processing, human vocalizations seem to be special for humans (conspecifics). In general, conspecific vocalizations for a given species may recruit greater expanses of cortical hierarchies for signal processing due to life-long familiarity and experience, and the relatively greater degrees of attention devoted to extracting behaviorally relevant information. This notably includes the processing of harmonic-vocal sounds produced by caretakers that may even begin *in utero* (Lee et al., 2007; Webb et al., 2015).

Neuroimaging studies of non-human animals have identified brain regions along auditory cortices showing specificity, or at least preferential activation, in response to that species conspecific calls and vocalizations, including dogs (Andics et al., 2014), macaque monkeys (Rauschecker, 1998; Petkov et al., 2008; Ortiz-Rios et al., 2015), and chimpanzees (Taglialatela et al., 2009). Based on behavioral studies, acute sensitivity to conspecific calls further seems to be a feature of other mammals including ungulates (e.g. deer, elk) (Lingle and Riede, 2014) and even aquatic mammals such as pinnipeds (e.g. walruses, seals) (Cunningham et al., 2014; Reichmuth and Casey, 2014). Note, however, that some emotional calls, especially distress vocalizations, appear to reflect evolutionary conserved mechanisms underlying both production and behavioral responses (Lingle et al., 2012). Thus, responses to some categories of emotional calls may reflect innate or ‘reflexive’ circuit mechanisms that could dissociate from the proposed hearing perception model, which emphasizes processing of natural sounds that are learned to have meaning to the listener through experience. Notwithstanding, in human neuroimaging studies, spoken language sounds lead to activation along specific brain regions, with comprehensible language being more lateralized to the left hemisphere (Binder et al., 1997, 2000; Belin et al., 2000; Belin and Zatorre, 2003; Obleser et al., 2006), and pitch, prosody, and emotional qualities in human vocals more heavily involving the right hemisphere (Zatorre et al., 1992; Buchanan et al., 2000; Gandour et al., 2004). Human voice sounds, in contrast to instrumental sounds, are also reported to show distinct responses in electroencephalography (EEG) studies (Levy et al., 2001). In infants, human voice leads to preferential activations (Fifer and Moon, 1989; Dehaene-Lambertz et al., 2002; Pena et al., 2003; Grossmann et al., 2010; Sato et al., 2012), attesting to their special status as a conspecific (or at least familiar) acoustic subcategory of vocalization that may be learned very early in life.

To identify which brain regions might show specificity for processing non-linguistic human vocalizations, Talkington et al. (2012) used fMRI to examine cortical responses to the processing of sounds produced by humans mimicking animal calls in contrast to hearing the original animal calls themselves as a critical control. Interestingly, hearing the human mimic sounds more strongly activated auditory belt/parabelt regions in the *left* cortical hemisphere (contributing to Fig. 3, red hues). These results suggested that a left lateralization bias may exist not only for speech sound processing but perhaps more generally for processing the spectro-temporal attributes characteristic of human (conspecific) vocal tract sounds, though this finding and its relation to lateralizations for spoken language processing remain to be further elucidated.

Conversely, hearing animal vocalizations, in contrast to humans mimicking the corresponding sounds in the above study, led to greater activation along a number of right lateralized brain regions (Fig. 3, contributing to red hues). These higher processing tiers were consistent with earlier described networks for processing affective prosodic cues of vocalization stimuli, such as slow pitch-contour modulations (Zatorre and Belin, 2001; Kotz et al., 2003; Friederici and Alter, 2004; Ethofer et al., 2006; Ross and Monnot, 2008; Grossmann et al., 2010). Thus, some of the right lateralized regions may have been related to processing the prosodic cues and other signal features in the original animal calls (Wilden et al., 1998; Farago et al., 2014) that were not adequately, or at least differently, conveyed by the human actors mimicking them. These findings were consistent with neuropsychological studies indicating that lesions to the right inferior frontal gyrus (IFG) and anterior insula region can lead to sensory aprosodia, an inability to interpret emotion in voice (Heilman et al., 1975).

In sum, pathways for non-linguistic vocalization processing are robustly present in both cortical hemispheres out to the STG/STS regions (Fig. 3, red hues). The left hemisphere appears to show relatively greater sensitivity to the processing of human (conspecific) vocal tract sounds and for linguistic content, while the right hemisphere shows processing biases for prosody and other emotional attributes. This lateralization effect for higher level conspecific communication signals is argued to be related to hemisphere specializations originally related to handedness, praxis skills, and gestural origins of language/ communication, as addressed in later Sections. In contrast to vocalization sounds, however, we next consider how the auditory system deals with non-vocal “action sounds”.

2.2. Action events produced by living things

The next two categories, living and non-living action sounds, are defined here as representing all other types of natural sound signals other than vocalizations, and share many features that correlate in time across visual and/or sensorimotor modalities (e.g. cross-modal correlated changes in stimulus intensities, such as impact sounds). The intermodal invariant features of action sounds produced by non-living things, such as wind blowing through trees, often include complex visual motions that correlate (for sighted listeners) with sound signal textures and amplitudes (addressed in Section 2.3 below). Action sounds produced by living things, such as when hearing and viewing an individual dribbling a basketball, entail changes in signal energy that correlate in time across visual, auditory, and often with sensorimotor systems. Thus, a listener’s experience with producing similar sounds themselves are likely to establish associations (“embodiments”) that further convey potential intention or meaning behind the sounds when heard in isolation. These multisensory and audio-motor association distinctions, we argue, help establish other sets of acoustic-semantic universals used for organizing the mammalian auditory system to process action sounds produced by living things, as addressed below.

A recent fMRI study by our group directly tested where auditory pathways for processing action sounds by living things versus vocalizations might diverge (Webster et al., 2017). Non-human animal vocalizations and non-human action sounds were used to minimize confounds associated with potentially greater semantic processing of conspecific sounds. Relative to primary auditory cortices, vocalizations activated the STG regions bilaterally as expected (Fig. 3, contribution to red regions), while animal action sounds preferentially activated the posterior insulae bilaterally (Fig. 3, contributing to yellow hues). These regions were presumed to be associated with audio-tactile or audio-sensorimotor cortices, similar to the organization reported in macaques (Schroeder et al., 2001; Fu et al., 2003; Hackett et al., 2007). Thus, the divergence in processing for this major acoustic-semantic boundary was along intermediate auditory cortical stages, overlapping classically defined parabelt regions. However, at lower threshold settings, additional left-lateralized cortical regions were also preferentially activated by the action sounds, including frontal, parietal and posterior temporal regions that have previously been associated with mirror neuron systems (MNS) in human (Rizzolatti and Craighero, 2007; Molenberghs et al., 2012) and non-human primates (Rizzolatti and Craighero, 2004). While the implications of the MNS-like representations will be addressed further in Section 3 from a top-down perspective, these results support the notion that the audio-motor associations of biological action sounds reflect a form of acoustic-semantic attribute that may serve as universal signals for organizing or refining the mammalian “auditory” system to mediate hearing perception.

In dual stream vision models, motion processing has a strong dorsally directed component, involving a number of occipito-parietal and parietal cortices plus the lateral temporal visual motion area hMT—terminology derived from the macaque monkey area MT (Van Essen et al., 1981; Tootell et al., 1995). The visual processing of articulated biological motions preferentially occurs just anterior to hMT along the posterior middle temporal gyri (pMTG) and posterior superior temporal gyri (pSTS) regions (Grossman et al., 2000; Beauchamp et al., 2002; Grossman and Blake, 2002; Thompson et al., 2005; Han et al., 2013), and are reported to represent primary loci for complex natural motion processing (Martin, 2007; Lewis, 2010), most notably including human (conspecific) actions. Activation in the pSTS/pMTG complexes (Fig. 3, labeled yellow region) shows interaction or integration effects when corresponding sounds are also present (Calvert et al., 2000; Beauchamp et al., 2004b, 2004a; Taylor et al., 2006, 2009; Campanella and Belin, 2007; Campbell, 2008), and are generally activated by human action sounds in the absence of visual input (Lewis et al., 2004, 2006; Bidet-Caulet et al., 2005; Gazzola et al., 2006; Galati et al., 2008; Engel et al., 2009). These regions were further shown to be more strongly activated by human action sounds relative to non-human animal action sounds, and lesser still by non-living action sounds (Engel et al., 2009) or vocalizations (Webster et al., 2017). Thus, from a bottom-up signal processing perspective, these complexes appear to play a prominent perceptual role in transforming the spatially and temporally dynamic features of natural auditory (and visual) action information into a common neural code, conveying symbolic associations of physically matched audio-visual features. The pSTS/pMTG regions are also activated in association with hearing tool-use sounds and with manipulating virtual tools (Lewis et al., 2005, 2006). Hence, multisensory and sensorimotor association processing appears to be central to the supramodal processing functions of the pSTS/pMTG complexes.

Human action sound processing at a categorical level appears to develop prior to spoken language abilities, as assessed from a study of prelingual infants (Geangu et al., 2015). Using sound stimuli consistent with the category boundaries defined in the proposed neurobiological model, they examined event related potentials (electroencephalography methodology) of infants at 7 months of age, as they listened to human action sounds, human vocalizations, house-hold mechanical action sounds and sounds of the natural environment. Their results indicated that human action sounds were being differentially processed as a distinct category of socially relevant sound. This category-specific sound processing was occurring prior to the development of top-down language system influences, and also prior to formation of audio-motor associations with human action sounds, although they likely have visually observed other human conspecifics in their world using tools, with some level of pantomime or play tool-use movements.

In sum, a complex interplay of audio-visual associations and/or audio-motor associations can help shape the differential responsiveness of brain regions for processing different categories of natural sound (though also see Section 3), and namely action sounds produced by living things. However, associations with non-living sound sources are qualitatively different, leading to the third major category of natural sounds.

2.3. Action events produced by non-living things

Non-living action sounds, such as rain, ocean waves, fire and wind, produce sound in a manner that cannot be fully or meaningfully emulated by a listener's own motor system (which makes the X-Men character "Storm", who can control weather, a particularly interesting fictional superhero from the scientific perspective of this review). Consequently, the *human* brain must rely more heavily on learning and associating the acoustic signal attributes of non-living action sounds with visual motion cues (if sighted) and with tactile inputs. However, early blind individuals readily learn to recognize non-living action sounds (see Section 3), indicating that visual associations are not absolutely required.

In another line of reasoning, natural sound events, living or non-living, are usually not heard in complete isolation in real-world settings. Hence, the auditory system was proposed to have evolved to focus on streaming sound components that likely belong to a given sound-source (Bregman, 1990; Teki et al., 2011). This processing strategy shares analogies with segmenting of figure-versus-ground and structure-from-motion in the visual system (Parks, 1984; Yantis and Jonides, 1990; Rubin, 2001). Thus, one might similarly expect a model of hearing perception to show a dichotomy, or at least a gradation, in cortical organization for representing auditory objects versus auditory scenes (acoustic soundscapes and/or distracting ambient noises). In this regard, an important role of the auditory system is not only to alert and attend to potential auditory objects of interest, but also to perform dynamic acoustic accommodation by simultaneously "filtering out" the drone of less relevant background acoustic noise (Bregman, 1990), thereby freeing up attentional resources for other sensory or cognitive processes. In ferrets, directing attention to particular types of sound as either foreground versus background were shown to modulate primary auditory cortices (Fritz et al., 2007b, 2007a), and in humans detecting and recognizing animal calls amidst acoustically complex background scenes revealed activated foci along the left and right angular gyri (Maeder et al., 2001). Thus, some aspects of auditory object processing (in sighted and blind) may be addressed in terms of an object-vs-scene dimension.

Because the auditory system, we argue, places a premium on assessing whether or not a sound-producing action is being caused by a living agent (addressed in Sections 2.2 and 3), notably by engaging audio-motor embodiment mechanisms to ascertain possible intentful actions, a comparison between visual and auditory object-vs-scene processing systems might be more straightforward when examining sound-sources that fall within the category of non-living action sounds (Fig. 1, rightmost box). With this rationale in mind, another study by our group examined cortical networks showing sensitivity to auditory objects versus auditory scenes using 'non-living' environmental and mechanical sounds (Lewis et al., 2012), adhering to the category boundary definitions in the proposed model. Sounds were assessed psychophysically, revealing a continuum from object-like to scene-like exemplars that avoided perceptual processing issues related to embodiment and intention. This included environmental sounds rated as more object-like (e.g. water dripping in a cave) versus more scene-like (e.g. wind blowing through trees), plus mechanical sounds rated as more object-like (e.g. a ticking watch) versus more scene-like (e.g. industrial laundering machinery). Neuroimaging revealed a double dissociation of cortical processing regions on this object-vs-scene dimension, with foci along the left and right STG showing preferential activation to object-like sounds (Fig. 3, light blue hues) versus various cortical midline regions preferential for scene-like sounds (dark blue hues). What bottom-up acoustic or acoustic-semantic signals might have been contributing to this perceptual dichotomy of auditory objects versus scenes?

Prominent low-level acoustic attributes, such as loud sounds or salient three-dimensional spatial cues, can cause a sound-source to suddenly pop out as an "auditory object". This includes signal processing along reflexive circuits, involving brainstem regions such as the inferior colliculi (Belenkov and Goreva, 1969), though this does not lead to recognition *per se*. Sound-production necessarily implies some form of motion, which often includes fairly robust first order motion cues such as interaural intensity differences (IID), interaural time differences (ITD), amplitude changes (e.g. looming), and Doppler effects (e.g. changing frequency of sound of a train speeding by on a track). These acoustic cues, and even illusory spatial motion, are reported to activate primary auditory cortices (Griffiths et al., 1994; Mäkelä and McEvoy, 1996; Murray et al., 1998; Baumgart et al., 1999; Lewis et al., 2000; Warren et al., 2002). However, hearing stationary non-living sound sources, such as a watch ticking or listening in a noisy room for your computer muffin fan to determine if it might be overheating, may be largely devoid of distinctive first order motion cues, with only subtle or no visual motion cues.

A number of higher order signal attributes can capture the temporally homogeneous signals that represent scene-like sounds, and be used to help segment auditory objects from otherwise uninteresting acoustic scenes or soundscapes. This includes, for instance, "sound textures" (McDermott and Simoncelli, 2011). In vision, textures represent intermediate-level feature attributes that the system can use to define salient object boundaries or to fill in surfaces of perceived visual objects (Reppas et al., 1997; Kastner et al., 2000). Another form of higher order acoustic signal attribute related to textures includes spectral structure variation (SSV) measures, which quantifies changes in signal entropy over time (Reddy et al., 2009). Auditory object-like sounds tend to show relatively greater measures of SSV and lower mean entropy levels than scene-like sounds (Lewis et al., 2012), and an fMRI study indicated that the activation along the anterior STG regions for object-like sounds (Fig. 3, light blue) shows parametric sensitivity to SSV measures. Thus, the processing of sound textures, entropy, and SSV measures may be reflective of prägnanz computations in cortex used to probabilistically simplify the likely segmentation of sounds and auditory objects from acoustic background scenes (Cusack and Carlyon, 2003; McDermott and Oxenham, 2008).

Many scene-like sounds that have relatively homogeneous acoustic temporal structure over time (e.g. Box 1, rainfall) also have a power spectrum that can at least be grossly characterized by smoother $1/f^{\alpha}$ spectral shape: where f = sound wave frequency and α

ranges from 1 to 2 (Lewis et al., 2012). In other words, the physics of sound propagation is such that as the distance between an observer and a given sound source increases, the higher frequency sound pressure waves undergo disproportionately greater losses in intensity. Thus, distant sound-sources may effectively be filtered along cortical pathways as object-versus scene-like based on the learned relative shapes of their power spectra. In experienced adult listeners, many sound producing events that are located far away may tend to be less immediately relevant (though with dependence on survival/task demands and settings) and thus more likely relegated as sensory background noise or ambience rather than attention demanding closer range object-like status. Physically experiencing and interacting with up close sound-sources (living and non-living) presumably help develop representations for object-like acoustic signal attributes (through vision and touch)—and involve cortical regions such as the STG (Fig. 3, light blue hues). In this sense, the term “auditory object” seems to have its greatest relevance as a concept when comparing multisensory features to visual objects and haptic objects.

Scene-like action sounds, however, which are not tangible but may develop visual associations (in sighted individuals), appear to develop by establishing representations outside of auditory cortex proper (Fig. 3, dark blue hues) relating to “non-self” and/or episodic representations (Ries et al., 2007; Burianova et al., 2010). Probabilistically learned scene-like acoustic signatures based on relative loudness, $1/f^{\alpha}$ spectral shape, acoustic textures, SSV, and presumably other higher order signal attributes, are proposed here to reflect acoustic dimensionality reductions that could define *acoustic-semantic universals* that help shape the organization of the mammalian brain for natural sound processing. Notwithstanding, the auditory system may be prone to accommodate or calibrate to acoustic signal structures that are statistically more likely to represent a background scene, representing a form of acoustic accommodation. While this may be especially pertinent to representing environmental sounds of nature, this same mechanism may also apply to the processing of crowded social scenes of people talking as background (e.g. at a market place). The learned acoustic features of scene-like sounds may thus probabilistically lead to the differential processing in the brain, contributing to the perception of acoustic scenes that may apply to any of the major categories of natural sound, though may be most pertinent to non-living categories from an ethological perspective (Box 2).

Relaxing acoustic scenes—Incidentally, for relaxation purposes many people like to listen to soundscapes of nature (most commonly non-living things), such as the soothing sounds of a babbling brook in the woods, the tranquil crackling from a Yule Log fireplace video, music that emulates the above mentioned acoustic features (such as “alpha wave” music), and soundscapes of crickets or street traffic. Because these sounds represent actions that are generally well out of a person’s ability to directly physically control or influence, they are less readily represented in cortical networks relating to purposeful or intentful behaviors. Thus, they less readily activate motor repertoires affiliated with a listener’s sense of “self” or the embodiment of intentful actions of others, and instead engage other cognitive processes not related to self-representations. Though speculative, the processing of environmental sounds outside of motor-related networks (Fig. 3, dark blue hues) may thus help explain why subdued “sounds of nature” aid in relaxation and stress reduction, as they can help take one’s mind off of their self-ruminating thoughts.

In sum, research approaches using bottom-up acoustic signal processing similar to those described above, have led to tremendous strides in our understanding of how sound processing may be organized in cortex beyond tonotopic organizations. However, to more comprehensively address the neurobiology of perception other major lines of research have been the development of a variety of top-down models of brain organization that may mediate perception and cognition, as addressed next.

3. Top-down perspectives of vision and hearing models

Relying on bottom-up theories alone does not provide an adequate answer as to ‘why’ the brain is organized the way it is. A main category boundary of our model is living (action sounds and/or vocalizations) versus non-living action sounds. When one considers what may be at the root of this difference, network representations conveying meaning and intent form a solid explanation. Animate sources are much more likely to be imbued with intent. Thus, animate sounds, with a generally greater sense of meaning to the subject, are more likely to require complex behavioral responses such as social engagement, fleeing, approach, or engagement as predator or prey. But how do these sounds become cognitively interpreted so that correct behavioral action can be affected?

Some theories of cognition have posited that high level object and action knowledge resides in a central semantic memory system (a central store) separate from the brain’s sensory, or modal, systems for perception (Pylyshyn, 1984; Fodor, 2001). However, grounded (or embodied) cognition theories (Barsalou, 2008) together with neuroimaging studies of semantic systems (Martin, 2007; Binder et al., 2009) espouse another extreme, that knowledge representations emerge from weighted activity within property-based brain regions. The past century of neuropsychological research, including human brain lesion studies, has demonstrated that word form knowledge of different object categories is represented in part along distinct brain regions (Damasio et al., 1996; Martin et al., 1996; Moore and Price, 1999; Grossman et al., 2002), notably for representing living versus non-living things (Warrington and Shallice, 1984; Hillis and Caramazza, 1991; Silveri et al., 1997). This living versus non-living category boundary in neuropsychological models further inspired research regarding other semantic category representations in the brain. In the realm of visual object recognition, specific brain lesions were found to disproportionately impair, or spare, semantic knowledge of various visual object categories, including animals, tools and artifacts, famous people, and fruits & vegetables (Tranel et al., 1997; Caramazza and Mahon, 2003). Questions remained, however, as to how all these visual object and action categories might relate to representations of potentially distinct categories of “auditory objects”.

From the perspective of grounded cognition theories, a couple studies sought to assess whether a cortical organization respecting a living versus non-living boundary would hold true of real-world sounds (Engel et al., 2009; Lewis et al., 2011b). However, because vocalizations are acoustically characterized and distinguished by strong harmonic content, to facilitate data interpretation (addressed in Section 2.1) the sound stimuli used were restricted to those devoid of any vocal content. The category of living things included human action sounds (readily recognized as a human agent instigating the action) versus non-human action sounds, which were not as easily emulated or mimicked by humans (e.g. horse galloping). The non-living category included sounds of the natural environment (e.g. wind blowing through trees) versus mechanical sounds of automated machinery (e.g. watch ticking or laundry machine tumbling). Importantly, the machinery and mechanical action sounds were judged as not being instigated by a human or living agent. In short, brain regions responsive to correctly categorized living (biological) action sounds (Fig. 3, contributions to yellow hues) were strikingly distinct from those for non-living (non-biological) sounds (contributions to blue hues). The perception of biological action sounds was correlated with activation in numerous motor-related and audio-motor association regions, plus portions of the thalamus, basal ganglia and cerebellum. This was consistent with the idea that the listener was “embodying” the sounds (especially human and to a lesser extent animal actions). Presumably, participants were effectively comparing or probabilistically matching the sound stimuli (incoming sound streams) to their own repertoire of sound-producing motor actions to attain a sense of recognition. The findings that human action sound processing recruits premotor and motor-related cortices has also received support from other studies using fMRI (Bidet-Caulet et al., 2005; Gazzola et al., 2006) and EEG (Pizzamiglio et al., 2005; De Lucia et al., 2009).

In contrast to living things, the processing of sounds produced by non-living things, which are not as readily or easily embodied, preferentially recruit brain regions commonly associated with high level visual processing, episodic memory, and other network processes that remain to be fully resolved (Engel et al., 2009). Sounds produced by non-living things preferentially activated occipito-parietal cortical regions typically regarded as visual areas, plus the parahippocampal gyri and posterior cingulate cortices. In a follow-up study the same participants were subsequently highly familiarized with all the sound stimuli and re-tested in the fMRI scanner (Lewis et al., 2011a), which also resulted in preferentially activated regions of cortex (though with some variations after perceptual learning) that further supported the four-fold dissociation of action sound subcategories in the proposed model. Thus, the living versus non-living double-dissociation for cognitive level processing organization (i.e. for word form and visual object processing) also persisted in the realm of hearing perception.

Conspecific versus non-conspecific action sound processing differences thus far appear to be mostly reflected as a matter of degree of activation of regions in a network rather than the recruitment of any unique brain regions outright. For instance, the left and right posterior insulae were more strongly activated by action sounds that were clearly perceived as being caused by an animal versus a human agent under different listening tasks or with different listening experiences, and were presumed to have functions related to audio-tactile or audio-sensorimotor associations (Engel et al., 2009; Lewis et al., 2011a; Webster et al., 2017). The human action sounds, relative to animal actions, more prominently activated a left-lateralized fronto-parietal and pSTS/pMTG network, overlapping mirror neuron systems (MNS), classically defined as involving the inferior parietal lobule, inferior frontal gyrus plus ventral premotor cortex (Rizzolatti and Craighero, 2004; Molenberghs et al., 2012).

This supported the notion that action sounds produced by living things, human conspecifics, and to some extent non-human animals, were effectively being 'embodied' for purposes of sound categorization. As mentioned in Section 2.2, the pSTS/pMTG complexes have roles in dynamic temporal processing of object actions, whether viewed, heard, or manipulated, and thus have metamodal or supramodal functions. They are also regions reported to have prominent roles in social cognition (Pelphrey et al., 2004; Jellema and Perrett, 2006; Zilbovicius et al., 2006). Thus, reading subtleties of human expressions and body language together with associated sounds produced by biological actions may in part be represented there, conveying information that helps guide social interactions.

3.1. Listener extremes

To further probe the possible function(s) of the different brain regions and boundaries identified for processing action sounds in the proposed model, we next examine evidence from the brain organization in listeners who effectively grew up in different multisensory environments, including left-handers and early blind listeners.

3.1.1. Influence of handedness on action sound processing—Language and music processing in the human cortical hemispheres show strong lateralizations that are proposed to have their origins in gestural networks (Johannesson, 1950; Hewes, 1973; Grigor'eva and Deriagina, 1987; Corballis, 1999; Li et al., 2000; Zatorre et al., 2002; McNamara et al., 2008; Morillon et al., 2010), which in turn are thought to have evolved from adaptations leading to handedness (Lausberg et al., 2003; Meguerditchian et al., 2013). In right handed listeners, hearing and categorizing uni-manual tool-use sounds (a subcategory of human action sounds) led to network activation that overlapped with regions independently activated during pantomime of tool use with the dominant hand (Lewis et al., 2005). This included activation of the left hemisphere somatosensory- and motor-related networks, most notably including the left inferior parietal lobule (IPL; Fig. 3, green solid outline). This was consistent with reported left-lateralized MNS-like networks (Kohler et al., 2002; Rizzolatti and Craighero, 2004; Molenberghs et al., 2012), and thus consistent with brain organizations typical of right-handed individuals.

To examine the effects of handedness on these hearing perception networks, strongly left-handed participants were recruited to perform the exact same listening task involving uni-manual tool-use sounds, and performing and the virtual tool manipulation task with their dominant left hand (Lewis et al., 2006). The left-handers, who grew up associating the sounds of uni-manual tool use predominantly with their dominant left hand, recruited strong activation along the right hemisphere IPL (cf. Fig. 3, dotted versus solid green outlines)—this IPL activation focus thus effectively flipped sides! This difference in activation pattern was concluded to be reflecting learned audio-motor associations, linking hearing perception with hand and arm movements (motor action schemas) associated with the dominant hand. This lent strong support for embodied cognition accounts of knowledge representation relating to audition. Knowledge of tool making and tool use, and thus of tool use sound processing, is arguably more highly developed in humans than any other species (Paillard, 1993). Thus, the proposed model regards tool-use sound processing as an extension of human action sound representations (Fig. 1, plain text), wherein body schemas may become adapted to perceptually assimilate objects as extensions of one's body (Paillard, 1993; Iriki et al., 1996). This effect of handedness on sound processing pathways illustrates a compelling instance of how top-down or cross-sensory processing influences the organization of cortical networks mediating hearing perception.

Interestingly, in some ancient Indian philosophical systems, such as the Abhidharma, classifications of sound objects originating some 2500 years ago expressed main categories similar to some of those in the proposed model (Fig. 1), including: (1) sounds that originate from conscious elemental causes such as the voice of a sentient being or a finger snap; (2) sounds that originate from unconscious elemental causes such as the sounds of a river and the wind; and (3) sounds that originate from both conscious and unconscious elements such as a drum beat (Mipham, 2000). The latter category is similar to tool use sounds which we classified in our model as an action sound by a 'living thing'. This is in accord with the clarification that animate sounds can "express meaning" while inanimate sounds do not, and is generally supported by neuroimaging data, which ancient philosophers would not have had access to. The high correspondence of our model with how philosophers thought about sound object categories in earlier millennia supports the universality of the conceptual qualities of the model.

3.1.2. Influence of blindness on action sound processing—*As* addressed earlier, the pSTS/pMTG complexes overlapped with regions involved in visual biological motion processing (e.g. lip reading, running), clearly encroaching on cortices regarded as visual areas. However, individuals who have never had visual experience (early blind listeners) are clearly fully capable of recognizing and identifying action sounds. Examining brain responses of congenitally blind listeners (Lewis et al., 2011b), human action sounds deemed as recognized, in contrast to backward played version that were not recognized, also recruited the pSTS/pMTG complex in addition to portions of occipital cortices associated with visual processing (Fig. 3, yellow outlines). Thus, bilateral pSTS/pMTG regions appear to be recruited even in the absence of visual motion experience with the action sounds. Incidentally, when testing for brain activation in response to the four action sound subcategories described earlier, using human, animal, mechanical, and environmental sound stimuli (Engel et al., 2009), a four-fold dissociation of cortical networks was similarly observed in some early blind individuals (n = 2; unpublished data). Thus, while visual experience influences cortical network pattern recruitment, categorical boundaries for hearing perception nonetheless persist in the absence of vision. How might one reconcile the audio-visual representations of objects given that categorical sound perception organizations appear to develop in cortex even in the absence of visual experience?

The functional roles of the pSTS/pMTG in hearing perception are perhaps best interpreted in the context of 'metamodal operators' (Pascual-Leone and Hamilton, 2001). In this theoretical framework, different brain regions may be genetically "pre-wired", developing microcircuitry that happens to be efficient for conducting certain types of operations. Thus, for instance, if an individual has visual input, then regions such as the pSTS/pMTG complex will compete to perform relevant processing to establish representations related to the meaningfulness of the sensory event, typically being recruited for biological visual motion processing functions in sighted individuals. However, in the absence of vision, these cortices still compete to perform certain operations germane to biological action processing. Thus, the pSTS/pMTG complexes appear to play a functional role in transforming the spatially and temporally dynamic features of natural action event information into a common neural code (for audition, vision or touch), and may form a reference frame for probabilistically comparing the predicted or expected incoming auditory (and/or visual and haptic) information based on what actions have already occurred in real time (Lewis, 2010). These regions appear to be ideally suited for processing action sound sequences, especially for sounds produced by living (biological) things, less so for non-embodiable non-living things, and lesser still for vocalizations that cannot be directly viewed (at the vocal cords/laryngeal source)–the latter two of which have fewer intermodal invariant attributes in general. Thus, features of biological motion sequence processing appear to generally relate to operations of the pSTS/pMTG regions.

Interestingly, our meta-analysis revealed regions activated by human action sound processing in the early blind listeners that overlapped with regions activated by non-living action sound processing by the sighted group (Fig. 3, yellow outlines overlapping blue cortex). Thus, a number of cortical territories commonly allotted to the visual system may be more accurately regarded as metamodal operators that compete to process signals from whatever sensory input happens to be available, which in sighted individuals would typically be bottom-up visual signal inputs.

Another study with congenitally blind listeners examined brain regions activated upon hearing hand-made human action sounds, which revealed activation of motor-related networks defined as MNS networks (Ricciardi et al., 2009). However, in the congenitally blind study addressed early (Lewis et al., 2011b) only the sighted control group recruited significant activation in MNS-like networks relative to the early blind group. The activation pattern differences between groups appeared to differ in degree of activation, suggesting that blind participants may have effectively opted to use a different cortical processing strategy for “recognizing” the action sounds given the two-alternative forced choice task (recognized or not recognized). In particular, blind listeners recruited network structures implicated in episodic memory, rather than MNS-like networks implicated in embodiment or procedural memory, as their default strategy (Wagner et al., 2001; Yonelinas et al., 2005). Hence, the brains of blind individuals may adapt to the lack of visual motion input by preferentially using different encoding/decoding strategies to more efficiently represent auditory objects (or action events) for purposes of recognition. Thus, a sound-source may be deemed as “recognized” by different individuals by using completely different strategies (an issue of qualia). This further complicates what is meant by auditory object “recognition” in the human mind (Box 3).

In sum, the data pertaining to the brains of blind listeners to date buttress the idea that acoustic-semantic universals are driving, or are being utilized by, the cortical organization for auditory processing that respects the three major categories of natural sound (Fig. 1). Moreover, this organization is not critically dependent on visual experience. Furthermore, handedness influences which networks are ultimately adapted as extended portions of the “auditory system”. Thus, the underlying cortical network architecture genetically established at birth (prior to audio-visual and audio-motor associative learning) appears to include metamodal operator network architectures that are well suited for processing acoustic-semantic signals of natural sounds, as outlined in the proposed model, which is formalized next.

4. General tenants of the acoustic-semantic model for hearing perception

Now having reviewed much of the literature leading to the development and constraints of the proposed model of hearing perception for natural sounds, we next formalize some of the general tenants (4.1–4.4) that will have broader implications for the study of sensory perception, neurolinguistics, spoken language evolution, auditory cognition, and potentially other fields of neurocognition.

4.1. Parallel hierarchies process increasing degrees of information content

Our model presumes that, through experience, the brain further organizes beyond nascent networks in order to optimize the representation and processing of sensory information (leading to memory formation) that may occur through classic Hebbian-like mechanisms or other network-level mechanisms of sensory encoding (Dosher and Lu, 2009). The neural representations of sound events thus likely propagate simultaneously along many of the cortical pathways, reaching various intermediate processing stages or tiers (Fig. 3, all colored cortices). Network activations may continue until leading to a “stable” activation pattern that matches a ‘memory trace’ (episodic, procedural, semantic), reflecting a local minimum state (Hopfield and Tank, 1985) that mediates or confers a sense of successful recognition.

Conspecific, versus non-conspecific, action sounds and vocalizations in general would arguably be more familiar and behaviorally relevant for each given species. Thus, hearing conspecific sounds in isolation would likely tap into the greater depths of how they had been encoded over life-long experiences by the individual, which in turn often translates to more expansive network activations in neuroimaging paradigms. However, there are a few caveats with this simplistic interpretation. Perceptual learning studies using visual, tactile, or auditory processing have revealed a number of cortical network mechanisms for memory encoding and retrieval. One is that perceptual learning can lead to greater activation as the stimulus takes on greater degrees of behavioral relevance, newly engaging regions previously not recruited or not to as great an extent (Gauthier et al., 1999). A second is that greater familiarity can lead to less, rather than more, network activation over time due to the networks becoming more efficient and/or faster (sharpening and facilitation models) at processing specific stimuli (Wiggs and Martin, 1998). A third is that network patterns may undergo outright changes in which different brain regions become recruited (scaffolding mechanisms) with experience (Petersen et al., 1998; Lewis et al., 2011a). Another consideration is that when a sound is heard and it is unclear exactly what the sound is (but not confabulated as belonging to a wrong category) it can lead to greater degrees of activation in a given network as it effectively continues to “try” to settle on a probabilistic solution (Lewis et al., 2005). Consequently, the neural correlates of sound perception may be obscured in a given study depending on the specific stimuli used, the nature of the task demands (Bracci et al., 2017), and listener biases, as well as the spatial and temporal resolution of the imaging modality (Santoro et al., 2014). Thus, when exploring the processing of different semantic categories, one may be activating or revealing a cross section through several parallel hierarchical stages of sound representation, which can easily lead to complicated interpretations of the specific functional roles of different brain regions.

4.2. Metamodal operators guide sound processing network organizations

Various cortical regions may be genetically established and interconnected so as to excel at performing specific types of operations, regardless of what sensory input they happen to receive (or fail to receive), and have been termed “metamodal operators” (Pascual-Leone and Hamilton, 2001). For the studies of hearing perception in the blind, this concept of metamodal operator mechanisms is particularly elegant for explaining patterns of activation in what are traditionally regarded as “visual cortices” (Section 3; Fig. 3, yellow outlines). Such processing regions may compete to perform operations with certain types of acoustic attributes, or universal acoustic-semantic attributes, which consequently shape the organization(s) of networks that ultimately mediate hearing perception. The natural sound categories in the proposed model may thus be reflective of whole-brain level strategies optimized for encoding meaningfulness to learned sound stimuli.

4.3. Natural sounds are embodied when possible

Outside of innately pre-wired acoustic reflex circuits (e.g. startle reflexes, emotional communication sounds with newborn infants), a primary cortical mechanism behind encoding and recognizing natural sounds appears to be to “embody the sound if possible”. In this regard, when a sound is heard in isolation the brain attempts to match acoustic inputs with learned audio-motor or audio-sensorimotor association representations, which appear to be lateralized to the left hemisphere. This notion is consistent with the idea that a number of cognitive functions become more lateralized to one hemisphere (e.g. spatial skills, handedness) to minimize interhemispheric “wiring” and thereby being more efficient (Preuss, 2011). Evidence of embodied representations for natural sound processing was perhaps most striking when comparing left-handed versus right-handed listeners upon hearing uni-manual tool-use sound (Section 3), in that certain functional loci flipped sides (especially in inferior parietal cortex), consistent with dominant hand audio-motor association embodiment.

Even music can be embodied in perceptual systems. For example, one study found that individuals listening to piano pieces after (versus before) training in how to play the piece with their own hands newly led to activation in motor-related networks (Lahav et al., 2007). Thus, while the complexities of music appreciation may be represented among widespread brain regions (Koelsch et al., 2004; Zatorre et al., 2007), experience with musical sound production can influence which brain networks are recruited when hearing a musical piece —“feeling” the music emotionally, and/or embodying it technically in motor systems (also see Section 6).

4.4. Categorical perception emerges in neurotypical listeners

The model presumes that neuronal organizations or representations of certain fundamental acoustic-semantic categories develop both in neurotypical human listeners and presumably other mammalian species with hearing ability (and perhaps necessarily with auditory communication ability as well). We further presume that other sub-categories may develop with listening expertise (e.g. birders, hunters, musicians), which may develop with dependence on visual, sensory-motor, and other multisensory inputs or contexts. Conversely, categorical perception may fail to fully develop, such as for some individuals on the autism spectrum who have difficulty with generalizing objects into categories (e.g. difficulty conceiving all dog barks as belonging to one specific subcategory; “dog barks”) (Grandin, 2008), which presumably renders some of the model boundaries as less distinct for those listeners. Thus, targeted interventions may focus on training young individuals to hear different natural sounds as belonging to distinct semantic categories in an effort to tap in to nascent cortical circuitries that may ultimately develop to become more efficient for representing acoustic-semantic knowledge or meaning at a categorical level, as addressed with other issues of potential clinical significance in the following sections.

5. Model implications and predictions

As mentioned in the introduction, the proposed model of hearing perception should have impact on multiple lines of thinking across fields of sensory, multisensory, and cognitive neurosciences in both human and non-human animals, as addressed below.

5.1. Fundamental advances in understanding how perception functions

As our model focuses on a lesser studied modality, audition, and incorporates both bottom-up and top-down influences on object perception, we hope to spur thinking about how different modalities can inform perception, and how their unique properties can shape how that modality is used by the individual and why. Because sound, in contrast to vision, only comes from motion (changes in energy to produce sound pressure waves), the auditory system is likely to be more heavily dominated by representations for agent intention and meaning, potentially as a vestige of sound processing crucial for survival. Stationary visual objects may become of interest to us from a top-down perspective, so although meaning is still key for object perception in any modality, the visual world can accommodate many more immediately meaningful perceptions of segmented objects compared to audition, where auditory object perception may require several seconds of sound events to unfold to accurately convey a sense of recognition. Thus, this may make audition an ideal model system for advancing fundamental constraints of multimodal models of object perception.

We further assert that our model may contribute to the study of object representation in general, as well as how factors such as intention, learning, and action influence perception. Section 4 elucidated multiple tenants set forth by the model that are likely applicable across other sensory modalities. As the auditory faculty has attributes that are either absent or novel compared to vision, one can utilize audition to further tease apart what qualities of object perception are truly modality invariant versus dependent as it relates to object perception in brain network representations.

5.2. Understanding spoken language development in children

Vocal imitation and mimicry, together with sound symbolism, are known to play a crucial role in a child's spoken language neurodevelopment (Kuhl and Meltzoff, 1982; Rhoades, 2007; Imai et al., 2008, 2015; Ozturk et al., 2013). A number of theories suggest that aspects of vocal communication should show a resemblance to properties of sensory referents, as formalized in theories of sound symbolism and iconicity (Imai and Kita, 2014; Perniss and Vigliocco, 2014).

Humans infants are known to be sensitive to voices and speech sounds shortly after birth. Two-day-olds prefer their native language (Moon et al., 1993; Beauchemin et al., 2011; Sato et al., 2012), and are thought to be learning the significance of maternal voices as early as during fetal development, showing responsiveness (e.g. heart rate changes) to speech produced in the child's mother tongue (DeCasper et al., 1994), which may be influencing the development of hearing perception proto-networks *in utero* (DeCasper and Fifer, 1980; DeCasper et al., 1994; Kisilevsky et al., 2003, 2009). Thus, one future direction would be to determine if acoustic-semantic universals of vocalizations are the first to start driving auditory system development (perhaps *in utero*), followed by later stages that are driven by processing of other acoustic-semantic universals that may emerge later in post-natal life. In this regard, some aspects of the proto-language networks that will develop to mediate spoken language processing and perception may depend on the development of rudimentary semantic networks (for categorical perception) before symbolic linguistic representations can be properly formed and organized. This neurodevelopmental mechanism might prove to help explain some etiologies of spoken language delays in children (Sheridan, 1959; Stothard et al., 1998).

5.3. Understanding central hearing deficits and recovery after brain injury

Outside of peripheral hearing loss, we know from examples of central auditory disorders that damage to specific brain regions can lead to a variety of different hearing deficits, including agnosias that may be specific for environmental sounds, timbre, rhythm, words, melody in music, and in rare cases for more specialized sound object categories (Goll et al., 2011; Trumpp et al., 2013). This indicates that there are separable functions and processes mediating hearing perception in the human brain, similar to the visual system, which has led to the idea that brain networks mediating *categorical* perception are also a likely hallmark feature of the human auditory system (e.g. Section 4.4). However, pure associative agnosias for specific categories of sound other than voice and melodies are rare (Saygin et al., 2003). This may be due to the nature of how natural sounds, which may require relatively longer stimulus durations to unambiguously identify, and can be confabulated and completely mis-categorized to the satisfaction of the listener. For instance, animal vocalizations that were misperceived by individual listeners as representing tool sounds (in a two alternative forced choice task) were correlated with cortical activation of "tool sound processing" networks (Lewis et al., 2005). Thus, sound confabulation represents a feature of the auditory system that may make auditory object processing deficits trickier to assess neuropsychologically relative to visual object processing deficits. Further advances in defining the processing mechanisms associated with the different category boundaries for hearing perception relative to other modalities will likely help develop clearer taxonomies for describing, diagnosing and developing interventions for individuals recovering from auditory cognitive deficits, and guide targeted interventions at perhaps more rudimentary semantic levels as a key to neurorehabilitation.

5.4. Advances in biomimetic hearing aid designs

In an effort to help people who suffer from hearing loss, models of how the brain processes sound have led to ideas for engineering biologically-inspired ("biomimetic") hearing aid algorithm designs (Wang and Shamma, 1994; Smith and Fraser, 2004; Coath et al., 2008; Shannon, 2012). A need still persists for the continued development of both smaller devices and more "intelligent" designs (NIDCD, 2009) that are effective for different listening environments. Selectively amplifying frequency bandwidths characteristic of human speech represents one strategy (Harkins and Tucker, 2007; Takahashi et al., 2007), though this does not always allow a listener to segregate, for instance, the sounds of one person speaking in the presence of a crowd of people, or to tolerate the noises heard while chewing food. Further efforts have and will continue to require not only considerable improvements to the hardware implementing them, such as small size, ultralow power consumption, field programmability (Rumberg and Graham, 2015), but also to the simultaneous development of more sophisticated algorithms that appropriately suppress background acoustic noise based on probabilistic sound signal profiles to better enhance signals of interest (Takahashi et al., 2007; Chung and McKibben, 2011; Lowery and Plyler, 2013)—potentially capitalizing on some of the putative acoustic-semantic universal attributes proposed herein. For instance, biomimetic designs may be able to capitalize on capturing temporally correlated signals of harmonic profiles and power spectra profiles to effectively enhance sounds that are characteristic of a single natural sound-source category, and filter out acoustic signatures characteristic of background acoustic noise, thereby performing acoustic accommodation filtering on the front end.

5.5. Advances in anthropological models of oral communication in hominins

The proposed model has implications regarding literature debating the acoustic versus gestural origins of spoken language systems (Darwin, 1871/1981; Hewes, 1973; Liberman and Mattingly, 1985). *Mimesis*, the ability to produce self-initiated representational acts (also see glossary), is thought to represent one of the earliest forms of cognitive-motor abilities that distinguished hominins (e.g. *homo erectus* and possibly *homo habilis*) from the great apes (Hewes, 1973; Grigor'eva and Deriagina, 1987; Donald, 1991), and the above theory purports gesture movements as an early form of communication that predated vocal language.

Mimicking the events and sounds of the natural world, and conveying propositional communications, may have constituted a form of pre-linguistic communication. This includes big game pre-hunt organizations, teaching complex skills to other troop members, interpretive dance, and numerous other socializing events that would help stabilize larger groups of individuals in a community. Some oral communication advancements presumably enabled hominins to detect and interpret increasing degrees in nuances of emotional states and intentions of conspecifics through an ability to produce and perceive non-stereotyped oral communications (e.g. acoustically conveying jealousy, love, triumph) (Donald, 1991), which in extant humans are largely processed in cortical networks lateralized to the right hemisphere. Over roughly the last 100,000 years, the auditory and semantic systems of humans are thought to have evolved further to accommodate high speed articulated speech perception, likely through exapting circuits used for gestural communication planning and generative praxis (MacNeilage, 1998; Corballis, 1999; Rizzolatti and Craighero, 2004; Arbib, 2005; Corina and Knapp, 2008; Stout et al., 2008), in that in most individuals language is left lateralized. Eventually, vocal sounds largely supplanted manual gestures as the main form of communication, permitting substantially faster communication of ideas, communication in total darkness, and communication over greater distances through visual barriers like dense forests. Recent theories posit that some of the earliest categories of natural sounds that needed to be orally mimicked would likely have included incidental sounds of locomotion, tool-use sounds, and vocal calls of other animals (Falk, 2004b; Larsson, 2014, 2015), which are consistent with the boundaries of the proposed model.

The brain regions related to language processing may be rooted in evolutionarily earlier systems when gestural mimesis prevailed. This ‘default’ gestural origin theory is supported by the proposed model and meta-analysis data, which shows that *living action sounds*, and not conspecific vocalizations per se, predominantly activate fronto-parietal (motor-related) regions in the left hemisphere (Fig. 3, yellow): This includes regions commonly associated with language reception (e.g. Wernicke’s area) and production (e.g. Broca’s area). Thus, the evolution of hearing perception systems in modern humans, based on vestiges of how the brain appears to be organized for natural sound processing (Fig. 3), appears to have been closely tied to the ability to produce and interpret communicative action sounds produced by living things as a semantic category.

With regard to human and primate communication, the proposed model thus leads to a number of predictions or questions. One is that conspecific action sounds for other species, including sound producing body-action asymmetries of great ape species (Cashmore et al., 2008), might also show evidence of lateralized networks for processing and conveying a sense of meaning (and perhaps related to the degree of primate handedness). Another prediction is that the organizational principles for sound processing at a categorical level should also be respected in organizations for sound production or mimesis at a categorical level. For instance, oral production and orchestration of non-linguistic natural sound events (imitation, mimicry and/or mimesis) might entail motor planning networks that also respect the category boundaries of the proposed model.

Regarding language, a third prediction is that cognitive architectures for phrase-level language comprehension should at some level also respect the major category boundaries. Models of language and cognition suggest that parallel hierarchies entail perceptual-semantic links ranging from lower sensory signal features, to auditory/visual object representations, to situations/events, and to abstract ideas (Perlovsky, 2011). Onomatopoeia represents one level of linkage (Hashimoto et al., 2006) though this does not adequately explain the more highly symbolic levels of language (Imai and Kita, 2014). However, natural sounds at a category level more generally may prove to correlate with the linguistic concept representations associated with short spoken phrases or utterances. For instance, grounded cognition models (Barsalou, 2008) would predict that the comprehension of short spoken phrases describing sound-producing events (e.g. “wind storm”) should engage at least some of the same brain regions that demonstrated category-specificity to perceptual-level processing of those corresponding sound events (i.e. perceiving the sounds of a wind storm), independent of the language(s) used. If verified, this would lead to new lines of research for characterizing the nature of perceptual-linguistic links in the brain; with the idea that linguistic-semantic systems may largely be grounded in perceptual-semantic systems. Acoustic-semantic universals may thus represent one form of natural sensory signal attribute to help bootstrap cortical networks for cognition. These avenues of research may thus shed light on the theories behind both the phylogeny and ontology of spoken language systems from more of a “bottom-up” perspective.

6. Limitations of the model and issues for future research

While our neurobiological model accounts for much of the neuroimaging data from humans and other mammals to date, there are a number of limitations of the model and need for future research, as addressed below.

6.1. Category refinement and task dependency

The chosen categories and subcategories of sound-producing events illustrated in the model accounted for most if not all of the observed brain processing organizations for non-linguistic hearing perception. However, when considering a global model that extends to other sensory perception domains, these may not necessarily be the most informative category division or subdivision definitions. In order to make direct comparisons across modalities, one must ask if these categories fully correlate with those proposed in the visual and haptic modalities, and in what semantic contexts. Studies that compare objects of our included categories that can be recognized and separately presented as auditory, visual and ideally haptic stimuli may help address the cross-modality applicability of this model. Because audition is different, certain unique attributes may not extend to other modalities and it is worth exploring these similarities and differences. The weighted degree of activation of a specific brain region in a given network may vary based on experience, as mentioned earlier, and studies examining expertise versus novice observers, as well as behaviorally manipulating context and task demands may help refine this model.

6.2. Processing of music, emotional and threatening sounds

The top-down influences of the affective and reward systems have only been superficially addressed in this model. Sounds from any of the categories can take on a highly positive or negative valence, and any of the categories may be used to elicit musical forms (rhythm and melodies), wherein music appreciation appears to be relatively unique to humans. Conceivably, the missing ‘fourth’ category of sound of the model, non-living vocalizations, may be reflective of music as a sound category at a conceptual level that can be found or learned to be appealing by different listeners. In the context of this review, both speech processing and music appreciation (enjoyable or disagreeable) are regarded as tapping into higher forms of acoustic communication that utilize brain systems that extend beyond the more rudimentary levels of natural sound recognition in the proposed model. This corresponds reasonably well with brain imaging studies of music in general (Zatorre et al., 2002; Salmi et al., 2016). Understanding how lower and higher level acoustic regions interact in the context of music has implications for evolutionary theories, as complex forms of music production and appreciation have also appeared to evolve in hominins dating back 2–3 million years ago (Donald, 1991).

Of course, music is an important part of human culture and experience, and even has therapeutic implications (Raglio et al., 2016). Testing the boundaries of the proposed model with different instruments from voice and voice-like (“non-living”) wind instruments, versus background percussion or artificially created sounds with manipulated attributes, could have interesting implications for testing category fluidity and figure/ground distinctions and music therapies. In addition, auditory processing of music may interact with speech vocal networks at both low and high levels. It may also have a salient link with affective processing, such as with song being used as a tool for memorization. As to how lower level auditory cortical processing serves as a bridge to these higher functions will be an important area of future research.

Further research is also needed to understand how the limbic system (for affect in general) and auditory perception systems interrelate beyond reflexive circuits, which should have important clinical implications. For example, in post-traumatic stress disorder (PTSD) specific acoustic-limbic circuits may become overly interactive (Schechter et al., 2012; Suo et al., 2015). Conversely, in conditions such as autism spectrum disorder (ASD) there may be a relative lack of acoustic-limbic interconnections that may lead to inappropriate (or a lack of) responses to threat or social/emotional sounds (Tecchio et al., 2003; Baranek et al., 2007; Linke et al., 2017; Lortie et al., 2017). The proposed model may provide a framework for understanding how learned acoustic signals affect, and are affected by, interactions with the limbic system, thereby leading to evidence-based research on targeted intervention for clinicians.

6.3. Cross-species comparisons of the model

A comparative cross-species study of brain organization for processing conspecific versus non-conspecific action sounds or vocalizations, perhaps testing both relative to non-living environmental sounds, would provide a robust method for further testing of whether this theoretical framework truly reflects a fundamental model underlying perception. For example, studying vocalization versus action sound processing in species who rely more on one or the other sound category as a form of communication or environmental interaction, or are more visual versus auditory dominant, would help refine the model.

6.4. Category and object-scene fluidity

The category of sound and how it is processed in the brain for a given auditory event may have some “fluidity”. As an example, the sound of an unseen cricket (a living, non-conspecific animal) in a closed environment, such as a tent or yurt, could be readily identified as an object. However, even though it is an action sound (with strong harmonic content), the general background of the cricket(s) or other nature sounds for most people might become a background acoustic scene, as addressed in Section 2.2. Even human conspecific speech sounds can become part of the cacophony background noise of a restaurant, evidenced perhaps by the fact background restaurant babble is a sound type that is rated as emotionally neutral in the International Affective Directory of Sounds (IADS) (Fernandez-Abascal et al., 2008). However, if in that restaurant one is attempting to covertly overhear a conversation, one's relationship to the processing stream would change and one may carefully attempt to parse the frequency, intensity and duration characteristics of the individual voice they are attempting to decipher.

The potential for fluidity is likely similar with all categories in our model, and it is currently unclear whether manipulating low level signal attributes within a category (such as harmonicity in vocalizations, constancy, or $1/f^{\alpha}$ in environmental sounds) can lead to category manipulation. Exploring cortical activation to the same set of natural sound stimuli from different categories when they are attentively or contextually cued as being either more 'object-like' versus more 'ground/scene-like' might also be an important test of the fluidity of the bottom-up versus top-down drivers of auditory cortical processing across categories.

6.5. Limitations of neuroimaging technology

Thus far, neuroimaging of auditory systems has been limited by the constraints of the methodologies available. Whole brain imaging currently can only be achieved when the participant is extremely still. Also, the environment of the fMRI is very noisy, requiring creative methodologies, such as event-related sparse sampling in which short clips of auditory stimuli are presented in quiet periods between brain slice acquisitions. Any action of the participant is limited to a short button press or utterance. Studying participants making natural vocalizations or action sounds, or naturally interacting with sound stimuli in conventional neuroimaging devices is thus rather difficult given the artificial constraints of such environments. Novel neuroimaging technologies currently being developed, including a wearable, upright positron emission tomography (PET) helmet with greater head motion tolerance that our group is developing (Bauer et al., 2016), will allow one to probe deeper into predictions generated by the model, accommodating whole brain imaging during natural movement such as speech, gestures, and many types of tool use. Virtual reality (VR) further lends promise for more immersive studies that can be well controlled, and one could use VR to study how the model functions in relation to action, motivation, emotion and attention in more natural settings. Wearable brain imaging technologies will allow for imaging with robust natural movements, and compatibility with headphones, VR, as well as EEG, and with neurotransmitter targeted ligands to examine different systems (Fig. 4). Advances in wearable brain imaging technology that allows greater spatial resolution and depth penetration (Boto et al., 2017), may mean exciting developments for future investigation of auditory processing in natural contexts.

7. Concluding remarks

We have proposed a simple, neurobiological acoustic-semantic model of hearing perception in which bottom-up and top-down influences interact, resulting in perception of auditory objects. The model takes into account bottom-up acoustic properties that are either instinctual or learned as belonging to either of three main types of semantically meaningful categories of natural sound: living action sounds, non-living action sounds, or vocalizations (living sources). Such principles may underlie neuronal processing mechanisms to efficiently direct sound signal processing, based probabilistically on a number of acoustic-semantic universals, to cortices or networks best adapted for conveying a sense of meaningfulness to the listener. Our model further accounts for top-down influences that reflect grounded ("embodied") cognition principles for both vocalization and action sound processing. We additionally explored multiple fields of research that could benefit from the model's framework. Auditory processing should ultimately be studied in the context of how a listener interacts with the sound-sources, and thus advances in interactive auditory stimulus delivery and neuroimaging technologies, such as wearable PET neuroimaging systems, will likely be crucial. The proposed model could serve as a test bed for predictions made by cross-species comparisons, for developing or refining taxonomies for cognitive deficits, for advancing models of spoken language evolution, and for refining models of auditory processing neurodevelopment trajectories in children.

Acknowledgments

This work was supported by the NCRR/NIH COBRE grant P20 RR15574 (to the Center for Neuroscience of West Virginia University) and subproject to JWL, and the West Virginia Clinical and Translational Science Institute (WVCTSI) research scholar program supported by the National Institute of General Medical Sciences, U54GM104942 to JBL.

Glossary of terms

Acoustic-semantic universals

A quantifiable acoustic parameter or set of parameters inherent to sounds of the natural world that probabilistically assigns a sound-source to membership of a distinct semantic category. In principle, the auditory system of all mammals has evolved intrinsic cortical micro-circuitry that efficiently develops to extract or segment sound-sources based on a number of universal signal attributes

Agnosia

The loss of ability to recognize the import of sensory stimuli or impressions. Different varieties of agnosia (e.g. pure auditory agnosia) are distinguished by different sensory modalities of functions

Auditory object

A collection of acoustic data bound in a common perceptual representation and disambiguated from other events in an auditory scene. This term has variations in definition in different fields of study that are addressed in this review, and thus is an operationally defined term

Episodic memory

The system that allows one to remember (consciously recollect) past experience of autobiographical events, reflecting concrete or time-bound memory

Event perception

The ability to perceive complex, usually moving, clusters and patterns of stimuli as a unit. This contrasts with characteristics of object perception in that it further takes into account motion and context

Hominin

Humans (*Homo sapiens sapiens*) and their closest non-extant relatives (e.g. *homo habilis, homo erectus*). The term hominid includes the great apes

Mimesis

The ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic (e.g. charades, pantomime, ritual dance). Mimicry is different from mimesis in that it is more literal as an attempt to render an exact duplicate of an observed act. Imitation, found especially in monkeys and apes, is also different, wherein mimesis adds the element of invention of intentional representations

Object

An *object* is loosely defined as “a thing, person, or matter to which thought or action is directed” (Random House dictionary). All objects (visual, auditory or haptic), however are seemingly defined by ‘meaning’ and therefore, what constitutes an object implies some degree of *fluidity* as meaning can be different to different organisms and may even change over the lifespan of a given organism. The status of an object (auditory or otherwise) may depend on the perceivers a) experience, b) faculties and capabilities, and c) circumstances

References

Adams RB, Janata P. A comparison of neural circuits underlying auditory and visual object categorization. Neuroimage. 2002; 16:361–377. [PubMed: 12030822]

Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW. Task-modulated “what” and “where” pathways in human auditory cortex. Proc Natl Acad Sci USA. 2006; 103:14608–14613. [PubMed: 16983092]

Andics A, Gacsi M, Farago T, Kis A, Miklosi A. Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Curr Biol. 2014; 24:574–578. [PubMed: 24560578]

Arbib MA. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav Brain Sci. 2005; 28:105–124. (discussion 125–167). [PubMed: 16201457]

Arnott SR, Binns MA, Grady CL, Alain C. Assessing the auditory dual-pathway model in humans. Neuroimage. 2004; 22:401–408. [PubMed: 15110033]

Arnott SR, Cant JS, Dutton GN, Goodale MA. Crinkling and crumpling: an auditory fMRI study of material properties. Neuroimage. 2008; 43:368–378. [PubMed: 18718543]

Austin, JL. How to do things with words. 2d. Oxford Eng.: Clarendon Press; 1975.

Bar M. Visual objects in context. Nat Rev Neurosci. 2004; 5:617–629. [PubMed: 15263892]

Bar M, Tootell RBH, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM. Cortical mechanisms specific to explicit visual object recognition. Neuron. 2001; 29:529–535. [PubMed: 11239441]

Baranek GT, Boyd BA, Poe MD, David FJ, Watson LR. Hyperresponsive sensory patterns in young children with autism, developmental delay, and typical development. Am J Ment Retard. 2007; 112:233–245. [PubMed: 17559291]

Barrett DJ, Hall DA. Response preferences for “what” and “where” in human non-primary auditory cortex. Neuroimage. 2006; 32:968–977. [PubMed: 16733092]

Barsalou LW. Grounded cognition. Annu Rev Psychol. 2008; 59:617–645. [PubMed: 17705682]

Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. Trends Cogn Sci. 2003; 7:84–91. [PubMed: 12584027]

Bauer CE, Brefczynski-Lewis J, Marano G, Mandich MB, Stolin A, Martone P, Lewis JW, Jaliparthi G, Raylman RR, Majewski S. Concept of an upright wearable positron emission tomography imager in humans. Brain Behav. 2016; 6:e00530. [PubMed: 27688946]

Baumgart F, Gaschler-Markefski B, Woldorff MG, Heinze HJ, Scheich H. A movement-sensitive area in auditory cortex. Nature. 1999; 400:724–725. [PubMed: 10466721]

Beauchamp M, Lee K, Haxby J, Martin A. Parallel visual motion processing streams for manipulable objects and human movements. Neuron. 2002; 34:149–159. [PubMed: 11931749]

Beauchamp MS, Lee KM, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 2004a; 41:809–823. [PubMed: 15003179]

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 2004b; 7:1190–1192. [PubMed: 15475952]

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]

Belenkov N, Goreva OA. [The role of the posterior colliculi of the corpora quadrigemina in realizing the orienting reflex]. Zh Vyss Nerv Deiat Im I P Pavlov. 1969; 19:453–461.

Belin P, Zatorre R. ‘What’, ‘where’ and ‘how’ in auditory cortex. Nat Neurosci. 2000; 3:965–966. [PubMed: 11017161]

Belin P, Zatorre RJ. Adaptation to speaker’s voice in right anterior temporal lobe. Neuroreport. 2003; 14:2105–2109. [PubMed: 14600506]

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

Bi Y, Wang X, Caramazza A. Object Domain and Modality in the Ventral Visual Pathway. Trends Cogn Sci. 2016; 20:282–290. [PubMed: 26944219]

Bidet-Caulet A, Voisin J, Bertrand O, Fonlupt P. Listening to a walking human activates the temporal biological motion area. Neuroimage. 2005; 28:132–139. [PubMed: 16027008]

Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing E. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10:512–528. [PubMed: 10847601]

Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 2009; 19:2767–2796. [PubMed: 19329570]

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. J Neurosci. 1997; 17:353–362. [PubMed: 8987760]

Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences. 1993; 17:97–110.

Boto E, Meyer SS, Shah V, Alem O, Knappe S, Kruger P, Fromhold TM, Lim M, Glover PM, Morris PG, Bowtell R, Barnes GR, Brookes MJ. A new generation of magnetoencephalography: room temperature measurements using optically-pumped magnetometers. Neuroimage. 2017; 149:404–414. [PubMed: 28131890]

Bracci S, Daniels N, Op de Beeck H. Task Context Overrules Object- and Category-Related Representational Content in the Human Parietal Cortex. Cereb Cortex. 2017

Bregman, AS. Auditory Scene Analysis. MIT Press; Cambridge, MA: 1990.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Brain Res Cogn Brain Res. 2000; 9:227–238. [PubMed: 10808134]

Burianova H, McIntosh AR, Grady CL. A common functional brain network for autobiographical, episodic, and semantic memory retrieval. Neuroimage. 2010; 49:865–874. [PubMed: 19744566]

Calvert, GA., Lewis, JW. Hemodynamic studies of audio-visual interactions. In: Calvert, GA.Spence, C., Stein, B., editors. Handbook of multisensory processing. MIT Press; Cambridge, Massachusetts: 2004. p. 483-502.

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol. 2000; 10:649–657. [PubMed: 10837246]

Campanella S, Belin P. Integrating face and voice in person perception. Trends Cogn Sci. 2007; 11:535–543. [PubMed: 17997124]

Campbell R. The processing of audio-visual speech: empirical and neural bases. Philos Trans R Soc Lond B Biol Sci. 2008; 363:1001–1010. [PubMed: 17827105]

Caramazza A, Shelton JR. Domain-specific knowledge systems in the brain the animate-inanimate distinction. J Cogn Neurosci. 1998; 10:1–34. [PubMed: 9526080]

Caramazza A, Mahon BZ. The organization of conceptual knowledge: the evidence from category-specific semantic deficits. Trends Cogn Sci. 2003; 7:354–361. [PubMed: 12907231]

Cashmore L, Uomini N, Chapelain A. The evolution of handedness in humans and great apes: a review and current issues. J Anthropol Sci. 2008; 86:7–35. [PubMed: 19934467]

Chung K, McKibben N. Microphone directionality, pre-emphasis filter, and wind noise in cochlear implants. J Am Acad Audiol. 2011; 22:586–600. [PubMed: 22192604]

Clarke S, Thiran AB, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran JP. What and where in human audition: selective deficits following focal hemispheric lesions. Exp Brain Res. 2002; 147:8–15. [PubMed: 12373363]

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]

Cooper RP, Aslin RN. Preference for infant-directed speech in the first month after birth. Child Dev. 1990; 61:1584–1595. [PubMed: 2245748]

Corballis MC. The gestural origins of language. Am Sci. 1999; 87:138–145.

Corina DP, Knapp HP. Signed language and human action processing: evidence for functional constraints on the human mirror-neuron system. Ann N Y Acad Sci. 2008; 1145:100–112. [PubMed: 19076392]

Cunningham KA, Southall BL, Reichmuth C. Auditory sensitivity of seals and sea lions in complex listening scenarios. J Acoust Soc Am. 2014; 136:3410. [PubMed: 25480085]

Cusack R, Carlyon RP. Perceptual asymmetries in audition. J Exp Psychol Hum Percept Perform. 2003; 29:713–725. [PubMed: 12848335]

Damasio AR, Damasio H, Van Hoeson GW. Prosopagnosia: anatomical basis and behavioral mechanisms. Neurology. 1982; 32:331–341. [PubMed: 7199655]

Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio RD. A neural basis for lexical retrieval. Nature. 1996; 380:499–505. [PubMed: 8606767]

Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A. Neural systems behind word and concept retrieval. Cognition. 2004; 92:179–229. [PubMed: 15037130]

Darwin C. Descent. Man, Sel Relat Sex. 1871/1981

De Lucia M, Camen C, Clarke S, Murray MM. The role of actions in auditory object discrimination. Neuroimage. 2009; 48:475–485. [PubMed: 19559091]

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]

DeYoe EA, Felleman DJ, Van Essen DC, McClendon E. Multiple processing streams in occipitotemporal visual cortex. Nature. 1994; 371:151–154. [PubMed: 8072543]

Dick F, Saygin AP, Galati G, Pitzalis S, Bentrovato S, D'Amico S, Wilson S, Bates E, Pizzamiglio L. What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. J Cogn Neurosci. 2007; 19:799–816. [PubMed: 17488205]

Donald, M. Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press; 1991.

Dosher BA, Lu ZL. Hebbian Reweighting on Stable Representations in Perceptual Learning. Learn Percept. 2009; 1:37–58. [PubMed: 20305755]

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]

Engelien A, Huber W, Silbersweig D, Stern E, Frith CD, Doring W, Thron A, Frackowiak RS. The neural correlates of 'deaf-hearing' in man: conscious sensory awareness enabled by attentional modulation. Brain. 2000; 123(Pt 3):532–545. [PubMed: 10686176]

Epstein R, Harris A, Stanley D, Kanwisher N. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 1999; 23:115–125. [PubMed: 10402198]

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. The roles of infant crying and motherese during prelinguistic evolution in early hominins. Am J Phys Anthropol. 2004a:93–93.

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

Farago T, Andics A, Devecseri V, Kis A, Gacsi M, Miklosi A. Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. Biol Lett. 2014; 10:20130926. [PubMed: 24402716]

Farah MJ, Aguirre GK. Imaging visual recognition: pet and fMRI studies of the functional anatomy of human visual recognition. Trends Cogn Sci. 1999; 3:179–186. [PubMed: 10322474]

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex. 1991; 1:1–47. [PubMed: 1822724]

Fernandez-Abascal EG, Guerra P, Martinez F, Dominguez FJ, Munoz MA, Egea DA, Martin MD, Mata JL, Rodriguez S, Vila J. [The International Affective Digitized Sounds (IADS): spanish norms]. Psicothema. 2008; 20:104–113. [PubMed: 18206072]

Ferrand CT. Harmonics-to-noise ratio: an index of vocal aging. J Voice. 2002; 16:480–487. [PubMed: 12512635]

Fifer WP, Moon C. Psychobiology of newborn auditory preferences. Semin Perinatol. 1989; 13:430–433. [PubMed: 2814529]

Fodor, J. The mind doesn't work that way: the scope and limits of computational psychology. Cambridge, MA: "A Bradford book" MIT Press; 2001.

Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R. Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron. 2003; 40:859–869. [PubMed: 14622588]

Freud E, Plaut DC, Behrmann M. 'What' Is Happening in the Dorsal Visual Pathway. Trends Cogn Sci. 2016; 20:773–784. [PubMed: 27615805]

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

Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. J Neurophysiol. 2007a; 98:2337–2346. [PubMed: 17699691]

Fritz JB, Elhilali M, David SV, Shamma SA. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? Hear Res. 2007b; 229:186–203. [PubMed: 17329048]

Fritz JB, Malloy M, Mishkin M, Saunders RC. Monkeys short-term auditory memory nearly abolished by combined removal of the rostral superior temporal gyrus and rhinal cortices. Brain Res. 2016; 1640:289–298. [PubMed: 26707975]

Fu KMG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE. Auditory cortical neurons respond to somatosensory stimulation. J Neurosci. 2003; 23:7510–7515. [PubMed: 12930789]

Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, Pizzamiglio L. A selective representation of the meaning of actions in the auditory mirror system. Neuroimage. 2008; 40:1274–1286. [PubMed: 18276163]

Gandour J, Tong Y, Wong D, Talavage T, Dzemidzic M, Xu Y, Li X, Lowe M. Hemispheric roles in the perception of speech prosody. Neuroimage. 2004; 23:344–357. [PubMed: 15325382]

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat Neurosci. 1999; 2:568–573. [PubMed: 10448223]

Gazzola V, Aziz-Zadeh L, Keysers C. Empathy and the somatotopic auditory mirror system in humans. Curr Biol. 2006; 16:1824–1829. [PubMed: 16979560]

Geangu E, Quadrelli E, Lewis JW, Macchi Cassia V, Turati C. By the sound of it An ERP investigation of human action sound processing in 7-month-old infants. Dev Cogn Neurosci. 2015; 12:134–144. [PubMed: 25732377]

Geschwind N. Disconnexion syndromes in animals and man. I Brain. 1965; 88:237–294. [PubMed: 5318481]

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

Goodale MA, Milner AD. Separate visual pathways for perception and action. Trends Neurosci. 1992; 15:20–25. [PubMed: 1374953]

Grandin, T. The way I see it: a personal look at autism & Asperger's. Future Horizons Inc; Arlington, TX: 2008.

Griffiths TD, Bench CJ, Frackowiak RSJ. Human cortical areas selectively activated by apparent sound movement. Curr Biol. 1994; 4:892–895. [PubMed: 7850422]

Grigor'eva OM, Deriagina MA. Gestural forms of communication in primates. II. The relation of gestures to acoustic signals and tool activities in primate phylogeny. Nauchnye Doki Vyss Shkoly Biol Nauk. 1987:56–60.

Grill-Spector K. The neural basis of object perception. Curr Opin Neurobiol. 2003; 13:159–166. [PubMed: 12744968]

Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. Brain areas involved in perception of biological motion. J Cogn Neurosci. 2000; 12:711–720. [PubMed: 11054914]

Grossman ED, Blake R. Brain areas active during visual perception of biological motion. Neuron. 2002; 35:1167–1175. [PubMed: 12354405]

Grossman M, Koenig P, DeVita C, Glosser G, Alsop D, Detre J, Gee J. The neural basis for category-specific knowledge: an fMRI study. NeuroImage. 2002; 15:936–948. [PubMed: 11906234]

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]

Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE. Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. J Comp Neurol. 2007; 502:924–952. [PubMed: 17444488]

Han Z, Bi Y, Chen J, Chen Q, He Y, Caramazza A. Distinct regions of right temporal cortex are associated with biological and human-agent motion: functional magnetic resonance imaging and neuropsychological evidence. J Neurosci. 2013; 33:15442–15453. [PubMed: 24068813]

Harkins J, Tucker P. An internet survey of individuals with hearing loss regarding assistive listening devices. Trends Amplif. 2007; 11:91–100. [PubMed: 17494875]

Hashimoto T, Usui N, Taira M, Nose I, Haji T, Kojima S. The neural mechanism associated with the processing of onomatopoeic sounds. Neuroimage. 2006; 31:1762–1770. [PubMed: 16616863]

Heilman KM, Scholes R, Watson RT. Auditory affective agnosia. Disturbed comprehension of affective speech. J Neurol Neurosurg Psychiatry. 1975; 38:69–72. [PubMed: 1117301]

Hewes GW. Primate communication and the gestural origin of language. Curr Anthropol. 1973; 14:5–24.

Hillis AE, Caramazza A. Category-specific naming and comprehension impairment: a double dissociation. Brain. 1991; 114(Pt 5):2081–2094. [PubMed: 1933235]

Hopfield JJ, Tank DW. “Neural” computation of decisions in optimization problems. Biol Cybern. 1985; 52:141–152. [PubMed: 4027280]

Imai M, Kita S. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130298. [PubMed: 25092666]

Imai M, Kita S, Nagumo M, Okada H. Sound symbolism facilitates early verb learning. Cognition. 2008; 109:54–65. [PubMed: 18835600]

Imai M, Miyazaki M, Yeung HH, Hidaka S, Kantartzis K, Okada H, Kita S. Sound symbolism facilitates word learning in 14-month-olds. PLoS One. 2015; 10:e0116494. [PubMed: 25695741]

Iriki A, Tanaka M, Iwamura Y. Coding of modified body schema during tool use by macaque postcentral neurones. Neuroreport. 1996; 7:2325–2330. [PubMed: 8951846]

Jellema T, Perrett DI. Neural representations of perceived bodily actions using a categorical frame of reference. Neuropsychologia. 2006; 44:1535–1546. [PubMed: 16530792]

Johannesson A. The gestural origin of language; evidence from six ‘unrelated’ languages. Nature. 1950; 166:60–61. [PubMed: 15439119]

Johnson-Frey SH. What’s so special about human tool use? Neuron. 2003; 39:201–204. [PubMed: 12873378]

Kaas JH, Hackett TA. Subdivisions of auditory cortex and levels of processing in primates. Audiol Neurootol. 1998; 3:73–85. [PubMed: 9575378]

Kaas JH, Hackett TA. ‘What’ and ‘where’ processing in auditory cortex. Nat Neurosci. 1999; 2:1045–1047. [PubMed: 10570476]

Kaas JH, Hackett TA. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci USA. 2000a; 97:11793–11799. [PubMed: 11050211]

Kaas JH, Hackett TA. How the visual projection map instructs the auditory computational map. J Comp Neurol. 2000b; 421:143–145. [PubMed: 10813777]

Kaas JH, Collins CE. The organization of sensory cortex. Curr Opin Neurobiol. 2001; 11:498–504. [PubMed: 11502398]

Kandel, E., Schwartz, J., Jessel, T. Principles of Neural Science. McGraw-Hill; New York: 2000.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 1997; 17:4302–4311. [PubMed: 9151747]

Karnath HO, Ruter J, Mandler A, Himmelbach M. The anatomy of object recognition-visual form agnosia caused by medial occipitotemporal stroke. J Neurosci. 2009; 29:5854–5862. [PubMed: 19420252]

Kastner S, De Weerd P, Ungerleider LG. Texture segregation in the human visual cortex: a functional MRI study. J Neurophysiol. 2000; 83:2453–2457. [PubMed: 10758146]

Kellenbach ML, Brett M, Patterson K. Actions speak louder than functions: the importance of manipulability and action in tool representation. J Cogn Neurosci. 2003; 15:30–46. [PubMed: 12590841]

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]

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]

Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD. Music, language and meaning: brain signatures of semantic processing. Nat Neurosci. 2004; 7:302–307. [PubMed: 14983184]

Kohler E, Keysers C, Umilta A, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. Science. 2002; 297:846–848. [PubMed: 12161656]

Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD. On the lateralization of emotional prosody: an event-related functional MR investigation. Brain Lang. 2003; 86:366–376. [PubMed: 12972367]

Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. Science. 1982; 218:1138–1141. [PubMed: 7146899]

Lahav A, Saltzman E, Schlaug G. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci. 2007; 27:308–314. [PubMed: 17215391]

Larsson M. Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. Anim Cogn. 2014; 17:1–14. [PubMed: 23990063]

Larsson M. Tool-use-associated sound in the evolution of language. Anim Cogn. 2015; 18:993–1005. [PubMed: 26118672]

Lausberg H, Kita S, Zaidel E, Ptito A. Split-brain patients neglect left personal space during right-handed gestures. Neuropsychologia. 2003; 41:1317–1329. [PubMed: 12757905]

Lee CT, Brown CA, Hains SM, Kisilevsky BS. Fetal development: voice processing in normotensive and hypertensive pregnancies. Biol Res Nurs. 2007; 8:272–282. [PubMed: 17456588]

Levy DA, Granot R, Bentin S. Processing specificity for human voice stimuli: electrophysiological evidence. Neuroreport. 2001; 28:2653–2657.

Lewis, JW. Audio-visual perception of everyday natural objects—hemodynamic studies in humans. In: Naumer, MJ., Kaiser, J., editors. Multisensory object perception in the primate brain. Oxford University Press: Springer; 2010. p. 155-190.

Lewis JW, Beauchamp MS, DeYoe EA. A comparison of visual and auditory motion processing in human cerebral cortex. Cereb Cortex. 2000; 10:873–888. [PubMed: 10982748]

Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA. Lefties get it “right” when hearing tool sounds. J Cogn Neurosci. 2006; 18:1314–1330. [PubMed: 16859417]

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, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. Distinct cortical pathways for processing tool versus animal sounds. J Neurosci. 2005; 25:5148–5158. [PubMed: 15917455]

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. J Cogn Neurosci. 2011a; 23:2079–2101. [PubMed: 20812786]

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. Human brain regions involved in recognizing environmental sounds. Cereb Cortex. 2004; 14:1008–1021. [PubMed: 15166097]

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. J Neurosci. 2009; 29:2283–2296. [PubMed: 19228981]

Lewis JW, Frum C, Brefczynski-Lewis JA, Talkington WJ, Walker NA, Rapuano KM, Kovach AL. Cortical network differences in the sighted versus early blind for recognition of human-produced action sounds. Hum Brain Mapp. 2011b; 32:2241–2255. [PubMed: 21305666]

Li E, Weng X, Han Y, Wu S, Zhuang J, Chen C, Feng L, Zhang K. Asymmetry of brain functional activation: fMRI study under language and music stimulation. Chin Med J (Engl). 2000; 113:154–158. [PubMed: 11775542]

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

Liepmann H. Agnosic disorders (1908) [classical article]. Cortex. 2001; 37:547–553. [PubMed: 11721866]

Lingle S, Riede T. Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. Am Nat. 2014; 184:510–522. [PubMed: 25226186]

Lingle S, Wyman MT, Kotrba R, Teichroeb LJ, Romanow CA. What makes a cry a cry? A review of infant distress vocalizations. Curr Zool. 2012; 58:698–726.

Linke AC, Jao Keehn RJ, Pueschel EB, Fishman I, Muller RA. Children with ASD show links between aberrant sound processing, social symptoms, and atypical auditory interhemispheric and thalamocortical functional connectivity. Dev Cogn Neurosci. 2017

Lortie M, Proulx-Begin L, Saint-Amour D, Cousineau D, Theoret H, Lepage JF. Brief report: biological sound processing in children with autistic spectrum disorder. J Autism Dev Disord. 2017

Lowery KJ, Plyler PN. The effects of noise reduction technologies on the acceptance of background noise. J Am Acad Audiol. 2013; 24:649–659. [PubMed: 24131601]

MacNeilage PF. The frame/content theory of evolution of speech production. Behav Brain Sci. 1998; 21:499–511. (discussion 511–446). [PubMed: 10097020]

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]

Mäkelä JP, McEvoy L. Auditory evoked fields to illusory sound source movements. Exp Brain Res. 1996; 110:446–453. [PubMed: 8871103]

Martin A. The representation of object concepts in the brain. Annu Rev Psychol. 2007; 58:25–45. [PubMed: 16968210]

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. Neural correlates of category-specific knowledge. Nature. 1996; 379(6566):649–652. [PubMed: 8628399]

Mastropieri D, Turkewitz G. Prenatal experience and neonatal responsiveness to vocal expressions of emotion. Dev Psychobiol. 1999; 35:204–214. [PubMed: 10531533]

McDermott JH, Oxenham AJ. Spectral completion of partially masked sounds. Proc Natl Acad Sci USA. 2008; 105:5939–5944. [PubMed: 18391210]

McDermott JH, Simoncelli EP. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron. 2011; 71:926–940. [PubMed: 21903084]

McNamara A, Buccino G, Menz MM, Glascher J, Wolbers T, Baumgartner A, Binkofski F. Neural dynamics of learning sound-action associations. PLoS One. 2008; 3:e3845. [PubMed: 19050764]

Meguerditchian A, Vauclair J, Hopkins WD. On the origins of human handedness and language: a comparative review of hand preferences for bimanual coordinated actions and gestural communication in nonhuman primates. Dev Psychobiol. 2013; 55:637–650. [PubMed: 23955015]

Mipham, JR. Gateway to knowledge. Hong Kong: Boudhanath & Esby: Rangjunk Yeshe Publications; 2000.

Molenberghs P, Cunnington R, Mattingley JB. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. Neurosci Biobehav Rev. 2012; 36:341–349. [PubMed: 21782846]

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

Moore CJ, Price CJ. A functional neuroimaging study of the variables that generate category-specific object processing differences. Brain. 1999; 122:943–962. [PubMed: 10355678]

Morillon B, Lehongre K, Frackowiak RS, Ducorps A, Kleinschmidt A, Poeppel D, Giraud AL. Neurophysiological origin of human brain asymmetry for speech and language. Proc Natl Acad Sci USA. 2010; 107:18688–18693. [PubMed: 20956297]

Munoz-Lopez M, Insausti R, Mohedano-Moriano A, Mishkin M, Saunders RC. Anatomical pathways for auditory memory II: information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. Front Neurosci. 2015; 9:158. [PubMed: 26041980]

Murray MM, Lewkowicz DJ, Amedi A, Wallace MT. Multisensory Processes: a Balancing Act across the Lifespan. Trends Neurosci. 2016; 39:567–579. [PubMed: 27282408]

Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. Rapid brain discrimination of sounds of objects. J Neurosci. 2006; 26:1293–1302. [PubMed: 16436617]

Murray SO, Newman AJ, Roder B, Mitchell TV, Takahashi T, Neville HJ. Functional organization of auditory motion processing in humans using fMRI. Soc Neurosci Abstr. 1998; 24:1401.

Disorders IoDaOC. , editor. NIDCD. Quick statistics retrieved September 11. 2009.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp. 2006; 27:562–571. [PubMed: 16281283]

Ortiz-Rios M, Kusmierek P, DeWitt I, Archakov D, Azevedo FA, Sams M, Jaaskelainen IP, Keliris GA, Rauschecker JP. Functional MRI of the vocalization-processing network in the macaque brain. Front Neurosci. 2015; 9:113. [PubMed: 25883546]

Ozturk O, Krehm M, Vouloumanos A. Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. J Exp Child Psychol. 2013; 114:173–186. [PubMed: 22960203]

Paillard, J. The hand and the tool: the functional architecture of human technical skills. In: Berthelet, A., Chavaillon, J., editors. The use of tools by human and non-human primates. Clarendon Press; Oxford: 1993. p. 36-50.

Palmeri TJ, Gauthier I. Visual object understanding. Nat Rev Neurosci. 2004; 5:291–303. [PubMed: 15034554]

Parks TE. Illusory figures: a (mostly) atheoretical review. Psychol Bull. 1984; 95:282–300. [PubMed: 6544435]

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

Pelphrey KA, Morris JP, McCarthy G. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J Cogn Neurosci. 2004; 16:1706–1716. [PubMed: 15701223]

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. Proc Natl Acad Sci USA. 2003; 100:11702–11705. [PubMed: 14500906]

Perlovsky L. Language and cognition interaction neural mechanisms. Comput Intell Neurosci. 2011; 2011:454587. [PubMed: 21876687]

Perniss P, Vigliocco G. The bridge of iconicity: from a world of experience to the experience of language. Philos Trans R Soc Lond B Biol Sci. 2014; 369:20130300. [PubMed: 25092668]

Petersen SE, Van Mier H, Fiez JA, Raichle ME. The effects of practice on the functional anatomy of task performance. Proceedings of the National Academy of Sciences USA. 1998; 95:853–860.

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]

Phillips DP. Auditory gap detection, perceptual channels, and temporal resolution in speech perception. J Am Acad Audiol. 1999; 10:343–354. [PubMed: 10385876]

Pietrini P, Furey ML, Ricciardi E, Gobbini MI, Wu WH, Cohen L, Guazzelli M, Haxby JV. Beyond sensory images: object-based representation in the human ventral pathway. Proc Natl Acad Sci USA. 2004; 101:5658–5663. [PubMed: 15064396]

Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D’Amico S, Di Russo F. Separate neural systems for processing action- or non-action-related sounds. Neuroimage. 2005; 24:852–861. [PubMed: 15652320]

Poremba A, Mishkin M. Exploring the extent and function of higher-order auditory cortex in rhesus monkeys. Hear Res. 2007; 229:14–23. [PubMed: 17321703]

Preuss TM. The human brain: rewired and running hot. Ann N Y Acad Sci. 2011; 1225(Suppl 1):E182–E191. [PubMed: 21599696]

Pylyshyn, ZW. Computation and Cognition: Toward a Foundation for Cognitive Science. MIT Press Cambridge: MA; 1984.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage. 2001; 13:669–683. [PubMed: 11305896]

Raglio A, Galandra C, Sibilla L, Esposito F, Gaeta F, Di Salle F, Moro L, Carne I, Bastianello S, Baldi M, Imbriani M. Effects of active music therapy on the normal brain: fmri based evidence. Brain Imaging Behav. 2016; 10:182–186. [PubMed: 25847861]

Rauschecker JP. Parallel processing in the auditory cortex of primates. Audiol Neurootol. 1998; 3:86–103. [PubMed: 9575379]

Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. Proceedings of the National Academy of Sciences USA. 2000; 97:11800–11806.

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]

Rauschecker JP, Tian B, Hauser M. Processing of Complex Sounds in the Macaque Nonprimary Auditory Cortex. Science. 1995; 268:111–114. [PubMed: 7701330]

Rauschecker JP, Tian B, Pons T, Mishkin M. Serial and parallel processing in rhesus monkey auditory cortex. J Comp Neurol. 1997; 382:89–103. [PubMed: 9136813]

Recanzone GH, Cohen YE. Serial and parallel processing in the primate auditory cortex revisited. Behav Brain Res. 2009; 206:1–7. [PubMed: 19686779]

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

Reichmuth C, Casey C. Vocal learning in seals, sea lions, and walruses. Curr Opin Neurobiol. 2014; 28:66–71. [PubMed: 25042930]

Reppas JB, Niyogi S, Dale AM, Sereno MI, Tootell RB. Representation of motion boundaries in retinotopic human visual cortical areas. Nature. 1997; 388:175–179. [PubMed: 9217157]

Rhoades, E. Sound-object associations. In: Easterbrooks, S., Estes, E., editors. Helping children who are deaf and hard of hearing learn spoken language. Thousand Oaks, CA: Corwin Press; 2007. p. 181-188.sound-object associations

Ricciardi E, Bonino D, Sani L, Vecchi T, Guazzelli M, Haxby JV, Fadiga L, Pietrini P. Do we really need vision? How blind people "see" the actions of others. J Neurosci. 2009; 29:9719–9724. [PubMed: 19657025]

Riede T, Herzel H, Hammerschmidt K, Brunnberg L, Tembrock G. The harmonic-to-noise ratio applied to dog barks. J Acoust Soc Am. 2001; 110:2191–2197. [PubMed: 11681395]

Ries ML, Jabbar BM, Schmitz TW, Trivedi MA, Gleason CE, Carlsson CM, Rowley HA, Asthana S, Johnson SC. Anosognosia in mild cognitive impairment: relationship to activation of cortical midline structures involved in selfappraisal. J Int Neuropsychol Soc. 2007; 13:450–461. [PubMed: 17445294]

Rizzolatti G, Craighero L. The mirror-neuron system. Annu Rev Neurosci. 2004; 27:169–192. [PubMed: 15217330]

Rizzolatti, G., Craighero, L. Language and mirror neurons. Gaskell, MG., editor. The Oxford Handbook of Psycholinguistics; Oxford: 2007.

Ross ED, Monnot M. Neurology of affective prosody and its functional-anatomic organization in right hemisphere. Brain Lang. 2008; 104:51–74. [PubMed: 17537499]

Rubin N. Figure and ground in the brain. Nat Neurosci. 2001; 4:857–858. [PubMed: 11528408]

Rumberg, B., Graham, D. Proceedings of the International Symposium on Quality Electronic Design. Santa Clara, CA: 2015. A low-power field-programmable analog array for wireless sensing; p. 542-546.

Salmi J, Koistinen OP, Glerean E, Jylanki P, Vehtari A, Jaaskelainen IP, Makela S, Nummenmaa L, Nummi-Kuisma K, Nummi I, Sams M. Distributed neural signatures of natural audiovisual speech and music in the human auditory cortex. Neuroimage. 2016

Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS Comput Biol. 2014; 10:e1003412. [PubMed: 24391486]

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]

Saygin AP, Dick F, Wilson SW, Dronkers NF, Bates E. Neural resources for processing language and environmental sounds. Brain. 2003; 126:928–945. [PubMed: 12615649]

Schechter DS, Moser DA, Wang Z, Marsh R, Hao X, Duan Y, Yu S, Gunter B, Murphy D, McCaw J, Kangarlu A, Willheim E, Myers MM, Hofer MA, Peterson BS. An fMRI study of the brain responses of traumatized mothers to viewing their toddlers during separation and play. Soc Cogn Affect Neurosci. 2012; 7:969–979. [PubMed: 22021653]

Schmitt V, Kroger I, Zinner D, Call J, Fischer J. Monkeys perform as well as apes and humans in a size discrimination task. Anim Cogn. 2013; 16:829–838. [PubMed: 23443407]

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC. Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol. 2001; 85:1322–1327. [PubMed: 11248001]

Sergent J, Ohta S, MacDonald B. Functional neuroanatomy of face and object processing. Brain. 1992; 115:15–36. [PubMed: 1559150]

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

Sheridan M. Delay or failure in the development of spoken language. Proc R Soc Med. 1959; 52:913–916. [PubMed: 14445870]

Silveri MC, Gainotti G, Perani D, Cappelletti JY, Carbone G, Fazio F. Naming deficit for non-living items: neuropsychological and PET study. Neuropsychologia. 1997; 35:359–367. [PubMed: 9051684]

Smith L, Fraser D. Robust sound onset detection using leaky integrate and fire neurons with depressing synapses. IEEE Trans Neural Netw. 2004; 15:1125–1134. [PubMed: 15484889]

Stothard SE, Snowling MJ, Bishop DV, Chipchase BB, Kaplan CA. Language-impaired preschoolers: a follow-up into adolescence. J Speech Lang Hear Res. 1998; 41:407–418. [PubMed: 9570592]

Stout D, Toth N, Schick K, Chaminade T. Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. Philos Trans R Soc Lond B Biol Sci. 2008; 363:1939–1949. [PubMed: 18292067]

Suo X, Lei D, Li K, Chen F, Li F, Li L, Huang X, Lui S, Li L, Kemp GJ, Gong Q. Disrupted brain network topology in pediatric posttraumatic stress disorder: a resting-state fMRI study. Hum Brain Mapp. 2015; 36:3677–3686. [PubMed: 26096541]

Sweet RA, Dorph-Petersen KA, Lewis DA. Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J Comp Neurol. 2005; 491:270–289. [PubMed: 16134138]

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

Takahashi G, Martinez CD, Beamer S, Bridges J, Noffsinger D, Sugiura K, Bratt GW, Williams DW. Subjective measures of hearing aid benefit and satisfaction in the NIDCD/VA follow-up study. J Am Acad Audiol. 2007; 18:323–349. [PubMed: 17580727]

Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J Neurophysiol. 2004; 91:1282–1296. [PubMed: 14614108]

Talkington WJ, Rapuano KM, Hitt LA, Frum CA, Lewis JW. Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. J Neurosci. 2012; 32:8084–8093. [PubMed: 22674283]

Talkington WJ, Taglialatela JP, Lewis JW. Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. Hear Res. 2013; 305:74–85. [PubMed: 23994296]

Taylor KI, Stamatakis EA, Tyler LK. Crossmodal integration of object features: voxel-based correlations in brain-damaged patients. Brain. 2009; 132:671–683. [PubMed: 19190042]

Taylor KI, Moss HE, Stamatakis EA, Tyler LK. Binding crossmodal object features in perirhinal cortex. Proc Natl Acad Sci USA. 2006; 103:8239–8244. [PubMed: 16702554]

Tecchio F, Benassi F, Zappasodi F, Gialloreti LE, Palermo M, Seri S, Rossini PM. Auditory sensory processing in autism: a magnetoencephalographic study. Biol Psychiatry. 2003; 54:647–654. [PubMed: 13129660]

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD. Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci. 2011; 31:164–171. [PubMed: 21209201]

Thompson JC, Clarke M, Stewart T, Puce A. Configural processing of biological motion in human superior temporal sulcus. J Neurosci. 2005; 25:9059–9066. [PubMed: 16192397]

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature. 1996; 381:520–522. [PubMed: 8632824]

Tootell RBH, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci. 1995; 15:3215–3230. [PubMed: 7722658]

Tranel D, Damasio H, Damasio AR. A neural basis for the retrieval of conceptual knowledge. Neuropsychologia. 1997; 35:1319–1327. [PubMed: 9347478]

Trumpp NM, Kliese D, Hoenig K, Haarmeier T, Kiefer M. Losing the sound of concepts: damage to auditory association cortex impairs the processing of sound-related concepts. Cortex. 2013; 49:474–486. [PubMed: 22405961]

Ungerleider, LG., Mishkin, M., Goodale, MA., Mansfield, RJW. Two cortical visual systems. In: Ingle, DJ., editor. Analysis of Visual Behavior. MIT Press; Cambridge, MA: 1982. p. 549-586.

Van Essen DC, Maunsell JH, Bixby JL. The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. J Comp Neurol. 1981; 199:293–326. [PubMed: 7263951]

Varela, FJ., Thompson, E., Rosch, E. The embodied mind: cognitive science and human experience. Cambridge, Mass: MIT Press; 1991.

Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron. 2001; 31:329–338. [PubMed: 11502262]

Wang K, Shamma SA. Modeling the auditory functions in the primary cortex. Proceedings of the SPIE. 1994:692–703.

Wang WJ, Wu XH, Li L. The dual-pathway model of auditory signal processing. Neurosci Bull. 2008; 24:173–182. [PubMed: 18500391]

Warren J, Zielinski B, Green G, Rauschecker J, Griffiths T. Perception of sound-source motion by the human brain. Neuron. 2002; 34:139–148. [PubMed: 11931748]

Warrington EK, Shallice T. Category specific semantic impairments. Brain. 1984; 107(Pt 3):829–854. [PubMed: 6206910]

Webb AR, Heller HT, Benson CB, Lahav A. Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. Proc Natl Acad Sci USA. 2015; 112:3152–3157. [PubMed: 25713382]

Webster PJ, Skipper-Kallal LM, Frum CA, Still HN, Ward BD, Lewis JW. Divergent human cortical regions for processing distinct acoustic-semantic categories of natural sounds: animal action sounds vs. vocalizations. Front Neurosci. 2017; 10:579. [PubMed: 28111538]

Wessinger CM, Buonocore MH, Kussmaul CL, Mangun GR. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. Human Brain Mapp. 1997; 5:18–25.

Wiggs CL, Martin A. Properties and mechanisms of perceptual priming. Curr Opin Neurobiol. 1998; 8:227–233. [PubMed: 9635206]

Wilden I, Herzel H, Peters G, Tembrock G. Subharmonics, biphonation, and deterministic chaos in mammal vocalization. Int J Anim Sound its Rec. 1998; 9:171–196.

Yantis S, Jonides J. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. J Exp Psychol: Human Percept Perform. 1990; 16:121–134. [PubMed: 2137514]

Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. Separating the brain regions involved in recollection and familiarity in recognition memory. J Neurosci. 2005; 25:3002–3008. [PubMed: 15772360]

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 Cogn Sci. 2002; 6:37–46. [PubMed: 11849614]

Zatorre RJ, Chen JL, Penhune VB. When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci. 2007; 8:547–558. [PubMed: 17585307]

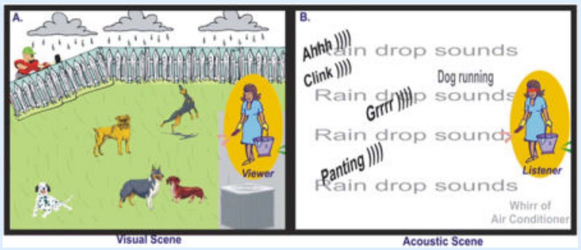
Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. Science. 1992; 256:846–849. [PubMed: 1589767]

Zilbovicius M, Meresse I, Chabane N, Brunelle F, Samson Y, Boddaert N. Autism, the superior temporal sulcus and social perception. Trends Neurosci. 2006; 29:359–366. [PubMed: 16806505]

Box 1

Comparison of real-world visual and auditory objects

In this box, we see a visual and auditory landscape, as experienced by the woman in blue as a viewer (A), and a listener (B). Note differences in the richness of spatial detail, number and types of objects perceived in the visual vs. auditory world. The visual world is more spatially precise and detailed, such that details of the grass, clouds and dogs' coats can be ascertained in high resolution due to retina properties. In contrast, only objects in motion make vibrations that activate the cochlea, but these sounds can travel even through a visual occlusion like a fence, and if loud enough can be detected even if looking away or asleep. Objects like the silent dogs are invisible to the blindfolded listener, while the hammering neighbor behind the fence would be undetected by a deaf viewer. All categories proposed in the model (Fig. 1) are present. This scenario will parallel discussion of the model in the text in Boxes 2 and 3.



Box 2

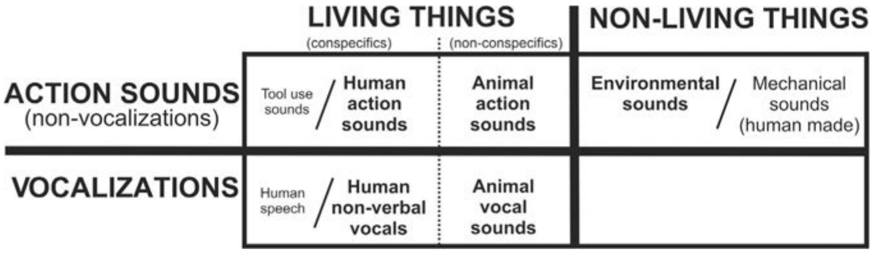
Auditory objects versus auditory scenes

Returning to the example scenario from Box 1, bottom up cues may direct processing of the perceived stimuli. For example, the ‘Ahh’ utterance depicted in Box 1 has high **harmonic content** that would engage signal processing along lateral aspect of medial temporal cortices, as depicted by the rainbow arrows in Fig. 2. Being a human utterance, the sound would also share other particular low level attributes that may further direct its processing to speech related regions. The **acoustic signal changing loudness over time** as the hammer clinks and the dog runs towards her is another specific spectro-temporal attribute characteristic of action sounds. Soundscapes characterized by the relatively constant drone (**constancy**) of the rain and air conditioner, along with relatively flatter **1/f^{α} spectral power**, would more likely be relegated as background status and thus less likely to require attentional processing priority.

Box 3

Top-down influences on the hearing model

To examine the influence of top down processing, using the picture example in Box 1, one can infer why the sounds from living things may be more relevant, why they have meaning, and how they are embodied to help process these meanings. When the listener embodies the human sounds, she will realize that when she has made such an utterance and loud clap it was usually to grab someone's attention. Such sounds would typically be accompanied by a physical state of arousal, apprehension or anticipation. Thus, she could match the activation pattern to earlier learned states and applying those to her friend (theory of mind). The growl would likely also warrant preferred processing, being a signal with relatively low harmonic content that is often associated with threat or negative valence. Humans make similar low harmonic utterances, which often indicate anger and aggression, and our listener in Box 1 likely has memories and other experiences with dogs and other animals making growls, perhaps associated with fear or even a painful bite. The acoustic signal changes evolving over time as the dog runs towards her direction is also relevant here (e.g. changes in loudness and other specific spectro-temporal attributes). Motor regions in the brain may be activated that embody running, and the increase in loudness over time may help recall earlier multisensory experiences when animals may have been running towards one. What the above sound-source examples have in common is intent, the kind of intent and meaning that comes from animate objects and makes them an important stimulus to process when planning action. Finally, although the sound of the rain may be relevant for whether she wants to go inside for shelter, it is inanimate and temporally homogeneous in quality, and will not respond to her actions - this requires a different kind of processing. The air conditioner whirr and rain are difficult to embody into a motor schema, and thus associative learning would rely more heavily on other processing mechanisms in the brain. However, rain, if taken as a deliberate object, may conjure bodily sensations associated with memories of rain (e.g. wet, cold). Both the rain and whirr sounds are more likely to be processed as background, until the situation involving the 'figure' animate object sounds is properly assessed.



	LIVING THINGS		NON-LIVING THINGS
	(conspecifics)	(non-conspecifics)	
ACTION SOUNDS (non-vocalizations)	Tool use sounds / Human action sounds	Animal action sounds	Environmental sounds / Mechanical sounds (human made)
VOCALIZATIONS	Human speech / Human non-verbal vocals	Animal vocal sounds	

Fig. 1. A neurobiological model of hearing perception for different categories of real-world, natural sounds

This model was refined largely from recent human neuroimaging studies, but should apply to most mammalian species with hearing and sound production capabilities.

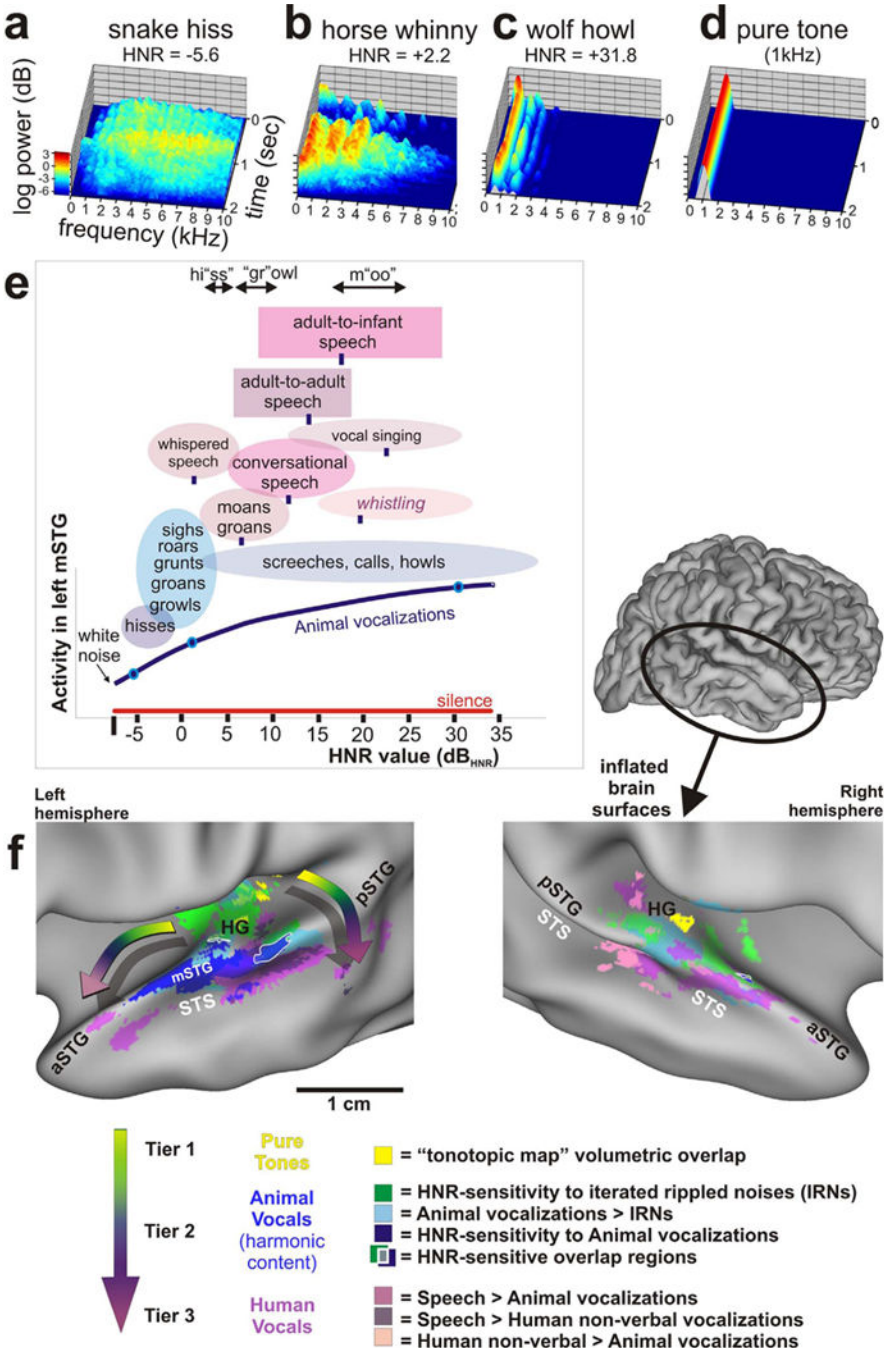


Fig. 2.

A bottom-up model for vocalization signal processing. (a–d) Three-dimensional spectrograms of example sound stimuli and a pure tone. Note the prominent stacks of energy along frequency bands of the vocalizations (b–c). (e) Chart illustrating harmonic content ranges (in dB_{HNR}) of various types or classes of communicative vocalizations. The y-axis depicts actual or relative degrees of fMRI activation in the left middle superior temporal gyrus (mSTG) region, with the blue curve depicting a response profile to animal vocalizations. Blue dots on curve correspond to sounds depicted in spectrograms. Ovals and boxes hovering above the curve depict dB_{HNR} values of different categories of animal vocalizations (blue hues) and human vocalizations (violet hues). (f) Progression of cortical pathways for processing harmonic content and information content of vocalizations. The "rainbow arrows" depict two prominent processing pathways showing increasing specificity for human vocalizations. Intermediate colors depict regions of overlap (refer to key for color codes). Data are illustrated on slightly inflated renderings of averaged cortical surface models (all at $p_{corrected} < 0.01$). Adapted and reprinted from Lewis et al. (2009) with permission from the publisher. Refer to text for other details.

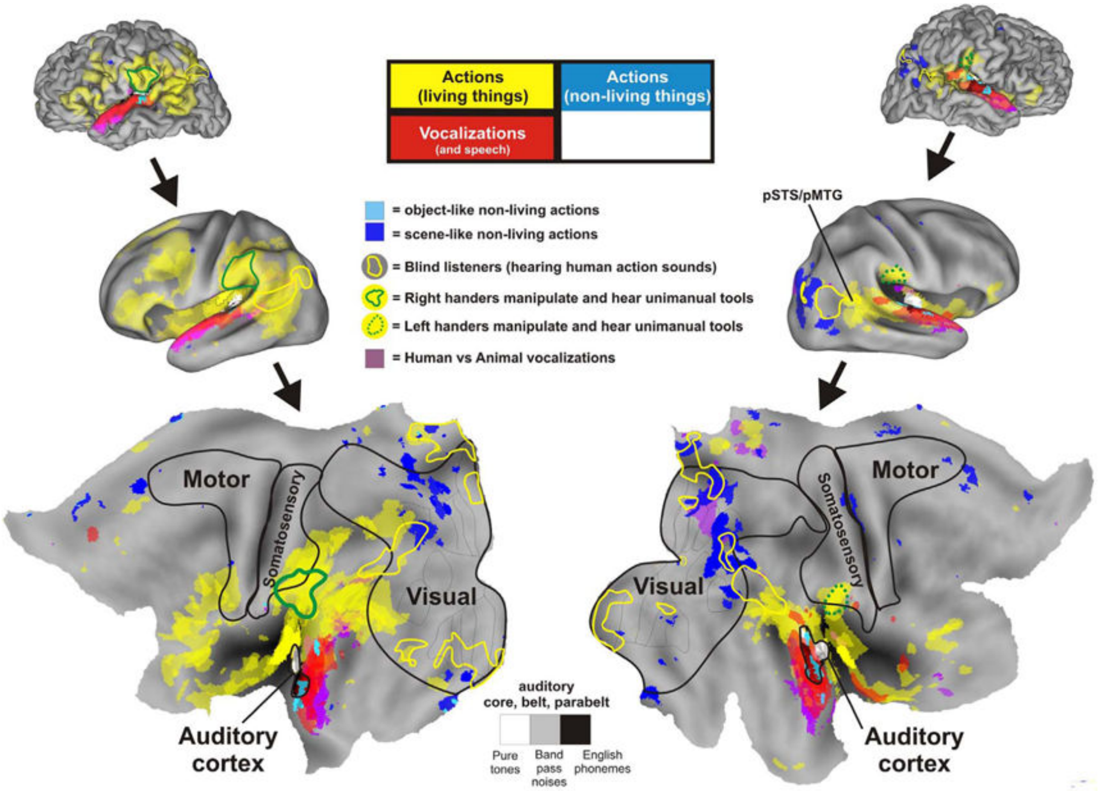

Actions (living things)
Actions (non-living things)
Vocalizations (and speech)
= object-like non-living actions
= scene-like non-living actions
= Blind listeners (hearing human action sounds)
= Right handers manipulate and hear unimanual tools
= Left handers manipulate and hear unimanual tools
= Human vs Animal vocalizations
pSTS/pMTG
Motor
Somatosensory
Visual
Auditory cortex
auditory core, belt, parabelt
Pure tones
Band pass noises
English phonemes

Fig. 3. Meta-analysis of brain regions preferential for each of the three major categories of natural sound in the model

The data from select published studies were adapted to be color coded and overlaid in transparent layers, revealing the major cortical regions and networks associated with natural sound processing. Red colored cortices correspond to vocalizations, yellow to biological action sounds, and blue to non-living environmental and mechanical action sounds. Studies include selected results restricted to the defined sound category boundaries from Lewis et al., (2004, 2005, 2006, 2009); Engel et al. (2009); Lewis et al. (2011a), (2011b), (2012); and Webster et al. (2017)). Refer to color key and text for other details.

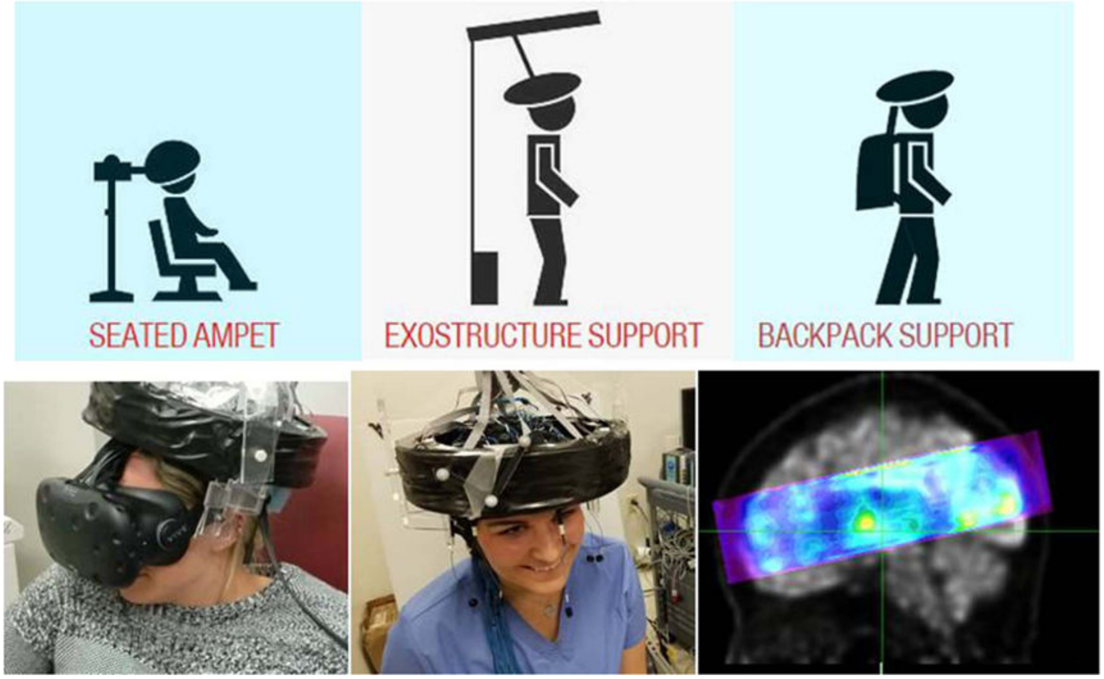


Fig. 4. Wearable PET technology that could be used for interactive auditory perception studies

The top panel shows the potential designs from seated to standing and walking, that could be utilized for behavioral studies of perception and action, and would allow for gestures, vocalizations and avoid/approach behaviors. Lower left shows our proof-of-concept device with virtual reality goggles which could be a mechanism for studying auditory and multimodal perception in an interactive environment (30 s - several minute temporal resolution) that could be combined with EEG to obtain higher temporal resolution (cite abstracts?). Lower right shows human patient data (one participant actively turning head from side to side) from our limited brain coverage prototype, but still demonstrating medial brain structures such as basal ganglia and thalamus.