2013

Early and Late Stage Mechanisms for Vocalization Processing in the Human Auditory System

William James Talkington
West Virginia University

Early and Late Stage Mechanisms for Vocalization Processing in the Human Auditory System

William James Talkington

Dissertation submitted to the School of Medicine at West Virginia University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Neuroscience

James W. Lewis, Ph.D., Chair
Albert S. Berrebi, Ph.D.
David W. Graham, Ph.D.
Marc W. Haut, Ph.D., A.B.P.P
George A. Spirou, Ph.D.

Department of Neurobiology and Anatomy

Morgantown, West Virginia
2013

Keywords: fMRI, EEG, human auditory cortex, conspecific vocalizations

ABSTRACT

Early and Late Stage Mechanisms for Vocalization Processing in the Human Auditory System

William James Talkington

The human auditory system is able to rapidly process incoming acoustic information, actively filtering, categorizing, or suppressing different elements of the incoming acoustic stream. Vocalizations produced by other humans (conspecifics) likely represent the most ethologically-relevant sounds encountered by hearing individuals. Subtle acoustic characteristics of these vocalizations aid in determining the identity, emotional state, health, intent, etc. of the producer. The ability to assess vocalizations is likely subserved by a specialized network of structures and functional connections that are optimized for this stimulus class. Early elements of this network would show sensitivity to the most basic acoustic features of these sounds; later elements may show categorically-selective response patterns that represent high-level semantic organization of different classes of vocalizations. A combination of functional magnetic resonance imaging and electrophysiological studies were performed to investigate and describe some of the earlier and later stage mechanisms of conspecific vocalization processing in human auditory cortices. Using fMRI, cortical representations of harmonic signal content were found along the middle superior temporal gyri between primary auditory cortices along Heschl's gyri and the superior temporal sulci, higher-order auditory regions. Additionally, electrophysiological findings also demonstrated a parametric response profile to harmonic signal content. Utilizing a novel class of vocalizations, human-mimicked versions of animal vocalizations, we demonstrated the presence of a left-lateralized cortical vocalization processing hierarchy to conspecific vocalizations, contrary to previous findings describing similar bilateral networks. This hierarchy originated near primary auditory cortices and was further supported by auditory evoked potential data that suggests differential temporal processing dynamics of conspecific human vocalizations versus those produced by other species. Taken together, these results suggest that there are auditory cortical networks that are highly optimized for processing utterances produced by the human vocal tract. Understanding the function and structure of these networks will be critical for advancing the development of novel communicative therapies and the design of future assistive hearing devices.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mentor, Dr. James W. Lewis for endless enthusiasm and passion as a scientist and as a mentor. I will ever be thankful for our time together, not only during my tenure in his lab, but also in the future as life-long friends. I am also grateful to our esteemed lab assistant Chris Frum, for being a great colleague, friend, fellow jokester, discusser of finance, and fishing partner.

I would also like to thank my committee members Drs. Marc Haut, Albert Berrebi, David Graham, and George Spirou for their time, input, and encouragement during my time at WVU.

Thank you to my wife, Brandi Talkington, who has seen me through each step of my graduate career more closely than anyone, first as fellow graduate students and until now and forever as my partner in life. Thank you for your enduring patience and support.

Thank you to all of the present and past individuals involved with the Center for Advanced Imaging at WVU for your input and assistance on everything MRI.

Without participants, I would not have data. Thanks to the many (often repeat) volunteers that made my science possible.

Finally, I would like to thank my family. To my parents, William and Gudrun, I cherish all of the life lessons (past and future), shared laughs, friendship, and most of all their support early in my life in the forms of fishing and hunting trips with Dad and countless long trips to our local public library with Mom. To my sister, my oldest friend, I hope that I can give as much love to you as you have given me. To my grandparents, James and Dorothea, thank you for all of the easy times in Buckhannon spent cheering, cherishing and discussing our Mountaineers and talk about "how it used to be". To my newest family members, parents William and Linda, siblings Rick, Jen, and Kristen, niece Vanessa, and nephews Tyler, Preston, Aaron, and Ethan, I look forward to many more years of laughter and fun.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABR	auditory brainstem response
AEP	auditory evoked potential(s)
AFC	alternative forced choice
AFNI	analysis of functional neuroimages
ANOVA	analysis of variance
AVCN	anterior ventral cochlear nucleus
BCV	broadcast call vocalization
BPN	band-pass noise
BOLD	blood-oxygen-level-dependent
CAPD	central auditory processing disorder
dB	decibel
DCN	dorsal cochlear nucleus
DTI	diffusion tensor imaging
EEG	electroencephalography
ERP	event-related potential(s) or evoked-response potential(s)
FDRR	frequency-dependent response region
FFA	fusiform face area
FFR	frequency following response
fNIRS	functional near infrared spectroscopy
fMRI	functional magnetic resonance imaging
FWHM	full-width at half max
FTPV	fronto-temporal positivity to vocalizations
GFP	global field power
HG	Heschl's gyrus
HNR	harmonics-to-noise ratio
Hz	Hertz
IC	inferior colliculus
IRN	iterated ripple noise
ISI	inter-stimulus interval
LI	lateralization index
LL	lateral lemnisci
MEG	magnetoencephalography
MGB	medial geniculate body
MI	maturation index
μV	microvolt
ms/msec	millisecond(s)
mSTG	middle superior temporal gyrus
OPTR	operation TR
PAC	primary auditory cortex
PET	positron emission tomography
PRV	proximal vocalization
PT	pure tone / planum temporale
ROI	region of interest
RMS	root mean square

s	second(s)
SD/s.d.	standard deviation
SE	standard error(s)
SFM	spectral flatness measure
SLI	specific language impairment
SOC	superior olivary complex
SPGR	spoiled gradient recalled
SSV	spectral structure variation
STG	superior temporal gyrus
STS	superior temporal sulcus
SUMA	surface mapping with AFNI
TE	echo time
TFCE	threshold-free cluster enhancement
TR	repetition time
TRV	temporally reversed vocalization
TTL	transistor-transistor logic
TVA	temporal voice areas
VLPFC	ventral lateral prefrontal cortex
VSR	voice specific response
WTA	winner take all

CHAPTER 1:

Introduction and Literature Review

Vocalizing species critically rely upon highly optimized neuronal circuits for effective auditory communication. The information within a vocal signal can include insight into the emotional state of an animal, its intentions, and especially in the case of humans, semantic content. Humans have arguably placed even greater demands and requirements on their auditory systems – a consequence of the added complexities and subtleties of spoken language and other non-verbal communication sounds. The transient nature of vocalizations also requires these processes to be very rapid; very subtle acoustic changes or events that impart significant meaning often evolve on a time-scale of milliseconds. Not surprisingly, neuronal preference to the vocalizations of one's species (conspecifics) has been demonstrated in the auditory structures of many primate and non-primate species. Sensitivity to specific acoustic attributes or combinations of attributes intrinsic to vocalizations likely form the early stages of these networks. These neuronal biases for specific acoustic features ostensibly aid the rapid segregation of vocalizations and other behaviorally relevant sounds from an auditory scene.

This literature review will provide background information for the concept of "conspecific vocalization sensitivity" and how it has been shown previously to be represented in neuronal structures, especially auditory cortices. The focus of this review concerns how these functions are represented in the human auditory system and how they form the foundation of auditory-based communicative skills. Numerous auditory and speech pathologies affecting communication skills arise as inadequate or compromised cortical representations of vocalization sounds (e.g presbycusis, central auditory processing disorders (CAPD), autism, etc.). Thus, the background information presented here will build the scientific rationale for the experiments discussed in subsequent chapters that investigated the network structure and physiology of these complicated and significant human auditory pathways.

I. The Human auditory system – Basic anatomy and physiology

Auditory systems rapidly process incoming acoustic information, subserving functions such as identifying sound sources, assessing source locations, and in some species, creating highly accurate spatial renderings of their environmental surroundings (echolocation). Humans, through non-verbal and spoken language faculties, use sound as an efficient medium to transmit large amounts of semantic information and to express nuanced emotions. These skills are supported by a network of anatomical structures between the cochleae and cortex. Even though the experiments described in the following chapters investigated neuronal responses of cortical origins, brief and general descriptions of certain major structures in the entire auditory pathway and their putative functions are given below as background.

The cochlea and auditory brainstem

Sound that enters the ears begins its transduction from physical phenomenon to biochemical/-electric signal at the tympanic membrane (ear drum) (Figure 1-1) (Noback et al., 2005). Oscillations of this membrane are transferred through an elegantly levered system of tiny bones (malleus, incus, and stapes) that in turn apply pressure to the primary sensory organ of the auditory system, the coiled and multi-chambered cochlea. Through a membrane called the oval window, these mechanical perturbations are transduced into fluid pressure waves. Along the length of the cochlea is a resonant structure called the basilar membrane (Figure 1-2) (*ibid*). The basilar membrane resonates at different frequencies along its length; following a logarithmic tonotopic (by frequency) distribution, high frequency sounds resonate the base, or oval-window end of the cochlea, and low frequency sounds resonate the other end apex. Resonation of the basilar membrane causes small hair cells (possessing rows of hair-like stereocilia) to produce a fluid shear force between the reticular and tectorial membranes. These structures and others form the basis of a structure referred to as the Organ of Corti that houses the majority of the molecular machinery of the cochlea (ten Donkelaar, 2011). The shearing

forces of the hair cells cause neurotransmitters to be released, exciting the dendritic end of the cochlear nerve; its cell bodies reside in the spiral ganglion. Action potentials then proceed down the cochlear nerve to the cochlear nucleus of the brainstem (Figure 1-3).

The cochlear nuclei are anatomically subdivided broadly into dorsal and ventral sections; this division results in to quasi-distinct information processing streams. The ventral section (especially the anterior section, AVCN) primarily functions for sound localization, projecting onto the bilateral superior olivary complexes (SOC) (ten Donkelaar, 2011). Projections between the SOCs and the inferior colliculus form the lateral lemnisci (LL). The dorsal section (DCN) primarily extracts spectral and temporal acoustic information for source identification, sending most of its projections onto the contralateral inferior colliculus (IC). Information from these two streams seems to predominantly converge and integrate at the level of the inferior colliculi. The IC primarily projects to the medial geniculate body (MGB), a portion of the thalamus that is dominated by auditory functions. The ventral portion of the MGB is laminar and displays tonotopic organization similar to its primary source of input, the laminar central nucleus of the IC (Morest, 1965, Oliver and Morest, 1984). The ventral MGB sends projections, via the acoustic radiations, to the transverse temporal gyri (Heschl's gyri) in cortex, proposed locations for primary auditory cortices in humans (Kaas et al., 1999). The dorsal and medial portions of the MGB (which are non-laminar) send projections to the planum temporale and other "higher-order" auditory cortical regions. The overall functional layout of the MGB is somewhat elusive (except for the ventral MGB) and it is highly influenced by the cortex through modulatory corticofugal inputs, perhaps owing to its purported role as an interface between the cortex and brainstem structures (Miller and Schreiner, 2000, Winer et al., 2001). Generally, the anatomical elements of the auditory brainstem show some well-defined functional organization. Notably, most of these structures possess tonotopic axes. This organization persists in some manner to the level of primary auditory cortices; we utilized cortical tonotopy maps as functional landmarks with respect to harmonic content processing in Chapter 3.

Auditory cortex

The primary auditory cortex in primates is purported to lie along the superior temporal plane; the cytoarchitectonic architectures of these cortices have been best described in the macaque (Figure 1-4). "Core" areas are usually defined in auditory cortices by their architectonic specificity, predominant thalamic input from the ventral MGB (Figure 1-5), and highly defined functional topographies; tonotopic representations in these regions usually display very systematic organization (Kaas et al., 1999). Surrounding these core regions are belt regions, forming a ring around the central core. Parabelt regions form another band of anatomically distinct regions that are located laterally to the core/belt structures. Further from the core, tonotopic organization becomes less organized, likely reflecting more specialized computations of auditory signals such as the processing of sound-source identity and vocalization features (Petkov et al., 2008). Additionally, belt and parabelt regions further project laterally into the superior temporal sulci (STS) and other temporal lobe structures as well as into frontal regions.

The auditory cortex in humans has similarly been described in architectonic and functional studies. Anatomically, the human cortical locations of primary auditory, or core regions, likely occur along the Heschl's gyri (Morosan et al., 2001, Rademacher et al., 2001). Belt and parabelts regions are likely to be found along the middle aspects of the STG, planum temporale, and into the STS (Morosan et al., 2005). Functionally, numerous attempts have been made to describe tonotopically-defined core regions in human auditory cortices (Formisano et al., 2003, Talavage et al., 2004, Lewis et al., 2009, Talkington et al., 2012). Generally, mirror-symmetric tonotopic maps can be revealed on or near Heschl's gyri, oftentimes posterior to the gyri. Woods et al. have performed very detailed functional analyses of the proposed auditory cortical fields in humans in an attempt to identify core-belt-parabelt organization and tuning similar to that seen in macaque models (Figure 1-6) (Kaas and Hackett, 2000, Woods et al., 2010).

The elucidation of the functions of auditory cortex and related regions has often paralleled work and theory accomplished in the visual faculties (Rauschecker and Scott, 2009). Similar to the visual system, two pathways for information were derived for the

auditory system: an anterior-ventral pathway for object identification ("what" pathways) and a posterior-dorsal pathway for sound localization and action ("where" and/or "how" pathways) (Goodale and Milner, 1992, Kaas and Hackett, 2000, Rauschecker and Tian, 2000, Tian et al., 2001, Arnott et al., 2004, Lomber and Malhotra, 2008). The anterior-ventral "what" pathway is thought to house cortical processing pathways that analyze the fine acoustic structure of incoming auditory stimuli, allowing for very precise identification of sound sources; in cases of living sound sources, the sex, health, emotional state, and even intent can be surmised by close analyses of their respective vocal signals.

One prominent dimension of vocalization sounds is the presence of strong harmonic content, combinations of acoustic frequency components that are interrelated by simple mathematical relationships. Vocal cords and homologous structures predominantly produce sound by vibrating columns of air, a physical arrangement that engenders strong harmonic acoustic energy (Riede et al., 2001, Fitch et al., 2002). Beyond tonotopic representations in primary auditory cortices, we surmised that early vocalization processing pathways in human auditory cortices should show sensitivity to the harmonic content of vocalizations and other sounds. Harmonics, or very specific combinations of frequency information, are proposed to form the basis of higher-order auditory circuits and networks in simulations as well as in biological networks that show "combination-sensitive" neuronal activation; combination-sensitive activation usually is strongest when stimuli contain similar acoustic profiles to conspecific vocalizations or other ethologically relevant stimuli (Suga et al., 1983, Lewicki and Konishi, 1995, Medvedev et al., 2002, Medvedev and Kanwal, 2004, Kumar et al., 2007). Thus, in Chapter 2, we describe an experiment in which fMRI was used to characterize auditory cortices that showed parametric sensitivity to the harmonic energy, quantified with a harmonics-to-noise ratio (HNR), of artificial iterated rippled noises (IRN) and animal vocalizations. Regions sensitive to HNR were anatomically compared to regions that were tonotopically organized as well as cortex that produced preferential fMRI BOLD activity to human-produced vocalizations. Additionally, in Chapter 3, we recorded auditory evoked potentials (AEP) in response to IRN stimuli to describe processing of harmonic content in

auditory cortices at a finer temporal scale; Chapter 4 investigated the influence of native language experiences on HNR-sensitive AEPs.

II. Voice-sensitivity in human auditory cortices

Functional neuroimaging findings

The ability to recognize, process, and produce non-verbal and verbal (speech) vocalizations form the auditory foundation of human communication skills. These abilities likely rely upon highly optimized neuronal circuits that are exquisitely sensitive to the acoustic signal characteristics of those sounds. The visual dominance of the fMRI field led to the early functional descriptions of cortical regions that produced greater BOLD activity to images to faces versus other categories of visual objects (Kanwisher et al., 1997). Specifically, Kanwisher et al. claimed that the fusiform face area (FFA) was selective for processing faces rather than being a region involved in more general visual object processing (though see below).

These findings in the visual realm drove researchers to investigate the presence of functionally homologous regions in the auditory system that would produce the greatest activity to human voices, rationalizing that voices are analogous to "auditory faces." A seminal study by Belin et al. revealed voices areas along the upper banks of the bilateral STS that preferentially responded to human vocal sounds (speech or non-verbal vocalizations) versus other categories of complex non-vocalization sounds (Belin et al., 2000). Previously, Binder et al. had demonstrated anterior and middle regions of the bilateral STSs that responded with greater BOLD intensity to speech sounds (words, pseudo-words, and reversed speech) versus frequency-modulated tones. Due to the speech or speech-like nature of the stimuli, left hemisphere activation often produced greater expanses of activity, consistent with historical notions of left-hemisphere dominance for speech processing (Parker et al., 2005), though see (Hickok and Poeppel, 2007). Others have also described hierarchies that are organized by the intelligibility of speech signals (Scott et al., 2000). These hierarchies are most organized in left

hemisphere temporal cortices (Davis and Johnsrude, 2003). Generally, as speech signals become more intelligible, preferential BOLD activity shifts spatially towards the STS.

The findings of Belin et al. more specifically showed that these central STS areas were sensitive to voices by including non-verbal vocalizations. The use of non-verbal human vocalizations also produced a slight right-hemisphere bias for BOLD activity (Belin et al., 2000). It was suggested that a dominant role for the right hemisphere is processing the more paralinguistic elements of vocal sounds. Specifically, the anterior temporal lobe been shown to be crucially involved with vocal identity of speakers using fMRI with both attention and adaptation paradigms (Belin and Zatorre, 2003, von Kriegstein et al., 2003). Interestingly, this skill is thought to be present before birth in humans and has been recorded in other non-human primate species (DeCasper and Fifer, 1980, Kisilevsky et al., 2003, Petkov et al., 2008). Future work in the field of cortical vocalization processing networks will greatly benefit from similar studies performed in developing infants and children as well as non-human primates. A greater anatomical and functional understanding of analogous cortical networks in these populations will be crucial to revealing the more "primitive" and foundational elements that become fully developed in adult humans. Reviews of vocalization processing studies in infants and non-human primates are provided in Chapters 5 and 6.

Electroencephalographic and magnetoencephalographic findings

Electromagnetic neuroimaging techniques (EEG, MEG, and derivatives) have also been used to describe human neuronal responses to vocalization sounds and speech. These methods are complementary to fMRI; both provide characterized neuronal behavior on a very fine temporal scale (millisecond resolution). Levy et al., motivated by the aforementioned fMRI voice-selective findings (Belin et al., 2000, Belin et al., 2002), designed an experiment to test the presence of a homologous electrophysiological effect (Levy et al., 2001). Comparing AEPs to the timbre of instrumental sounds (brass, wind, and string instruments) to those produced by the timbre of sung tones with matched fundamental frequencies, they described a "voice-specific response" (VSR) occurring at approximately 320ms after stimulus onset. This same group subsequently designed a set

of follow-up experiments to assess the effects of task and attention on the VSR. Participants ignoring the stimuli and performing a task focused on an auditory discrimination other than timbre had drastic effects on the VSR (Levy et al., 2003). In these instances, significant amplitude differences between voices and non-voice instrument sounds were non-existent. The authors interpreted the VSR to represent a marker for the allocation of attentional auditory resources.

MEG has also been used to assess the VSR (Gunji et al., 2003). Gunji et al. reported comparable N1m amplitudes in response to voice and non-voice stimuli. Sustained fields (SF) that occurred 400ms after stimulus onset produced somewhat stronger magnetic sources to voice stimuli. Though this response was relatively close to the VSR in time (~320ms), it was likely not produced by the same cortical generators and was much smaller in overall effect size. The cortical generators of the VSR were thought to be generated near the STS and radially-oriented, a configuration that does not tend to produce strong extracranial magnetic fields. Additionally, the subjects were not attending to the stimuli, which was shown to have an effect on the VSR (Levy et al., 2003).

The VSR seemed to insufficiently describe the neuronal mechanisms on voice processing as more findings were published. Murray et al. demonstrated that there are differential electrophysiological responses as early as 70ms to sounds produced by man-made objects versus those produced by living beings (including vocalizations and non-vocalizations) (Murray et al., 2006). Additionally, others had subsequently revealed various "voice-processing responses" to real and artificial vocalizations occurring earlier than the VSR related to paralinguistic features and/or vocal adaptation effects (Schweinberger, 2001, Lattner et al., 2003, Beauchemin et al., 2006, Zaske et al., 2009). Motivated by these findings, Charest et al. performed an oddball-detection (1000 Hz pure tone) experiment to directly compare the AEP responses to voice sounds with bird vocalizations and environmental sounds (Charest et al., 2009). They described a "Fronto-Temporal Positivity to Voices" (FTPV) produced at fronto-temporal electrode locations (e.g. FC5/6) that occurred at approximately 164ms after stimulus onset. However, speech and non-verbal vocalizations were used in the "voice" category; both categories of sounds produce similar activation maps on the scalp when compared to non-voice stimuli, but the responses were greater in amplitude and more widely spread to the

speech-containing stimuli. Notwithstanding, the authors did show a preferential response to voice occurring at a response time much earlier than the VSR.

Concerns about responses that depended upon attention to the "voice-ness" of stimuli or responses that seemed to be dominated by speech stimuli prompted one group to re-analyze previously reported data within the context of conspecific vocalization processing (De Lucia et al., 2010). De Lucia et al. investigated electrophysiological responses to animal vocalizations and non-verbal human vocalizations used in a previous study that investigated cortical processing differences between sounds produced by living and "man-made" sources (Murray et al., 2006). Their previous study required participants to discriminate between living and man-made sound sources. Thus, they reasoned that any discrimination between human and animal vocalizations would be implicit and attention would be primarily devoted to features other than "voice-ness". Additionally, the authors claim that similar studies involving ERPs (Levy et al., 2001, 2003, Charest et al., 2009) were critically dependent upon reference electrode choice, a decision that can drastically alter the statistical outcomes of voltage waveform findings (spatial scalp distributions and latencies) (Nunez and Srinivasan, 2006, Murray et al., 2008). De Lucia et al. therefore opted to use reference-independent global field power (GFP) calculations to avoid this potential confound (Lehmann and Skrandies, 1980, Murray et al., 2008). The earliest reliable GFP differences between animal and human vocalizations were found 169-219ms after stimulus presentation. Topographical analyses suggested that animal and human vocalization responses originated from indistinguishable networks; the GFP results reflected different strengths of network activation. Note, however, that a lack of topographical differences in this study does not preclude network differences at a finer spatial scale (Nunez and Srinivasan, 2006).

The methodology and findings of De Lucia et al., though more rigorous than previous studies, did not significantly change the current temporal landscape of conspecific vocalization processing in humans. The aforementioned Charest et al., using "simpler" ERP methods, reported very similar diverging electrophysiological responses between animal and human vocalization stimuli occurring after 164ms. Nonetheless, De Lucia et al. proposed a relatively comprehensive temporal hierarchy for human auditory processing based on their original study and subsequent analyses (Murray et al., 2006, De

Lucia et al., 2009). Their posited temporal hierarchy is four-tiered: (1) "general" sound processing (low-level spectrotemporal processing) occurs before approximately 70ms, (2) the differentiation between man-made (machinery and instruments) and living (non-verbal human vocalizations and animal vocalizations) sound sources occurs in a window near 70-119ms, (3) human versus animal vocalization discrimination occurs between approximately 169-219ms, and (4) music versus non-music discrimination occurs around 291-357ms. Note that the latter two tiers tend to support the previous findings of Charest et al. and Levy et al., respectively (Levy et al., 2001, 2003, Charest et al., 2009).

The currently proposed hierarchy of De Lucia et al. temporally situates the human brain's ability to discriminate between conspecific and non-conspecific vocalizations at a relatively early time point in cortical auditory processes. However, if the human auditory system is optimized – intrinsically, through development, or a combination of both – for processing human vocalizations, one could reasonably hypothesize that there should be detectable processing differences in even "earlier" auditory structures and processes. The time point for this phenomenon described by De Lucia et al. may have been a product of their chosen human stimuli, stereotypically produced human vocalizations, and their experimental task design, a cognitive discrimination between man-made or living sound sources.

III. Conspecific vocalization sensitivity

The aforementioned studies of vocalization processing in human auditory cortices have utilized numerous types of human and animal vocalizations, mechanical, and environmental sounds. However, none of the aforementioned studies have adequately controlled for conspecific vocal content when describing cortical regions as sensitive or preferential to human vocal signals. Specifically, the human vocalizations used have been *stereotypical*, that is, those sounds that are within the normal repertoire of human-produced vocalizations. Language sounds and non-verbal vocalizations such as humming, coughing, yawning, crying, screaming, etc. are all commonly encountered conspecific human vocalizations. As a result, the auditory systems of fully developed

adults are likely very adept at processing these sounds. Over-learned stimuli may unintentionally activate routinely used higher-order cortical networks (or "schemata") (Alain, 2007), overshadowing neuronal activity in more primary auditory regions. This would prevent investigations of the auditory networks that are most critical for discriminating between conspecific and non-conspecific vocalizations. Thus, we utilized a novel class of human vocalizations, human-mimicked versions of animal vocalizations, as a critical control for over-established auditory network activity. Human-mimicked animal vocalizations are an ideal platform for investigating vocalization processing in humans because they are naturally produced within the acoustic limits of the human vocal and articulatory structures (and thus are conspecific-produced), but they also minimize activity in networks involved with the processing of language and other stereotypical human vocalizations. Chapter 5 describes a human fMRI study that revealed conspecific vocalization sensitive regions predominantly near left primary auditory cortices (PACs), situated in much earlier auditory cortical stages than traditional voice-sensitive regions in the bilateral STS. Additionally, Chapter 7 describes the temporal processing dynamics for this phenomenon with a set of electrophysiology experiments. Together, these findings provide complementary results for building a more complete spatiotemporal hierarchy of vocalization processing in human auditory cortices.

FIGURES

FIGURE 1-1

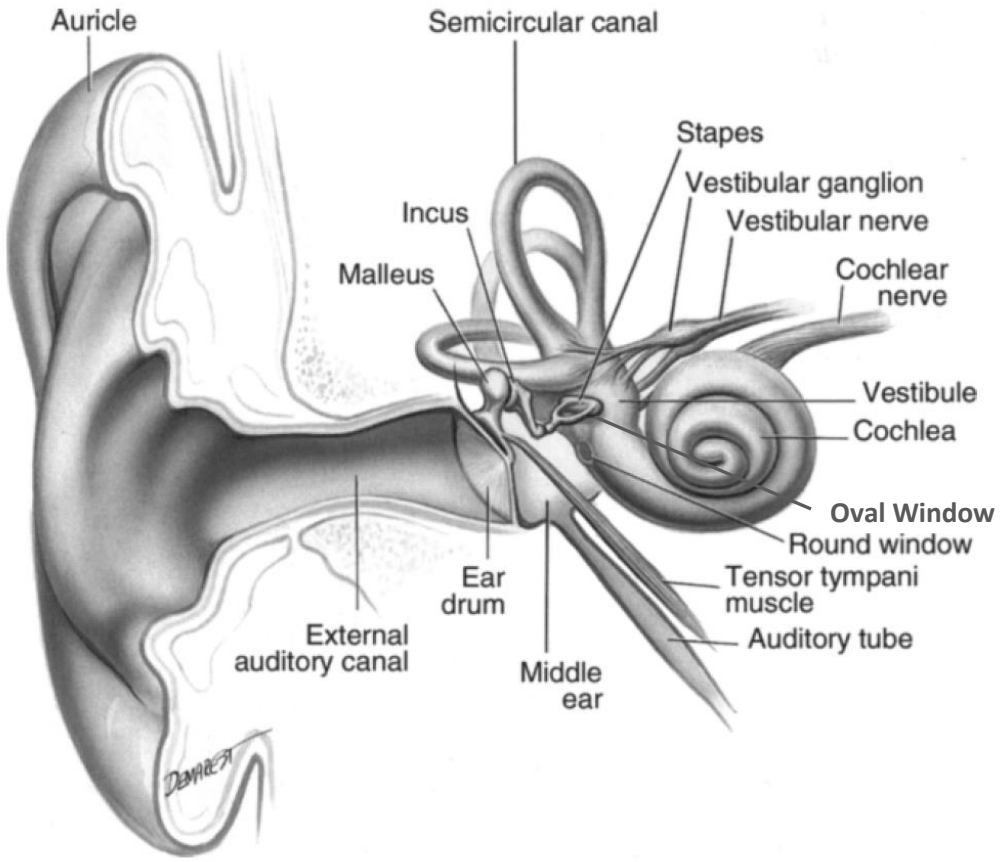


FIGURE 1-1. External ear, middle ear, and inner ear on right side viewed from the front. The oval window is positioned under the "face" of the stapes. This illustration and caption has been adapted from another source (Noback et al., 2005).

FIGURE 1-2

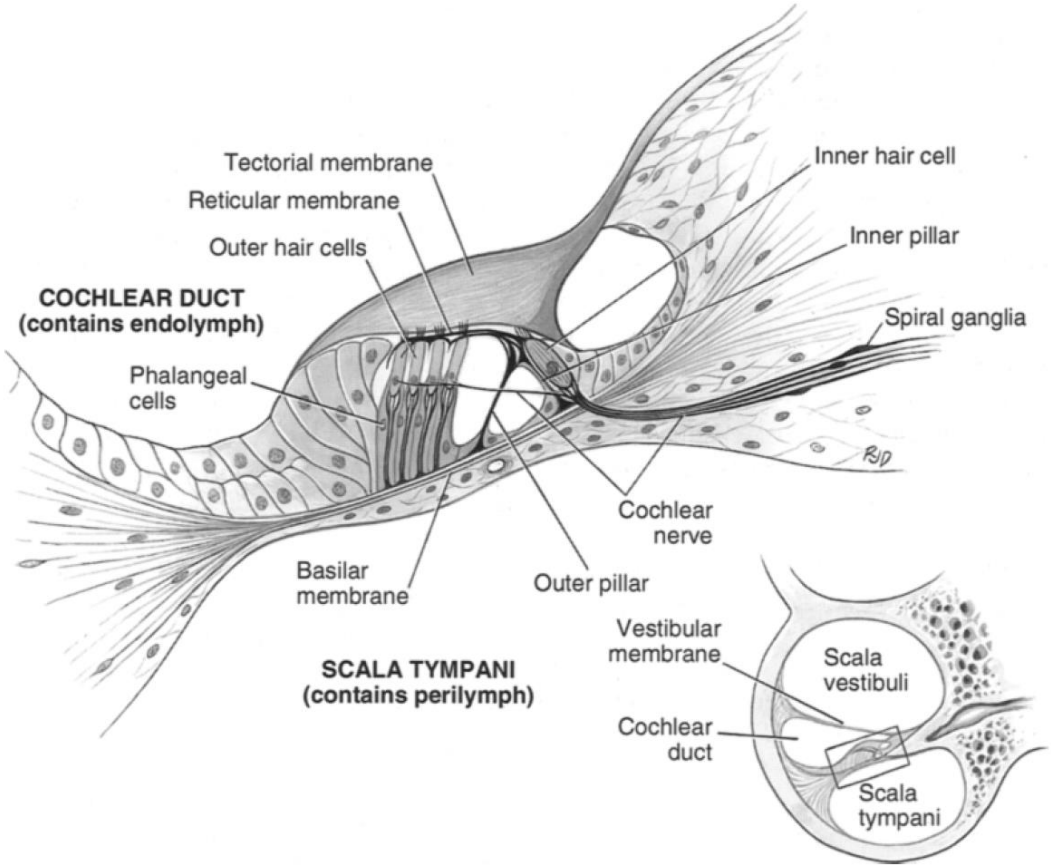


FIGURE 1-2. The organ of Corti in the middle turn of the cochlea with four rows of outer hair cells; there are three rows in the basal turn and five in the apical half-turn, reflecting the fact that the basilar membrane is wider at the apex. Inner and outer pillar cells enclose the tunnel of Corti, which contains perilymph and is traversed by cochlear nerve fibers. The pillars have fenestrated stiff processes that cover the apical surface of the hair cells. This illustration and caption has been adapted from another source (Noback et al., 2005).

FIGURE 1-3

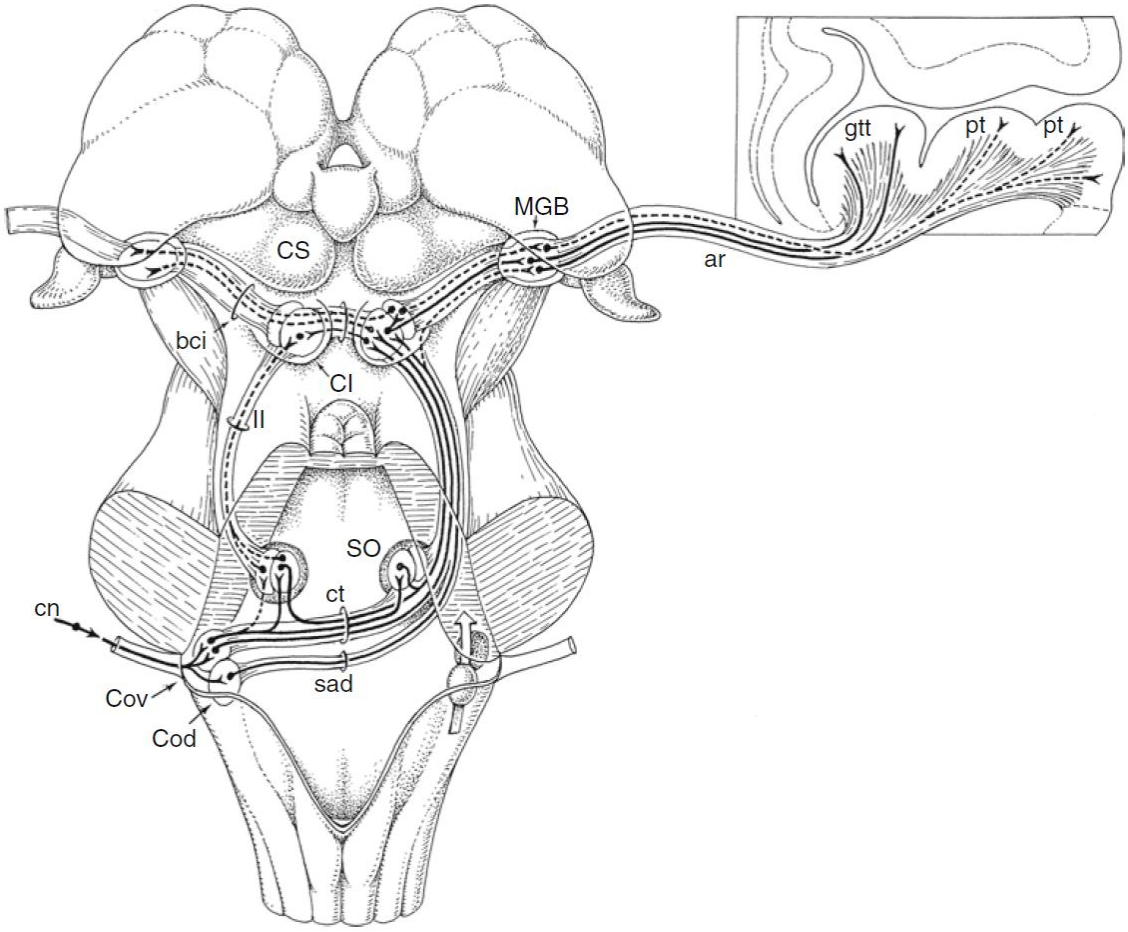


FIGURE 1-3. Overview of the nuclei and projections in the human auditory brainstem. *Abbreviations*: *ar* acoustic radiation; *bci* brachium of colliculus inferior; *CI* colliculus inferior; *cn* cochlear nerve; *Cod*, *Cov* dorsal and ventral cochlear nuclei; *CS* colliculus superior; *ct* corpus trapezoideum; *gtt* gyrus temporalis transversus (Heschl's or transverse temporal gyrus); *ll* lateral lemniscus; *MGB* medial geniculate body; *pt* planum temporale; *sad* stria acoustica dorsalis; *SO* superior olive. This illustration and caption has been adapted from another source (ten Donkelaar, 2011).

FIGURE 1-4

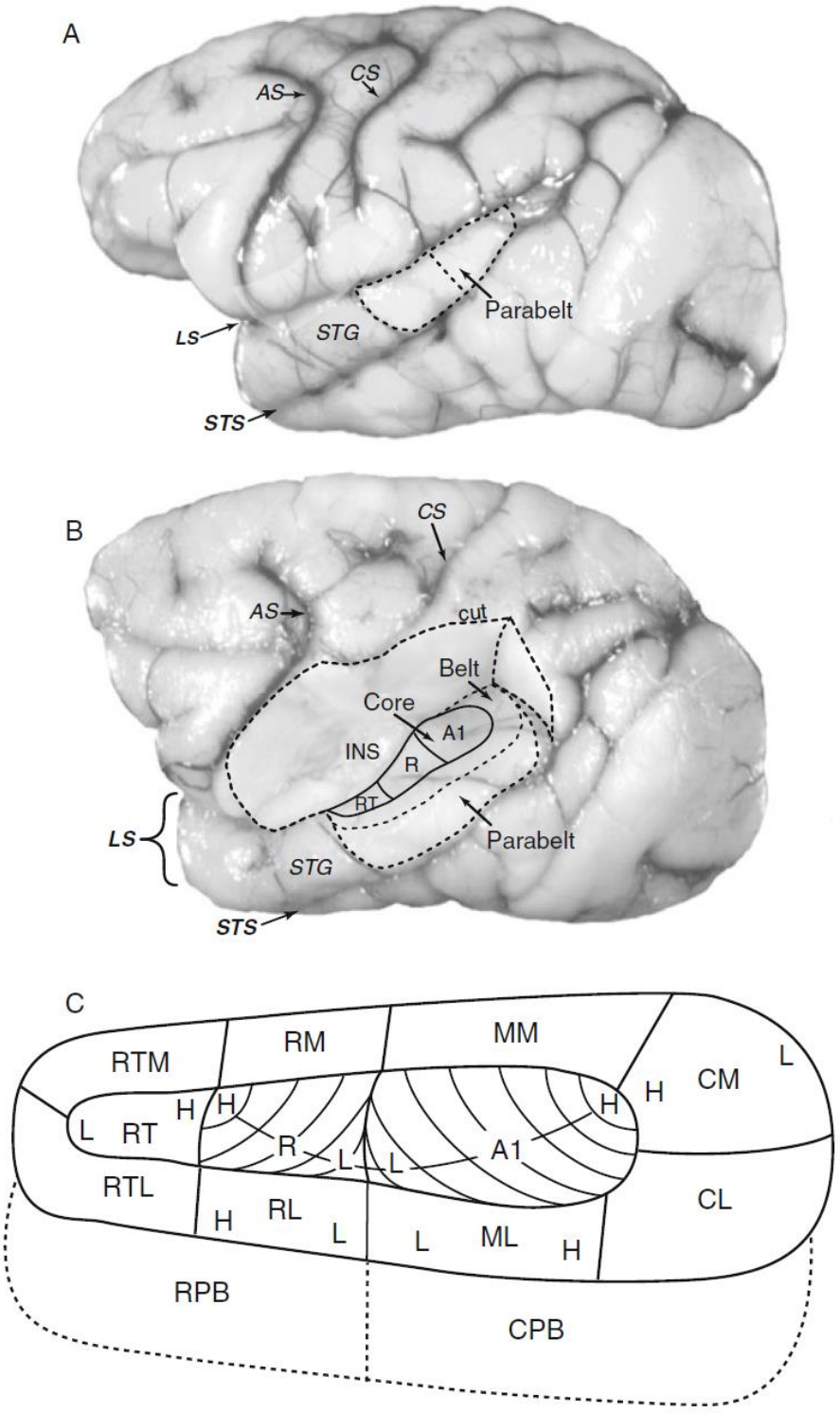

A
AS
CS
Parabelt
LS
STG
STS
B
CS
AS
cut
Belt
Core
A1
INS
R
RT
LS
Parabelt
STG
STS
C
RTM
RM
MM
L
CM
H
H
H
H
L
RT
R
L
L
A1
RTL
CL
H
RL
L
L
ML
H
RPB
CPB

FIGURE 1-4. The locations of primary and secondary auditory areas in the cortex of macaque monkeys. (A) The primary areas are within the ventral bank of the lateral sulcus, and are not apparent in this lateral view of the intact brain. Only the parabelt, a third level of auditory processing, is apparent. The lateral sulcus (LS), superior temporal sulcus (STS), and the central sulcus (CS) are indicated for reference. **(B)** Cortex of the upper bank of the lateral sulcus has been removed (*dashed line*) to reveal the auditory core and belt on the lower bank of the lateral sulcus. The insula (INS) is an island of cortex between the two banks. **c** A schematic of auditory cortex organization. A core of primary-like areas includes AI, a rostral area (R), and a rostrotemporal area (RT). Each of these areas is tonotopically organized from low (L) to high (H) frequencies. *Lines* of isorepresentation are shown for AI and R. The core is surrounded by a belt of secondary areas denoted by location: CL, caudolateral area; CM caudomedial area; ML, middle lateral area; RM, rostromedial area; AL, anterolateral area; RTL, lateral rostrotemporal area; RTM, medial rostrotemporal area. The lateral parabelt, a third level of processing, has been divided into rostral (RPB) and caudal (CPB) zones. Many of the belt areas are at least crudely tonotopically organized (Kaas and Hackett, 2000). This illustration and caption has been adapted from another source; Chapter 19 of (Winer and Schreiner, 2011) by Jon Kass.

FIGURE 1-5

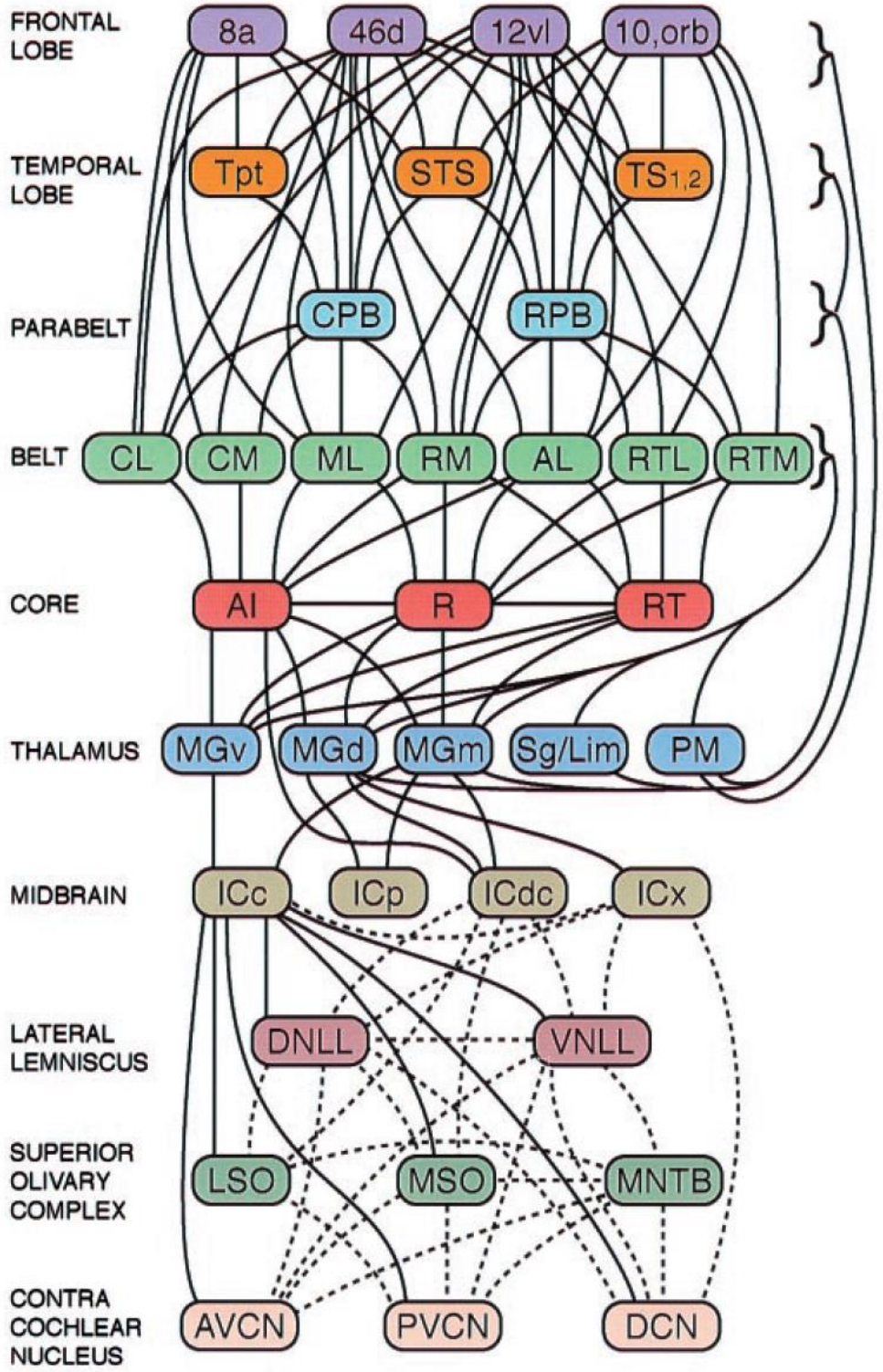

FRONTAL LOBE
8a
46d
12vl
10,orb
TEMPORAL LOBE
Tpt
STS
TS1,2
PARABELT
CPB
RPB
BELT
CL
CM
ML
RM
AL
RTL
RTM
CORE
Al
R
RT
THALAMUS
MGv
MGd
MGm
Sg/Lim
PM
MIDBRAIN
ICc
ICp
ICdc
ICx
LATERAL LEMNISCUS
DNLL
VNLL
SUPERIOR OLIVARY COMPLEX
LSO
MSO
MNTB
CONTRA COCHLEAR NUCLEUS
AVCN
PVCN
DCN

FIGURE 1-5. Cortical and subcortical connections of the primate auditory system. Major cortical and subcortical regions are color coded. Subdivisions within a region have the same color. Solid black lines denote established connections. Dashed lines indicate proposed connections based on findings in other mammals. Joined lines ending in brackets denotes connections with all fields in that region. The belt region may include an additional field, MM (see Fig. 5). Abbreviations of subcortical nuclei: AVCN, anteroventral cochlear nucleus; PVCN, posteroventral cochlear nucleus; DCN, dorsal cochlear nucleus; LSO, lateral superior olivary nucleus; MSO, medial superior olivary nucleus; MNTB, medial nucleus of the trapezoid body; DNLL, dorsal nucleus of the lateral lemniscus; VNLL, ventral nucleus of the lateral lemniscus; ICc, central nucleus of the inferior colliculus; ICp, pericentral nucleus of the inferior colliculus; ICdc, dorsal cortex of the inferior colliculus; ICx, external nucleus of the inferior colliculus; MGv, ventral nucleus of the medial geniculate complex; MGd, dorsal nucleus of the medial geniculate complex; MGm, medialymagnocellular nucleus of the medial geniculate complex; Sg, suprageniculate nucleus; Lim, limitans nucleus; PM, medial pulvinar nucleus. Abbreviations of cortical areas: AI, auditory area I; R, rostral area; RT; rostrotemporal area; CL, caudolateral area; CM, caudomedial area; ML, middle lateral area; RM, rostromedial area; AL, anterolateral area; RTL, lateral rostrotemporal area; RTM, medial rostrotemporal area; CPB, caudal parabelt; RPB, rostral parabelt; Tpt, temporoparietal area; TS1,2, superior temporal areas 1 and 2. Frontal lobe areas numbered after the tradition of Brodmann and based on the parcellation of (Preuss and Goldmanrakic, 1991): 8a, periarcuate; 46d, dorsal principal sulcus; 12vl, ventrolateral area; 10, frontal pole; orb, orbitofrontal areas. This illustration and caption has been adapted from another source (Kaas and Hackett, 2000).

FIGURE 1-6

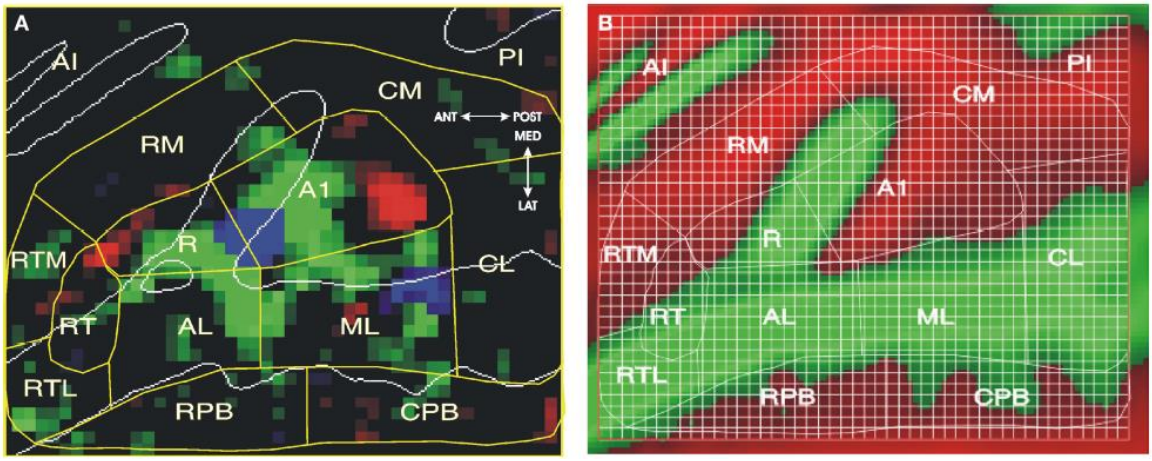


FIGURE 1-6. Auditory cortical fields (ACFs). (A) Best-frequency map, showing best frequency at each voxel relative to the two other frequencies. Saturation codes the magnitude of frequency preference (range: 0.07–0.15% difference). Red = 3600 Hz, Green = 900 Hz, Blue = 225 Hz. ACFs (yellow lines) were assigned following the model of Kaas et al. (1999). Auditory core fields were identified by their mirror-symmetric tonotopic organization with surrounding belt fields divided at the boundaries between adjacent core ACFs. White lines indicate gyral boundaries. See text for ACF labels. **(B)** Model projected on average curvature map of the superior temporal plane (green = gyri, red = sulci), showing anatomical structures and grids used for quantification. Abbreviations: HG: Heschl's gyrus; PT: planum temporale; STG: superior temporal gyrus; STS: superior temporal sulcus. This illustration and caption has been adapted from another source (Woods et al., 2010).

CHAPTER 2:

Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute

James W. Lewis[1], William J. Talkington[1], Nathan A. Walker[1],
George A. Spirou[1,2], Audrey Jajosky[1], and Chris Frum[1]

[1]Center for Advanced Imaging, Sensory Neuroscience Research Center, and
Department of Physiology and Pharmacology,
[2]Deptartment of Otolaryngology
West Virginia University, Morgantown, WV 26506, USA

This manuscript was published in the Journal of Neuroscience on February 18th, 2009.

ABSTRACT

The ability to detect and rapidly process harmonic sounds, which in nature are typical of animal vocalizations and speech, can be critical for communication among conspecifics and for survival. Single-unit studies have reported neurons in auditory cortex sensitive to specific combinations of frequencies (e.g. harmonics), theorized to rapidly abstract or filter for specific structures of incoming sounds, where large ensembles of such neurons may constitute spectral templates. We studied the contribution of harmonic structure to activation of putative spectral templates in human auditory cortex by using a wide variety of animal vocalizations, as well as artificially constructed iterated rippled noises (IRNs). Both the IRNs and vocalization sounds were quantitatively characterized by calculating a global harmonics-to-noise ratio (HNR). Using fMRI we identified HNR-sensitive regions when presenting either artificial IRNs and/or recordings of natural animal vocalizations. This activation included regions situated between functionally defined primary auditory cortices and regions preferential for processing human non-verbal vocalizations or speech sounds. These results demonstrate that the HNR of sound reflects an important second-order acoustic signal attribute that parametrically activates distinct pathways of human auditory cortex. Thus, these results provide novel support for putative spectral templates, which may subserve a major role in the hierarchical processing of vocalizations as a distinct category of behaviorally relevant sound.

INTRODUCTION

In the mammalian auditory system, recognizing and ascribing meaning to real-world sounds relies on a complex combination of both "bottom-up" and "top-down" grouping cues that segregate sounds into auditory streams, and ultimately lead to the perception of distinct auditory events or objects (Wang, 2000, Cooke and Ellis, 2001, Hall, 2005). To increase signal processing efficiency, different classes of sound may be directed along specific cortical pathways based on relatively low-level signal attributes. In humans, animal vocalizations, as a category of sound distinct from hand-tool sounds, are reported to more strongly activate the left and right middle superior temporal gyri (mSTG), independent of whether or not the sound is correctly perceived, and independent of handedness (Lewis et al., 2005, Lewis, 2006, Lewis et al., 2006, Altmann et al., 2007). Consequently, at least portions of the mSTG appear to process "bottom-up" acoustic signal features, or primitives, characteristic of vocalizations as a distinct *category* of sound. However, what organizational principles, beyond tonotopic organizations derived from cochlear processing, might generally facilitate segmentation and recognition of vocalizations?

One such second-order acoustic signal attribute is the sound's harmonic structure, which can be quantified by the harmonics-to-noise ratio (HNR) (Boersma, 1993, Riede et al., 2001). Sounds with greater HNR value generally correlate with the perception of greater pitch salience. For instance, a snake produces a hiss with a very low HNR value, near that of white noise (Fig. 2-1a-b). In contrast, sounds such as a wolf howl, and some artificially created iterated rippled noise sounds (IRNs; see Methods), tend to have a more tonal quality and greater pitch salience, being comprised of more prominent harmonically related frequency bands ("frequency stacks") that persist over time (hear Supplementary Audios 1-10 online). In mammals, the harmonic structure of vocalizations stem from air flow causing vibrations of the vocal folds in the larynx, resulting in periodic sounds (Langner, 1992, Wilden et al., 1998). In other species, this process similarly involves soft vibrating tissues such as the labia in the syrinx of birds, or phonic lips in the nose of dolphins, which underscores the ethological importance of this basic mechanism of

"vocal" harmonic sound production for purposes of communication. HNR measures have proven useful for analyzing features of animal vocal production (Riede et al., 2001, Riede et al., 2005). In humans, HNR measures have also been used clinically to monitor recovery from voice pathologies (Shama et al., 2007), and used to assess signal characteristics of different forms of speech, such as sarcasm (Cheang and Pell, 2008). We previously reported that the "global" HNR values for human and animal vocalizations were substantially greater than for other categories of natural sound, suggesting that this could be a critical signal attribute that is explicitly processed in cortex to facilitate sound segmentation and categorization of vocalizations (Lewis et al., 2005).

Moreover, HNR is an attractive signal attribute to study from the perspective of neural mechanisms for auditory object or sound-source segmentation. Because harmonically structured sounds are comprised of specific combinations of acoustic peaks of energy at different frequencies (cf. Fig. 2-1b-d), HNR-sensitivity could potentially build off of tonotopically organized representations, thereby increasing receptive field complexity, similar to intermediate processing stages in the cortex for other sensory modalities. In several animal species (e.g. frogs, birds, bats, and primates), neurons in auditory cortex, or analogous structures, show facilitative responses to specific combinations of frequencies, notably including the harmonic structures typically found in conspecific vocalizations (Lewicki and Konishi, 1995, Rauschecker et al., 1995, Medvedev et al., 2002, Medvedev and Kanwal, 2004, Petkov et al., 2008). Ensembles of "combination-sensitive" neurons could filter for or extract harmonic features (or primitives). Such representations may reflect elements of theorized spectro-temporal templates that serve to group spectral and temporal components of a sound-source, resulting in coherent percepts (Terhardt, 1974, Medvedev et al., 2002, Kumar et al., 2007). In humans, a substantial portion of auditory cortex presumably is, or becomes, optimized for processing *human* vocalizations and speech (Belin et al., 2000, Scott, 2005). Thus, the presentation of sounds with parametrically increasing harmonic structure (HNR value)—approaching those typical of speech sounds—should grossly lead to the recruitment of greater numbers of, or greater activity from, combination-sensitive neurons. If observed, this would provide evidence for HNR-sensitivity, and thus support

for spectral templates in representing a neural mechanism for extracting and streaming vocalizations.

The above working model indicates that HNR-sensitive regions, based on combination-sensitive neural mechanisms, would require input from multiple frequency bands. Thus, HNR-sensitive regions, to minimize cortical wiring, should be located along or just outside of tonotopically organized areas, and so we mapped tonotopic functional landmarks in some individuals. Additionally, this hierarchical model indicates that HNR-sensitive regions should largely be located along the cortical surface between tonotopically organized regions and regions preferential for human-produced vocalizations. Thus, as additional functional landmarks, we also mapped cortices sensitive to human non-verbal vocalizations and to speech.

MATERIALS AND METHODS

Participants

We studied 16 right-handed adult English speaking participants (age 18 to 39 years; 10 women), who underwent one to five of our scanning paradigms. All participants were free of neurological, audiological, or medical illness, had normal structural MRI and audiometric examinations, and were paid for their participation. Informed consent was obtained following guidelines approved by the West Virginia University Institutional Review Board.

Iterated rippled noise (IRN) stimuli

As one measure for studying harmonic structure as an isolated signal attribute, we used iterated ripple noises, or IRNs (Yost, 1996b, Shofner, 1999), which have previously been used to study pitch and pitch salience processing in other human neuroimaging studies (Griffiths et al., 1998, Patterson et al., 2002, Penagos et al., 2004, Hall et al., 2005). By delaying and adding segments of white noise back to itself, IRN sounds with periodic harmonic structure can be constructed, producing sounds perceived to have a tonal quality embedded in white noise (Fig. 2-1b; hear Supplementary Audios 6-10 online). Wideband noise was systematically altered by temporal rippling, using custom Matlab code (V7.4, The Mathworks Inc., Natick MA, USA; Dr. William Shofner, personal communication). IRN stimuli were generated (44.1 kHz, 16-bit, monaural, ~6 sec duration) by a cascade of operations delaying and adding back to the original noise ("IRNO" in the terminology of Yost, 1996), with a given gain (g; ranging 0 to +1, in steps of 0.1) a delay (d; 0.25, 0.5, 1, 2, 4, and 8 msec), and a wide range of number of ripple iterations (n; including 1, 2, 4, 8, 16, 32, 64, 100, 200, 300, … to 2000). The perceived pitch of the IRN changes inversely with delay, and we included pitches of 125, 250, 500, 1000, 2000, and 4000 Hz, which were chosen to complement tonotopic mapping of cortex (see paradigm #1). Increasing the number of iterations and/or gain

qualitatively increases the clarity or strength of the perceived pitch (Penagos et al., 2004), which appears to be highly correlated with the harmonic content of the sound. In contrast to earlier studies using IRNs (ibid), we examined HNR measures of IRNs (see below), effectively manipulating pitch depth along the dimension of harmonic content. We created a much larger set of IRNs (~1700) so as to span a wide range of HNR values (see Supplementary Fig. 2-1). We then selected sixty-three IRNs to evenly sample across the dimension of HNR in steps of 3dB HNR (trimmed to 2.00 sec duration, and matched for overall root mean square (RMS) power: -12.0 ± 0.2 dB). More importantly, the quantitative HNR measure could be applied to behaviorally relevant real-world sounds (see below), and thus we sought to test a much wider range of IRN stimuli than have previously been studied, being comparable in HNR ranges observed for animal and human vocalizations.

Animal and human vocalization stimuli

We collected 160 professionally recorded animal vocalizations (Sound Ideas Inc. Richmond Hill, Ontario, Canada), which were typically recorded using stereo microphones containing two directional monaural microphones (44.1 or 48 kHz, 16-bit). Only one channel (left) was retained (down-sampled to 44.1 kHz) to remove binaural spatial cues (Cool Edit Pro v1.2, Syntrillium Software Co., now owned by Adobe Inc.), and the monaural recording was presented to both ears. Sounds included a wide variety of animals producing sound through a vocal tract or analogous structure. Care was taken to select sounds derived from only one animal with relatively little background or ambient noise, and to avoid aliasing, clipping and reverberation that could introduce spectrogram artifacts (Wilden et al., 1998). Most sounds were trimmed to 2.0 ± 0.2 sec duration, though a few sounds were of shorter duration (minimum 1.6 sec) to allow for more natural sounding acoustic epochs. Sound stimuli were ramped in intensity 20 msec to avoid spectral transients at onset and offset. Most of the animal sounds were matched in total RMS power to the IRN stimuli (at -12 dB). However, since some of the vocalization recordings included quiet or silent gaps, the overall intensity was necessarily lower for

some stimuli to avoid clipping (mean=-12.6 dB, range -8.2 to -20 dB total RMS power). Human spoken phrases and non-verbal vocalizations used in fMRI paradigm #5 were collected using the same techniques described above.

As part of an analysis of the potential behavioral relevance of the global HNR value of human vocalizations, we also recorded adult-to-adult and adult-to-infant speech from 10 participants, using professional recording equipment (44.1kHz, 16-bit, monaural) in a sound isolation booth (Industrial Acoustics, Inc., Bronx, NY). Each participant was provided with a brief script of topics for conversation, including describing weekend plans to another adult, and speaking to a baby (a baby doll was present) in an effort to make him smile. The script also included speaking onomatopoetic words describing different sub-categories of animal vocalizations, including phrases such as "a hissing snake" and "a growling lion". The stress phonemes, such as the "ss" in hiss, were selected and subjected to the same HNR analysis as the other sound stimuli, as described below.

Harmonics-to-noise ratio (HNR) calculation

We analyzed and calculated HNR values of all sound stimuli using freely available phonetic software (Praat, http://www.fon.hum.uva.nl/praat/). The HNR algorithm (below) determined the degree of periodicity within a sound signal, x(t), based on finding a maximum autocorrelation, $r'_x(\tau_{max})$, of the signal at a time lag (τ) greater than zero (Boersma, 1993):

$$HNR(\text{in dB}) = 10 * \log_{10} \frac{r'_x(\tau_{max})}{1 - r'_x(\tau_{max})}$$

This measure quantified the acoustic energy of the harmonics that were present within a sound over time, $r'_x(\tau_{max})$, relative to that of the remaining "noise", 1- $r'_x(\tau_{max})$, which represents non-harmonic, irregular, or chaotic acoustic energy. Three parameters influence the estimate of the harmonic structure of a sound, including a time step (10 msec), minimum pitch cutoff for its fundamental (75 Hz minimum pitch, 20kHz ceiling), and periods per window (1 per window). As extreme examples, white noise yielded an

HNR value of -7.6 with the above parameters, while a sample consisting of two pure tones (2 kHz and 4 kHz sine waves) produced an HNR value of +65.4.

Although no single set of HNR parameters is ideal for assessing all real-world sound stimuli (Riede et al., 2001) (Dr. Tobias Riede, personal communication), the periodic nature of the IRN stimuli lent themselves to a robust HNR-value estimate over the entire 2 second duration. The HNR values of the selected IRNs ranged from -3.5 to +25.2 dB HNR (grouped in increments of 3dB HNR), with ±1.3 dB HNR average standard deviation (range 0.3 to 7.9). For the animal vocalizations, we carefully selected those having a relatively stable pitch and cadence over time, ranging from -6.5 to +32.7 dB HNR with ±5.4 dB HNR average standard deviation (range 0.8 to 10.9). The estimated pitches of the animal vocalizations (Fig. 2-1f) were also derived using a 75 Hz floor and 5 kHz ceiling (Praat software).

Care must be taken in applying the HNR calculation. We derived HNR values over a two second duration, which proved to be adequate for relatively continuous or temporally homogeneous sounds. However, the HNR estimate was sensitive to abrupt acoustic transitions, such as fricatives and plosives, because it relies on providing a good estimate of the fundamental frequency of the sound sample (Boersma, 1993, Riede et al., 2005). We found that for some sound stimuli, and some sound categories such as sounds produced by hand tools, it was difficult to derive reliable HNR estimates, especially when using long (2 sec) duration sound samples. Thus, for many natural sound stimuli it may be more meaningful to examine shorter segments of time, characterizing discrete segments as the sound dynamically changes (Riede et al., 2001).

FMRI imaging paradigms

Each participant (n=16) performed one to five different scanning paradigms (41 scanning sessions total). In all paradigms we used a clustered acquisition design allowing sounds to be presented during scanner silence, and allowed a one-to-one correspondence between a stimulus presentation and a brain image acquisition (Edmister et al., 1999, Hall et al., 1999).

Paradigm #1. "Tonotopy" localizers. In one scanning session (12 scanning runs, ~8 min each; n=4 participants) we randomly presented 15 repetitions of 12 test sounds and 120 silent events as a control. The test sounds included six pure tones (PTs) at 125, 250, 2000, 4000, 12,000 and 16,000 Hz, plus six corresponding versions of band pass noise (BPN) stimuli having the same six center frequencies. The BPNs were generated from one white noise sample that was modified by 7^{th} order Butterworth filters to yield ±1 octave bandwidths (Fig. 2-1d). The sound intensity of the PT and BPN stimuli had been assessed psychophysically prior to scanning by three participants and equated for perceived loudness (Fig. 2-1e). All stimuli consisted of five 400 msec bursts with 35msec on/off ramps, spanning 2 sec duration.

For purposes of a task, a second PT or BPN (2 sec) was presented 200 msec after each respective PT or BPN test sound, having a lower, the same, or a higher center frequency. The task sounds spanned a gradient of roughly 3% difference at the lower and higher center frequencies and 0.5% difference at the middle center frequency ranges to match for approximate discrimination difficulty. During scanning, participants, with eyes closed, responded by three alternative forced choices (3AFC) as to whether the second sound was lower, the same, or higher in pitch, responding quickly before the second sound had stopped playing.

A multiple linear regression analysis modeled the contribution to the blood oxygen level-dependent (BOLD) signal time series data for each of the 6 PTs and 6 BPNs, plus and error term (also see "Image analysis" below). A winner-take-all algorithm identified voxels showing the greatest average BOLD signal magnitude responses, relative to silent events, to one of the three different frequency ranges presented— low (125+250 Hz, yellow), medium (2+4 kHz, orange), and high (12+16 kHz, red)—separately for the PT and BPN stimuli. We then masked the winner-take-all map for significant activation to the PT or BPN tonotopy data separately for each individual at two conservative threshold settings ($p<10^{-4}$ and $p<10^{-6}$), and projected these data onto the cortical surface models for each individual (see Image analysis). The surface models were then highly inflated and unfolded to facilitate viewing of the functional data (unfolded flat maps not shown), and these were used to guide the generation of outlines around tonotopic progressions (see Results for outlining criteria). For illustration purposes, individual cortical surface models

of the left and right hemisphere were slightly inflated, smoothed, and cut away so as to reveal each individual's unique cortical geography along Heschl's complex, including Heschl's gyrus (or gyri in some individual hemispheres), planum polare, and planum temporale.

Paradigm #2. IRN HNR paradigm. For the IRN paradigm we randomly presented 180 pairs of IRN stimuli and 60 silent events (6 runs, ~7 min each, n=16). The 60 IRN test stimuli included six pitches across ten 3 dB increments in HNR value, ranging from -3.6 to +25.2 dB (Fig. 2-1f). A second IRN "task" sound was presented 200 msec after the test sound, and included the above 60 sounds together with two additional IRNs at -6 dB HNR and one at +27. The test and task IRN sound pairs had the same pitch, but had either higher, the same, or lower HNR-value (ranging in difference from 0 to 5 dB HNR). Participants indicated whether the second sound was more tonal, the same, or more "noisy" than the first, responding (3AFC) before the second sound had stopped playing. A multiple linear regression analysis modeled the BOLD response using two terms plus an error term. The first term modeled variance in the time series data due to the presence of sound versus silent events. The second term assessed how much additional variance was accounted for by activity that linearly correlated with the HNR value of each sound (partial F-statistic). HNR-sensitive regions were selected based on the second term in the model. Individual data sets were thresholded to $p<0.01$, and whole-brain corrected for multiple comparisons using Monte Carlo randomization statistics (see Image analyses), yielding a whole-brain corrections of $\alpha<0.05$. The IRN HNR-sensitive ROIs were also separately modeled for sensitivity to the 6 different IRN pitches, using a second regression analysis similar to that described for paradigm #1.

Paradigm #3. Loudness biased IRN control paradigm. When assessed psychophysically in a sound isolation booth, the perceived loudness of the different IRN stimuli with differing pitches and HNR-values proved difficult to precisely balance across individuals. Because increases in sound intensity have generally been reported to activate larger and/or varying extents of auditory cortex (Jancke et al., 1998, Bilecen et al., 2002, Yetkin et al., 2004), a subset of participants (n=4) also underwent a separate scan to directly test the effects of sound intensity versus HNR value of the IRN stimuli. In one condition the 60 IRN stimuli (test and task sound pairs) were reverse-biased for

sound intensity (Supplemental Figure 2-5a), applying a linear gradient from -5dB to +5dB average RMS power to the lower to higher HNR valued IRN stimuli in steps of 3 dB HNR. In a second condition, the opposite forward-bias with intensity was applied. Scanning parameters and the listening task were identical to those for IRN paradigm #2 (sometimes conducted during the same scanning session as paradigm #2), and 6 runs of each condition were randomly intermixed (12 or 18 runs, ~7 min each). Multiple linear regression analyses modeled sensitivity to HNR value, as described in paradigm #2, for each of the separate loudness conditions.

Paradigm #4. Animal vocalization HNR paradigm. We randomly presented 160 unique animal vocalizations, 120 IRNs (the above described 60 IRNs, presented twice), and 40 silent events (7 runs, ~7 min each) using the same scanning parameters as those for paradigms #2-3 (n=11 of the 16 from paradigm #2, including the 4 participants from paradigm #1). For animal vocalization task sounds, the HNR values of the test sounds (original recordings) were modified by either adding white noise or by filtering out white noise (Cool Edit Pro v1.2 software). This allowed for the same 3AFC task as with the IRN paradigms, judging whether the task sound was more tonal, same, of noisier than the test sound. A multiple linear regression analysis included four terms: Two terms modeled variance due to the presence of vocalizations or IRN sounds versus silent events, respectively, while two additional terms assessed how much additional variance was accounted for by activity that linearly (positively or negatively) correlated with the HNR value of the vocalizations or IRN sounds. These latter two terms were used to generate HNR-sensitive ROIs, as described in paradigm #2.

A post-hoc non-linear regression analysis was additionally used to model the response profile between BOLD signal and HNR-value of the animal vocalizations (see Fig. 2-4, blue curves) using the equation:

$$BOLD = b_0 + g_0/(1 + g_1 * e(-g2 * \mathrm{HNR}))$$

Although the coefficients in this equation (b_0, g_0, g_1, g_2) do not necessarily reflect any physiologically relevant measures, this non-linear regression model was chosen as it could more closely fit the data (blue dots) and reflect biologically plausible floor and ceiling limits in BOLD signal "activation" levels than could a linear fit. This approach

also had the advantage of being able to reveal an HNR range where the slope might be changing more rapidly.

Paradigm #5. Human vocalization HNR paradigm. For this paradigm, we included unique samples of (a) 60 human speech phrases (balanced male and female speakers), (b) 60 human non-verbal vocalizations and utterances, (c) 60 animal vocalizations (a subset from paradigm #4), and (d) 60 IRNs (from paradigm #2), together with 60 silent events (8 runs, ~7 min each). Each sound category was matched for HNR value range (+3 to +27 dB HNR) and HNR mean (+11.6 dB HNR). Participants (n=6; five from paradigm #4) performed a 2AFC task, indicating whether the sound stimulus was produced by a human or not. A multiple linear regression analysis modeled the contribution to the BOLD signal from each of the four categories of sound, each relative to responses to silent events as the baseline control.

Stimulus Presentation

For all paradigms, the high fidelity sound stimuli were delivered via a Windows PC computer with a sound card interface (CDX01, Digital Audio), a sound mixer (1642VLZ pro mixer, Mackie Inc.) and MR compatible electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax LTD., Gardena, CA), worn under sound attenuating ear muffs. Sound stimuli were presented at 80-83 dBC-weighted, as assessed at the time of scanning (Brüel & Kjær 2239A sound meter) using one of the IRN stimuli (1 kHz pitch, 11.3 dB HNR) as a "standard" loudness test stimulus. The sound delivery system imparted a 75 Hz high pass filter (at rate of 18 dB/octave), and the ear buds exhibited a flat frequency response out to 20 kHz (±4 dB).

Image acquisition

Scanning was conducted with a 3 Tesla General Electric Horizon HD scanner equipped with a body gradient coil optimized to conduct whole-head, spiral imaging of BOLD signals (Glover and Law, 2001). For paradigms #2-5, a sound pair or silent event was presented every 10 seconds, and 4.4 sec after onset of the test sound there followed

the collection of BOLD signals from axial brain slices (28 spiral "in" and "out" images, with 1.87x1.87x2.00 mm^3 spatial resolution, TE = 36 msec, TR = 10 sec, 2.3 sec slice package, FOV = 24mm). The tonotopy localizer paradigms used a 12 second cycle to further minimize possible contamination of the sound frequencies emitted by the scanner itself. The presentation of each event was triggered by a TTL pulse from the MRI scanner. During every scanning session, T1-weighted anatomical MR images were collected using a spoiled GRASS pulse sequence; 1.2 mm slices, with 0.9375- x 0.9375-mm in-plane resolution.

Image analysis

Data were viewed and analyzed using AFNI (Cox, 1996) and related software plug-ins (http://afni.nimh.nih.gov/). For each paradigm, the scanning runs from a single session (6 to 18 scans) were concatenated into one time series. Brain volume images were motion corrected for global head translations and rotations, by re-registering them to the 20^{th} volume of the scan closest in time to the anatomical image acquisition. BOLD data of each participant were converted to percent signal changes relative to the mean of the responses to silent events on a voxel-wise basis for each scan run. Functional data (multiple regression coefficients) were thresholded based on partial F-statistic fits to the regression models, and significantly activated voxels were overlaid onto anatomical images.

Using the public domain software package Caret (http://brainmap.wustl.edu), three-dimension cortical surface models were constructed from the anatomical images for several individuals (Van Essen et al., 2001, Van Essen, 2003), onto which the volumetric fMRI data were projected. For all paradigms, the combination of individual voxel probability threshold (partial *F*-statistic, typically $p<0.01$ or $p<0.05$), a cluster size minimum (typically 9 or 50 voxels), and an estimate of signal variance correlation between neighboring voxels (filter width at half maximum of 2 to 4 mm) yielded the equivalent of a whole-brain corrected significance level of $\alpha<0.05$ (AFNI plug-in AlphaSim.

For group-average analyses, each individual's anatomical and functional brain maps were transformed into the standardized Talairach (AFNI-tlrc) coordinate space. Functional data were spatially low-pass filtered (4 mm Gaussian filter), then merged volumetrically by combining coefficient values for each interpolated voxel across all participants. Combined data sets were subjected to *t-tests* (typically $p<0.05$), and to a cluster size minimum (typically 9 voxels).

Averaged cortical hemisphere surface models were derived from three of our participants, using Caret software, on which the group-averaged fMRI results were illustrated. Briefly, six geographical landmarks, including the ridge of the STG, central sulcus, Sylvian fissure, the corpus callosum (defining dorsal and ventral wall divisions), and calcarine sulcus of each hemisphere of each participant were used to guide surface deformations to render averaged cortical surface models. Portions of these data can be viewed at http://sumsdb.wustl.edu/sums/directory.do?id=6694031&dir_name=LEWIS_JN09, which contains a database of surface-related data from other brain mapping studies.

RESULTS

The following progression of five experimental paradigms, using high spatial resolution fMRI (<2mm^3 voxels), was designed to test for HNR-sensitive patches of auditory cortex in humans, using both artificially constructed iterated rippled noises (IRNs) and real-world recordings of animal vocalizations. This included identifying tonotopically organized cortices and regions sensitive to human vocalizations within individuals, allowing for a direct test of our proposed hierarchical model for processing vocalizations. In order to explore the possible behavioral significance that the HNR signal attribute might generally have in vocal communication across species, we further investigated the harmonic content of various "sub-categories" of human and animal vocalizations to provide further context.

Estimated localizations of primary auditory cortices

Based on a cytoarchitectonic study (Rademacher et al., 2001), the location of primary auditory cortices (PAC; including A1, R, and possibly a 3rd subdivision), tends to overlap the medial two thirds of Heschl's gyrus (HG), though with considerable range in individual and hemispheric variability. Although the correspondence between functional estimates for PAC with histological and anatomical criteria remains to be resolved (Talavage et al., 2004), the identification of frequency-dependent response regions (FDRRs) allowed for more precise and direct localization of HNR-sensitive regions (addressed below) to tonotopically organized patches of auditory cortex within individual hemispheres. We identified the location of tonotopically organized cortices in a subset of our participants (n=4, paradigm #1), utilizing techniques similar to those described previously (Formisano et al., 2003).

We charted cortex sensitive to pure tones, and additionally to 1 octave band pass noises, at low (125 and 250 Hz, yellow in Fig. 2-2), medium (2,000 and 4000 Hz, orange), and high (12,000 and 16,000 Hz, red) center frequency ranges, wherein participants performed a three alternative forced choice (3AFC) tone or pitch

discrimination task. In contrast to previous fMRI tonotopy mapping studies (Wessinger et al., 2001, Schonwiesner et al., 2002, Formisano et al., 2003, Talavage et al., 2004, Langers et al., 2007), we derived perimeter boundary *outlines* of FDRRs based on the presence of tonotopic gradients at conservative threshold settings, as illustrated for three representative individuals who participated in three or more of our paradigms (Fig. 2-2a-c, black outlines; also see Methods). The tonotopic subdivisions of the FDRRs were characterized by cortex that responded preferentially, but not exclusively, to particular pure tone frequency bands (Fig. 2-2a, histograms). Three criteria were used to define FDRR outlines. First, a red to orange to yellow contiguous progression, in any direction, had to be present along the individual's cortical surface model using either the pure tone data or a combination of pure tone and band-pass noise data. However, outlines only encircled the high threshold ($p<10^{-6}$) pure tone data. Second, some of the FDRR progressions showed a mirror image organization with neighboring progressions, as reported previously in human and non-human animal studies (*ibid*). In those instances activation gradients were divided roughly midway between the two FDRRs (e.g. Fig 2-2a, left hemisphere midway along the yellow cortex). Third, FDRR progressions had to show continuity in both volumetric and surface projection maps to be included within an outline.

To our knowledge, this is the first fMRI study to chart the location of cortex sensitive to very high frequency tones (12,000 and 16,000 Hz, red). A right hemisphere bias for high frequencies (up to 14,000 Hz) and left hemisphere bias for low frequencies has been reported using auditory evoked potentials (Fujioka et al., 2002). However, the results of the present study demonstrated significant activation in both hemispheres to the high frequencies (red), which was even evident when examining responses to only the 16,000 Hz pure tones relative to silence (see Supplementary Fig. 2-2).

FDRR organizations defined by pure tones and band-pass noises were largely congruent with one another (Fig. 2-2a; upper vs. lower panels), although the band-pass noises generally activated a greater expanse of auditory cortex, which may include "belt" regions as reported previously in human (Wessinger et al., 2001) and non-human primates (Rauschecker et al., 1995). Note, however, that the functionally defined FDRR

outlines may not accurately reflect genuine boundaries between primary and non-primary areas since they were dependent on relative threshold settings (Hall, 2005). Nonetheless, we could reliably reveal one to three FDRRs located along Heschl's gyrus in each hemisphere of each participant, thereby refining estimated locations of PACs, and allowing for direct comparisons within individuals with the location of HNR-sensitive cortices, as addressed in the following section.

Iterated rippled noises reveal HNR-sensitive patches of cortex

Next, we investigated our hypothesis that portions of cortex outside of FDRRs would be characterized by activity that increased with increasing harmonic content (HNR value) of the sound stimuli, representing "intermediate" acoustic processing stages. We used IRN sounds because they could be systematically varied in HNR value, yet not be confounded by additional complex spectro-temporal signal attributes that are typically present in real-world sounds such as vocalizations. Sixty IRN stimuli were used, spanning 10 different ranges of HNR value for each of 6 different pitches (Fig. 2-1f, green). In contrast to previous studies using IRNs to study pitch depth or pitch salience (Griffiths et al., 1998, Hall et al., 2002, Patterson et al., 2002, Krumbholz et al., 2003, Penagos et al., 2004), we included a much broader range of effective HNR values (and pitches) that was more comparable to ranges observed with vocalization sounds. Participants heard sequential pairs of IRN stimuli and performed a 3AFC discrimination task indicating whether the second sound was more tonal, the same, or noisier than the first (n=16, paradigm #2; see Methods).

Relative to silent events, IRN stimuli activated a broad expanse of auditory cortex, including the FDRRs (not shown). More importantly, all sixteen participants revealed multiple foci in auditory cortex characterized by increasing activity that showed a significant positive, linear correlation with parametric increase in HNR-value of the IRN stimuli. All of the illustrated IRN HNR-sensitive regions-of-interest (ROIs) showed significantly greater, positive BOLD signal activation relative to silent events (e.g. Fig. 2-

3 error bars in charts). The topography of these regions was illustrated on cortical surface models generated for the same three individuals depicted previously (Fig. 2-3a-c, green).

In general, the IRN HNR-sensitive foci showed a patchy distribution along much of Heschl's complex, the superior temporal plane, and in some hemispheres included cortex extending out to the mSTG. Within individuals, some of these foci partially overlapped portions of the outlined FDRRs. In these regions of overlap, the tonotopically organized frequency sensitive ranges were sometimes congruent with the pitch range of the IRN (Supplementary Fig. 2-3), though the degree to which representations of periodicity pitch versus spectral pitch overlap remains a controversial issue outside the scope of the present study (Langner, 1992, Jones, 2006). Nonetheless, these results show, at high spatial resolution within individuals, that substantial portions of IRN HNR-sensitive regions were located along and just outside the FDRRs.

Group-averaged IRN HNR-sensitive regions were projected onto an averaged cortical surface model (Fig. 2-3d; see Methods). These results revealed a left hemisphere bias for IRN HNR-sensitive activation, evident as more significant and expansive areas (green and light green) involving portions of HG and cortex extending out to the mSTG. In contrast to previous studies that localized cortex sensitive to increasing pitch depth, or pitch strength, using rippled noises or other complex harmonic stimuli (Griffiths et al., 1998, Hall et al., 2002, Patterson et al., 2002, Penagos et al., 2004), the present results (i) indicated that the global HNR value of a sound represents a quantifiable acoustic signal attribute that is explicitly reflected in activation of human auditory cortex, (ii) demonstrated that IRN HNR-sensitive foci partially, but clearly did not completely overlap, with estimates of tonotopically organized cortices, suggestive of a hierarchical relationship, and (iii) showed that there was a left hemisphere lateralization bias for HNR-sensitivity, even though non-natural and relatively acoustically "simple" sound stimuli were used.

Control conditions for IRN pitch and loudness

As control measures, we explicitly examined IRN pitch and perceived loudness (intensity) as variables that might affect the cortical activation patterns (Bilecen et al., 2002). A secondary analysis restricted to the IRN HNR-sensitive regions-of-interest (ROIs) tested for linear correlations with increasing or decreasing IRN pitch sensitivity, and failed to show any significant correlations (Supplementary Fig. 2-4).

To directly assess the effects of parametric increases or decreases in IRN stimulus intensity, a subset of the participants (n=4, paradigm #3) were tested using IRN stimuli where the HNR-values were forward- or reverse-biased with intensity (Supplementary Fig. 2-5; see Methods). Both forward- and reverse-biased IRN sounds yielded positive, linearly correlated activation foci that overlapped one another, demonstrating that the identification of IRN HNR-sensitive regions was not simply due to unintended differences in perceived loudness of the IRNs with different HNR values.

Animal vocalizations also reveal HNR-sensitive cortices

Next, we investigated whether we could reveal HNR-sensitive regions using recordings of natural animal vocalizations, and, if they existed, whether they overlapped with IRN HNR-sensitive regions. One possibility was that there might be a single HNR-sensitive processing "module" that would show HNR-sensitivity independent of the type of sound presented. Alternatively, because animal vocalizations contain additional signal attributes statistically more similar to human vocalizations than to IRNs, HNR-sensitivity using vocalizations might reveal additional or different foci along "higher-level" stages of auditory cortex, such as mSTG (Lewis et al., 2005, Altmann et al., 2007). As in the previous paradigm, we employed a 3AFC harmonic discrimination task, and included IRNs and silent events as controls (n=11, paradigm #4, see Methods).

Relative to IRNs, animal vocalizations activated a wider expanse of auditory cortex, and with greater intensity, including near maximal BOLD signal responses within the FDRRs and IRN HNR-sensitive regions (see Fig. 2-4d IRN foci charts, and Fig. 2-5

histograms). Moreover, all participants revealed activation foci showing a significant positive, linear correlation with increase in HNR-value of the animal vocalizations (Fig. 2-4a-c, blue cortex). Similar to the IRN HNR-sensitive regions (green cortex), these foci also showed a patchy distribution. However, within individuals there was only a moderate degree of overlap between IRN and animal vocalization HNR-sensitive regions (blue-green intermediate color) at these threshold settings, despite similarity in the range of HNR values used. Most vocalization HNR-sensitive regions were located further peripheral (lateral and medial) to the FDRRs and IRN HNR-sensitive regions, including regions along the mSTG in both hemispheres. Response profiles for nearly all animal vocalization HNR-sensitive ROIs (Fig. 2-4a-c, charts) revealed at least a trend for also showing positive, linear correlations with the HNR value of the IRN sound stimuli. However, the IRNs were generally less effective at driving activity in these regions, which is evident in all the charts (green lines; also Fig. 2-5 histograms). In some hemispheres, the animal vocalization data points resembled more of a negative exponential or sigmoid-shaped response curve. Thus, in addition to linear fits we also modeled these data using an exponential function (see Methods), thereby constructing a more biologically plausible activation profile that respected floor and ceiling limits in BOLD signal (e.g. Fig 2-4a, right hemisphere; also see Supplementary Fig. 2-6 for additional individual charts).

Group-averaged data, similar to the individual data sets, demonstrated that the HNR-sensitive regions defined using animal vocalizations (Fig. 2-4d, blue), as opposed to using IRNs (green), were located further laterally, predominantly along the mSTG, with a strong left-lateralization. Moreover, animal vocalization HNR-sensitive regions in all participants showed greater response magnitudes than those defined using IRNs (Fig. 2-4d, charts). However, within the IRN HNR-sensitive ROIs (charts in green boxes) the linear correlations with animal vocalizations were relatively flat, appearing to have reached a ceiling plateau in both hemispheres. Within the animal vocalization HNR-sensitive ROIs (charts in blue boxes) both the IRN and animal vocalizations yielded positive, linear correlations with the HNR values, but the IRN data were of relatively lower response magnitudes and slightly less steep slopes, and thus tended to not meet statistical significance at our threshold settings.

In sum, these results revealed the existence of HNR-sensitive regions when using animal vocalizations and/or IRN stimuli, but the two respective activation patterns showed only a moderate degree of overlap (Fig. 2-4, blue vs. green). The extent of overlap appeared to be in part due to floor and ceiling effects with the BOLD signal, in that the animal vocalizations, regardless of HNR value, lead to near maximal activation (green boxed charts). However, other acoustic signal differences between vocalizations and IRNs are also likely to have contributed to the degree of overlap. Activation of the mSTG may have required sounds with more specific effective stimulus bandwidths, specific power spectral density distributions of different harmonic peaks (e.g. the $1/f^{\alpha}$-like power spectrum density in panels of Fig. 2-1a vs 2-1b; f= frequency, $1<\alpha<2$), and/or different specific frequencies, harmonics and sub-harmonics that are present in natural vocalizations but not IRNs (see Discussion). Nonetheless, these results demonstrated that there exists cortex, especially in the left hemisphere, that is generally sensitive to the degree of harmonic structure present in artificial sounds and real-world vocalizations.

HNR-sensitive regions lie between FDRRs and human voice-sensitive cortices

Our working model for HNR-sensitivity, as representing intermediate processing stages, assumes that portions of auditory cortex of adult human listeners are optimally organized to process the signal attributes characteristic of human vocalizations and speech. Thus, as a critical comparison, we localized cortex sensitive to human non-verbal vocalizations (Hvocs) and to human speech (Speech) in a subset of the participants (n=6, paradigm #5). In the same experimental session, we also presented animal vocalizations (Avocs), IRN stimuli, and silent events (see Methods). For this paradigm, all four sound categories (Speech, Hvocs, Avocs, and IRNs) had the same restricted range of HNR values (mean = +11.2, range +3 to +25 dB HNR), and participants indicated by 2AFC whether or not the sound was produced by a human (see Methods).

As expected, all four sound categories presented yielded significant activation throughout the FDRRs, IRN HNR-sensitive ROIs, and other portions of auditory cortex (not shown, though see group-average data in Fig. 2-5 histograms). More specifically, we

charted the locations of foci that showed differential activation to one versus another category of sound in relation to the previously charted FDRRs and HNR-sensitive regions (Fig. 2-5, colored cortical maps). In particular, regions sensitive to human non-verbal vocalizations (violet and pink) relative to animal vocalizations, speech (purple) relative to human non-verbal vocalizations, and regions preferential for animal vocalizations (light blue) relative to IRNs, were all superimposed onto averaged cortical surface maps. The HNR-sensitive regions (dark blue and green hues) and FDRRs (yellow and outlines) were those depicted previously (refer to Fig. 2-5 color key). Although the combined overlapping patterns of activation are complex, a clear progression of at least three tiers of activation was evident (Fig. 2-5a; rainbow colored arrows). FDRRs (yellow and outlines, derived from Fig. 2-2) represented the first tier, and were located mostly along the medial two thirds of Heschl's gyri, consistent with probabilistic locations for primary auditory cortices (Rademacher et al., 2001). FDRRs were surrounded by, and partially overlapped with, HNR-sensitive regions defined using IRNs (green), and those regions were flanked laterally by HNR-sensitive regions defined using animal vocalizations (dark blue). Together, these HNR-sensitive regions were tentatively regarded as encompassing a second tier, although they may be comprised of multiple processing stages.

Regions preferential for processing human vocalizations comprised a third tier, which included cortex extending into the STS. This included patches of cortex preferential for speech (purple) relative to human non-verbal vocalizations, which were strongly lateralized to the left STS, consistent with earlier studies (Zatorre et al., 1992, Belin et al., 2000, Binder et al., 2000, Scott and Wise, 2003), and patches of cortex preferential for human non-verbal vocalizations (pink) relative to animal vocalizations, which were lateralized to the right hemisphere, also consistent with earlier studies (Belin et al., 2000, Belin et al., 2002).

Within all ROIs representative of these three tiers (Fig. 2-5, color coded histograms), human vocalizations produced the greatest degree of activation, even within the IRN HNR-sensitive regions (green boxes). However, when progressing from IRN HNR-sensitive regions to animal vocalization HNR-sensitive regions to speech-sensitive regions, activation became significantly preferential for human vocalizations (e.g. purple and pink boxed histograms). This three tiered spatial progression was generally consistent

with proposed hierarchically organized pathways for processing conspecific vocalizations in both human (Binder et al., 2000, Davis and Johnsrude, 2003, Scott and Wise, 2003, Uppenkamp et al., 2006) and non-human primates (Rauschecker et al., 1995, Petkov et al., 2008), and with the identification of an auditory "what" stream for processing conspecific vocalizations and calls (Rauschecker et al., 1995, Wang, 2000).

Sub-categories of vocalizations fall along an HNR continuum

Do the global HNR values of human or non-human animal vocal communication sounds have any behavioral relevance? We further sought to determine whether our approach of exploring global HNR values could be useful for further characterizing different sub-categories of human and animal vocal communication sounds, concordant with ethological considerations in the evolution of vocal production (Wilden et al., 1998, Riede et al., 2005, Bass et al., 2008)

In addition to the vocalizations used in neuroimaging paradigms 4 and 5, we also derived HNR value ranges and means for several conceptually distinct sub-categories of human communication sounds (see Methods). Indeed, various sub-categories of vocalizations could be at least roughly organized along the HNR continuum (Fig. 2-6, colored ovals and boxes). In the lower HNR ranges this included hisses and a sub-category that included growls, grunts, and groans, most of which are vocalizations associated with threat warnings or negative emotional valence. Whispered speech, as a sub-category, was also characterized by relatively low HNR values, consistent with its social function as an acoustic signal with a low transmission range and reduced speech perceptibility (Cirillo, 2004). At the other extreme, vocal singing and whistling sounds (though not produced by vibrating tissue folds) were characterized by significantly higher HNR values than those typical for conversational speech. We also derived HNR values of spoken phrase segments from adults (n=10) when speaking in monologue to other adults versus when speaking to a realistic infant doll (Fig. 2-6, rectangles; see Methods). Interestingly, in addition to generally increasing in pitch, each participant's voice was characterized by significantly greater harmonic structure when speaking to an infant.

Also noteworthy was that the vocalization sub-categories tended to have onomatopoetic descriptors (in many languages), which when spoken stress phonemes that correlate with the HNR structure of the corresponding category of sound. For instance, we recorded phrases from multiple speakers and found the “ss” in “hissing” to be consistently lower in HNR value range than the “gr” in “growling”, which was lower than the “oo” in “mooing” (Fig. 2-6, top; see Methods). Moreover, onomatopoetic words (in Japanese) have previously been associated with activation of the bilateral (left>right) STG/STS (Hashimoto et al., 2006), overlapping blue to violet/purple regions in Figure 2-5. Together, these results suggest that variations of harmonic structure during vocal production, by animals or humans, can be used to convey fundamentally different types of behaviorally relevant information.

DISCUSSION

The main finding of the present study was that bilateral portions of Heschl's gyri and mSTG (left > right) showed significant increases in activation to parametric increases in overall harmonic structure of either artificially constructed IRNs and/or natural animal vocalizations. Within individuals, these HNR-sensitive foci were situated between functionally defined primary auditory cortices and regions preferential for human vocalizations in both hemispheres, but with a significant left-lateralization. We propose that the explicit processing of harmonic content serves as an important bottom-up, second-order signal attribute in a hierarchical model of auditory processing, which are comprised pathways optimized for extracting vocalizations. In particular, HNR-sensitive cortex may function as an integral component of computationally theorized spectro-temporal template staging, which serves as a basic neural mechanism for the segregation of acoustic events (Medvedev et al., 2002, Kumar et al., 2007). Thus, higher-order signal attributes, or primitives, that are characteristic of behaviorally relevant real-world sounds experienced by the listener may become encoded along intermediate processing stages leading to the formation of spectro-temporal templates, which dynamically develop to statistically reflect these acoustic structures. In the mature brain, matches between components of an incoming sound and these templates may subsequently convey information onto later processing stages to further group acoustic features, segment the sound, and ultimately lead to its identification, meaning or relevance.

However, why didn't the IRN and animal vocalization HNR-sensitive regions (i.e. Fig. 2-4; green vs blue foci) of auditory cortex completely overlap to indicate a single, centralized stage of HNR processing? Our results were consistent with previous neuroimaging studies manipulating pitch salience or temporal regularity of IRNs or complex tones (cf. Figs. 2-3 – 2-5; green), all of which revealed bilateral activation along lateral portions of Heschl's gyri and/or the STG (Griffiths et al., 1998, Patterson et al., 2002, Krumbholz et al., 2003, Penagos et al., 2004, Hall et al., 2005). HNR-sensitivity for animal vocalizations may not have overlapped the entire IRN HNR-sensitive region because other features of animal vocalizations, regardless of their HNR value,

contributed to the maximal or near maximal BOLD activation within both FDRRs and IRN HNR-sensitive locations (i.e. Fig. 2-5). As a result, animal vocalization HNR-sensitivity may not have been detectable. Conversely, IRN HNR-sensitive regions may not overlap animal vocalization HNR-sensitive regions due to serial hierarchical processing of acoustic features. IRNs, with relatively simple harmonic structure (equal power at every integer harmonic), appeared to be effectively driving early stages of frequency combination-sensitive processing. However, the IRNs were less capable of significantly driving subsequent stages along the mSTG, and thus were effectively filtered out from the pathways we identified for processing vocalizations. The other signal attributes required to drive higher stages (mSTG and STS) presumably include more specific combinations and distributions of power of harmonic and sub-harmonic frequencies that more closely reflect the statistical structure of components characteristic of vocalizations (Darwin, 1984, Shannon et al., 1995, Giraud et al., 2000). The series of acoustic paradigms that we employed at minimum serve to identify cortical regions for further study highlighting additional acoustic attributes. Although other higher-order signal attributes that would further test this model remain to be explored, the present data indicate that harmonic structure represents a major, quantifiable second-order attribute that can differentially drive intermediate processing stages of auditory cortex, consistent with a hierarchical stectro-temporal template model for sound processing.

The apparent hierarchical location of HNR-sensitive regions may be a corollary to the intermediate cortical stages of other sensory systems. For example, V2, V4 and TEO in human visual cortex (Kastner et al., 2000) and S2 in primate somatosensory cortex (Jiang et al., 1997) have “larger” and more complex receptive fields relative to their respective primary sensory areas, showing sensitivity to textures, shapes, and patterns leading to object segmentation. In all three modalities, these intermediate cortical stages may be integrating specific combinations (second-order features) of input energy across spatially organized maps corresponding to their respective sensory epithelia. In this regard, HNR-sensitive regions appear to represent cortical processing stages analogous to intermediate hierarchical stages in other sensory modalities, potentially reflecting a general processing mechanism of sensory cortex.

Cortical organization for processing different categories of real-world sounds

The present results supported and further extended our previous findings, in that the preferential activation of mSTG by animal vocalizations, compared to hand-tool sounds, was likely due to the greater degree of harmonic content in the vocalizations (Lewis et al., 2005, Lewis et al., 2006). Thus, HNR-sensitive stages could be facilitating the processing of vocalizations as a distinct category of real-world sound. However, an auditory evoked potential study examining responses to sounds representative of living objects (which included vocalizations) versus man-made objects, both of which were explicitly matched overall in HNR values, reported a differential processing component between the two categories starting ~70 msec from the onset of sound (Murray et al., 2006). Thus, it is clear that complex signal attributes other than global HNR value are contributing grossly to early stages of sound categorization. Nonetheless, HNR-sensitivity should be considered when exploring processing pathways for different categories of sound.

Human vocalizations, as a sub-category of sound distinct from animal vocalizations, are generally characterized by more idiosyncratic combinations of frequencies, specific relative power distributions, as well as other spectral and temporal attributes not taken into consideration here (Rosen, 1992, Shannon et al., 1995, Wilden et al., 1998, Belin et al., 2000, Cooke and Ellis, 2001, Belin et al., 2004). These other more subtle signal attribute differences appear to be necessary to evoke activation of the speech-sensitive regions we and others have observed along the STG/STS regions. Those regions are thought to represent subsequent hierarchical stages involved more with processing acoustic primitives or symbols just prior to extracting linguistic content (Binder et al., 1997, Cooke and Ellis, 2001, Scott and Wise, 2003, Price et al., 2005). Thus, the contributions of HNR relative to other higher-order signal attributes toward the processing of human vocalizations, as an apparently distinct sub-category of vocalizations, remains to be explored.

Relation of HNR-sensitivity to speech processing

Evidence for the presence of spectral templates in humans has significant implications for advancing our understanding how one may process and learn to recognize sounds, including speech. In early development, experience with behaviorally relevant vocalizations produced by one's caretakers, and perhaps one's own voice, could help establish the receptive fields of auditory neurons to exhibit sensitivity to their specific frequency combinations, thereby reflecting the statistical distributions of harmonic structure of human (*conspecific*) vocalizations. These experiences and subsequent cortical encodings will be unique to each individual's listening experience. Large cortical ensembles of frequency combination-sensitive neurons may thus develop (e.g. Fig. 2-4a-c HNR-sensitive patches unique to each individual) to comprise spectral and spectro-temporal templates, and these templates could serve as Bayesian-like networks to rapidly group or stream vocalizations from a person or sound-source (Medvedev et al., 2002, Kumar et al., 2007). As a side note, such principles have already been implemented in automated speech recognition algorithms, in the form of "weft-resynthesis" (Ellis, 1997), which may be an important biologically-inspired mechanism for the future development of hearing devices optimized for amplifying speech sounds.

On a larger scale of auditory cortex, and common across individuals, a hierarchical organization appears to become further established. In our data, sounds containing increasing degrees of acoustic structure, defined here as becoming more characteristic of human vocalizations, preferentially recruited cortex extending out to the mSTG and STS in both hemispheres (Fig. 2-5, rainbow colored progressions). However, the left hemisphere had more, and better organized, cortex devoted to HNR-sensitive processing, and also a stronger bias for processing human speech sounds (Binder et al., 2000, Boemio et al., 2005). Interestingly, at birth, humans are reported to already have a left hemisphere superiority for processing human linguistic stimuli (Pena et al., 2003). Thus, there may be a predisposition for the left hemisphere to process harmonic sounds, perhaps even being influenced by listening experiences *in utero*.

Interestingly, modifying one's voice to speak to infants, ostensibly to make them happy, was strongly associated with an increase in the harmonic structure of spoken

words and phrases (Fig. 2-6, rectangles). This largely appeared to be due to the elongation of vowel sounds, accompanied by a decrease in noise and other "complicated" acoustic features. Though speculative, this could serve as a socially interactive mechanism to help train the auditory system of a developing infant to recognize and perceive the basic statistical structure of human vocalizations. He or she would then eventually learn to process more complex variations in spectral, temporal, and spectro-temporal structure that convey more specific and behaviorally relevant meaning or communicative content, such as with phonemes, words, prosody, and other basic units of vocal communication and language.

In sum, although the HNR-value of a sound is by no means the only important acoustic signal attribute for processing real-world sounds, our results indicate that harmonic structure is parametrically reflected along human auditory cortical pathways for processing vocalizations. This attribute may serve as an integral component for hierarchical processing of sounds, notably including vocalizations as a distinct category of sound. Consequently, the HNR acoustic signal attribute should be considered when studying and distinguishing among neural pathways for processing and recognizing human vocalizations, auditory objects, and other "conceptually" distinct categories of real-world sounds.

ACKNOWLEDGMENTS

We thank Dr. Julie Brefczynski-Lewis for helpful comments on the text, Dr. Bill Shofner for Matlab code to construct IRN stimuli, Dr. Tobias Riede for suggestions on HNR calculations, and Doug Ward and Gerry Hobbs for assistance with statistical analyses. We thank Dr. Robert Cox for continual development of AFNI, and Dr. David Van Essen, Donna Hanlon, and John Harwell for continual development of cortical data analysis and presentation with CARET. This work was supported by the NCRR NIH COBRE grant RR015524 (to the Sensory Neuroscience Research Center of West Virginia University).

FIGURES AND TABLES

FIGURE 2-1

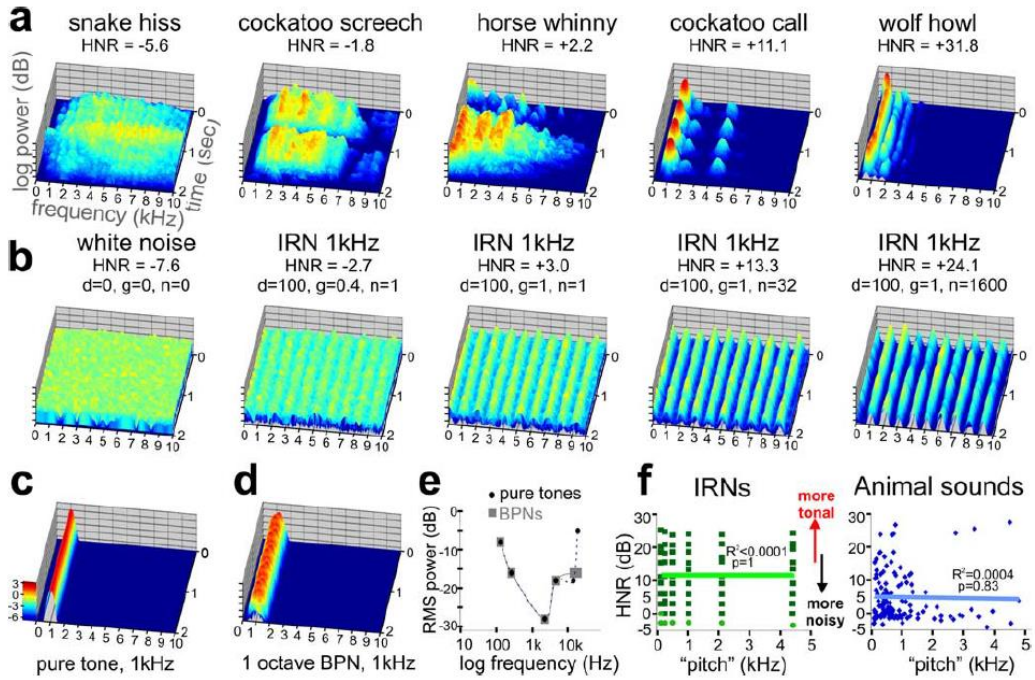


FIGURE 2-1. Sound stimulus attributes. a, 3D spectrograms of five vocalizations (2 sec duration), including one from a snake, two from birds, and two from mammals. In all plots, the frequency was limited to 10,000 Hz for illustration purposes, and the z-axis represents log-power (relative intensity, scale in panel **c** in log exponentials). The HNR value for each sound is indicated. **b**, Spectrograms of IRNs derived from one white noise sample (leftmost panel). The IRNs with greater HNR value correlate with more prominent frequency bands (peaks) at all harmonics (1 kHz in these examples), and had a more tonal quality. Spectrogram of an example (**c**) pure tone (PT) and (**d**) band pass noise (BPN) used for the frequency-dependent response regions (FDRR/tonotopy) localizer scans. Note the similarity of these peaks to those of the IRNs. **e**, Audiometric profile used to match perceived loudness of PT and BPN stimuli for the FDRR localizer scans. (**f**) Charts comparing "estimated pitch" versus HNR value of IRNs and animal vocalizations. Light green dots depict IRNs for which a pitch could not be accurately estimated computationally, though was determined by the IRN delay. There was no significant linear correlation between the pitch and HNR value for either stimulus set.

FIGURE 2-2

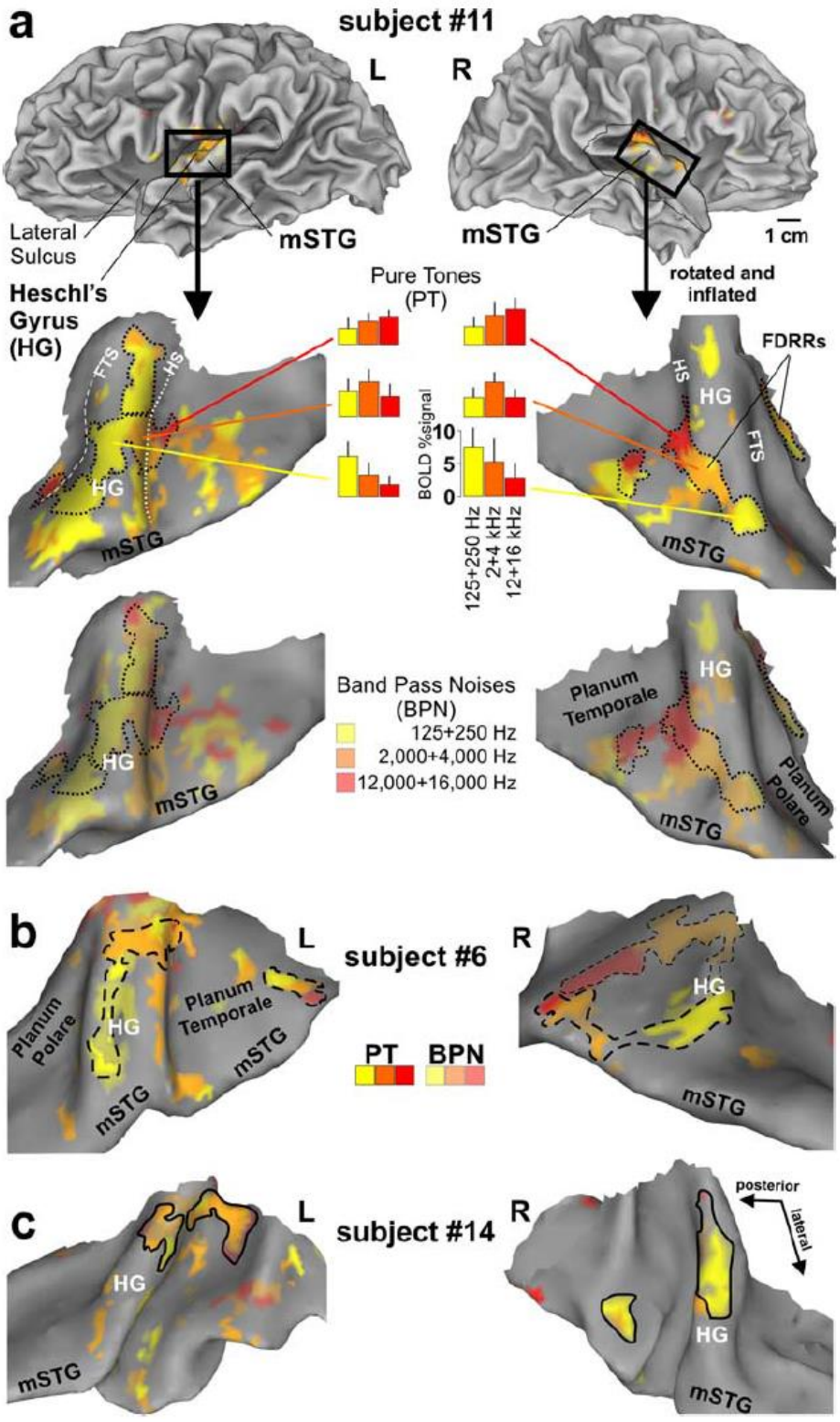

a
subject #11
L
R
Lateral Sulcus
Heschl's Gyrus (HG)
mSTG
mSTG
1 cm
rotated and inflated
Pure Tones (PT)
FTS
HS
HG
mSTG
BOLD %signal
10
5
0
125+250 Hz
2+4 kHz
12+16 kHz
HS
HG
FTS
FDRRs
mSTG
HG
mSTG
Band Pass Noises (BPN)
125+250 Hz
2,000+4,000 Hz
12,000+16,000 Hz
Planum Temporale
HG
Planum Polare
mSTG
b
L
subject #6
R
Planum Polare
HG
Planum Temporale
mSTG
mSTG
PT
BPN
HG
mSTG
c
L
subject #14
R
posterior
lateral
HG
mSTG
HG
mSTG

FIGURE 2-2. Functional localization of frequency-dependent response regions (FDRRs) in auditory cortex of three participants (a-c). Cortical hemisphere models of one participant (top panels) illustrate typical "cuts" (thin black outlines and black boxes) made to optimally view auditory cortex along the superior temporal plane and middle superior temporal gyri (mSTG) in this and subsequent figures. The cortical models of each hemisphere were slightly inflated and smoothed to facilitate viewing of Heschl's complex, including Heschl's gyrus (HG), Heschl's sulcus (HS; white dotted line), and the first transverse sulcus (FTS; white dashed line). The fainter dashed outline in panel **b** (right) depicts a prominent FDRR defined by the BPNs. The dotted, dashed, and solid black FDRR outlines distinguish these three representative individuals in this and subsequent figures. Refer to text for FDRR outlining criteria.

FIGURE 2-3

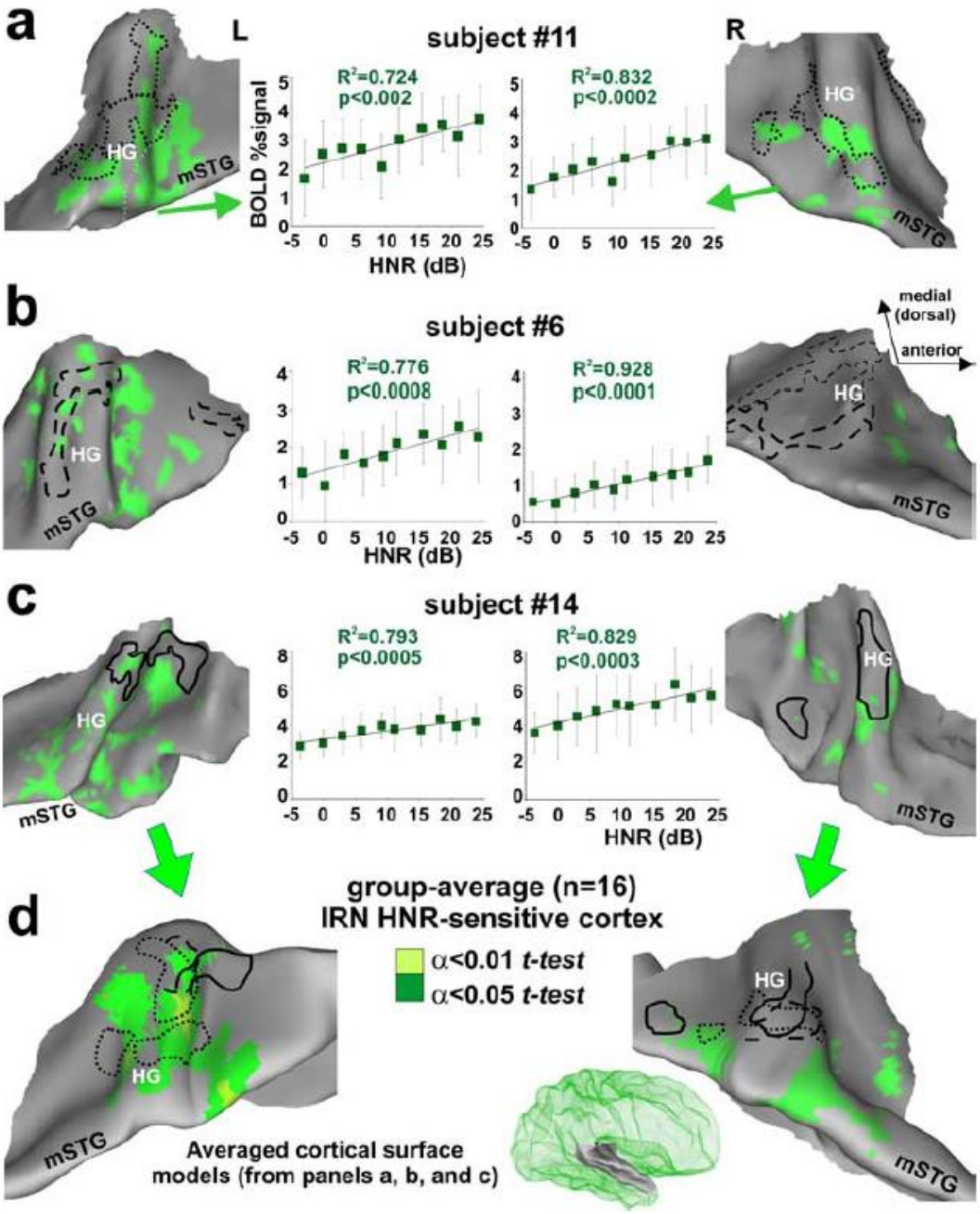

a
L
subject #11
R
HG
mSTG
BOLD %signal
R²=0.724
p<0.002
R²=0.832
p<0.0002
HNR (dB)
b
subject #6
medial (dorsal)
anterior
R²=0.776
p<0.0008
R²=0.928
p<0.0001
c
subject #14
R²=0.793
p<0.0005
R²=0.829
p<0.0003
d
group-average (n=16)
IRN HNR-sensitive cortex
α<0.01 t-test
α<0.05 t-test
Averaged cortical surface models (from panels a, b, and c)

FIGURE 2-3. Cortex sensitive to the degree of harmonic structure of iterated rippled noises (IRNs). a-c, Individual data sets showing location of IRN HNR-sensitive cortical foci ($\alpha<0.05$, corrected) relative to the location of FDRRs specific to each individual (dotted, dashed, and solid outlines from Fig. 2-2). Charts show the linear correlation between HNR value and BOLD activity (percent signal change relative to silent events) combined across the multiple foci along Heschl's complex and the mSTG (mean plus s.d.). The 180 IRN data points were binned at 3 dB HNR intervals for clarity. **d**, Group-average overlap of HNR-sensitive cortex after thresholding each individual data set (individual $\alpha<0.05$, and two *t-test* levels, $\alpha<0.05$ and $\alpha<0.01$, corrected) and projected onto averaged brain surface models derived from these three participants (right hemisphere model shown in green mesh inset).

FIGURE 2-4

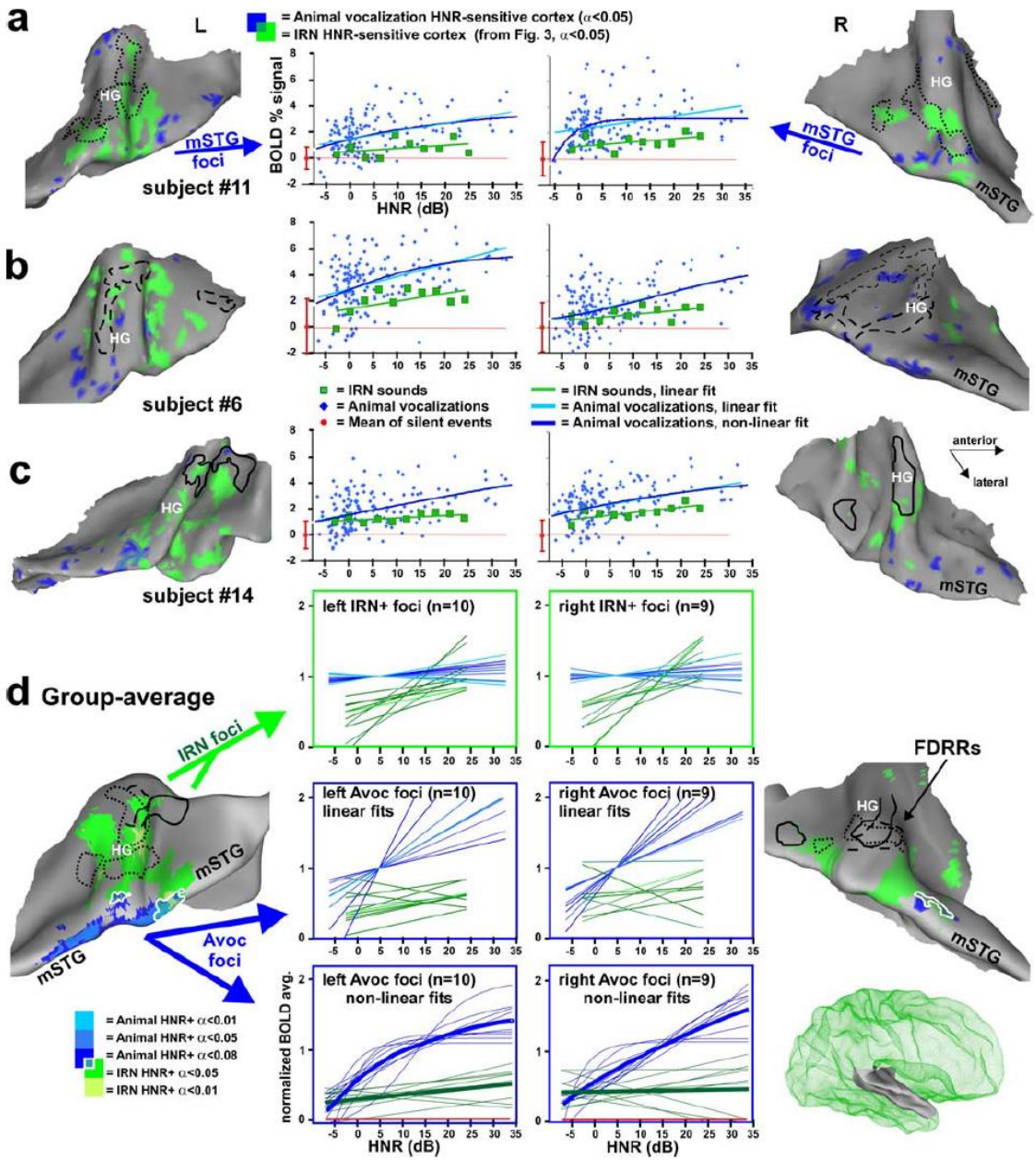

a
L
= Animal vocalization HNR-sensitive cortex (α<0.05)
= IRN HNR-sensitive cortex (from Fig. 3, α<0.05)
R
HG
mSTG
foci
subject #11
BOLD % signal
HNR (dB)
b
subject #6
= IRN sounds
= Animal vocalizations
= Mean of silent events
= IRN sounds, linear fit
= Animal vocalizations, linear fit
= Animal vocalizations, non-linear fit
c
subject #14
anterior
lateral
d Group-average
IRN foci
Avoc foci
left IRN+ foci (n=10)
right IRN+ foci (n=9)
left Avoc foci (n=10) linear fits
right Avoc foci (n=9) linear fits
left Avoc foci (n=10) non-linear fits
right Avoc foci (n=9) non-linear fits
normalized BOLD avg.
FDRRs
= Animal HNR+ α<0.01
= Animal HNR+ α<0.05
= Animal HNR+ α<0.08
= IRN HNR+ α<0.05
= IRN HNR+ α<0.01

FIGURE 2-4. Cortex sensitive to the degree of harmonic structure of animal vocalizations. a-c, Individual cortical maps illustrating animal vocalization HNR-sensitive cortex (blue), based on a linear regression model. IRN HNR-sensitive foci (green) and FDRR outlines (black) are from Fig. 2-3. Charts show the relation between HNR value and BOLD signal from the animal vocalization foci (blue) and IRN HNR-sensitive foci (green). The IRN data depicted in the charts were the control stimuli from paradigm #4 (as opposed to the data from paradigm #3 in Fig. 2-3), allowing for a direct comparison of relative activation response magnitudes (BOLD signal). All data are in percent BOLD signal change relative to the mean responses to silent events (red dot at zero, mean plus s.d.). **d**, Group-averaged maps of HNR-sensitive cortex to animal vocalizations (n=11, blue: *t-tests*, see color key) and to IRN stimuli (green, from Fig. 2-3d) on the averaged surface model from Fig. 2-3. White outlines encircle regions of overlap between IRN and animal vocalization HNR-sensitive regions. In the charts, thin curves are those from different individuals, normalized to the mean BOLD response within each ROI defined by the animal vocalization data. Not all participants showed significant bilateral activation (n=10 left, n=9 right hemisphere). Thick curves show the respective response averages. Some hemispheres revealed foci showing a significant *negative*, linear correlation with HNR value of the IRN and/or animal vocalizations (data not shown). When present, these foci were typically located along the medial wall of the lateral sulcus, and were more commonly observed in the right hemisphere. However, these negatively correlated HNR-sensitive foci were not significant in the group-averaged data.

FIGURE 2-5

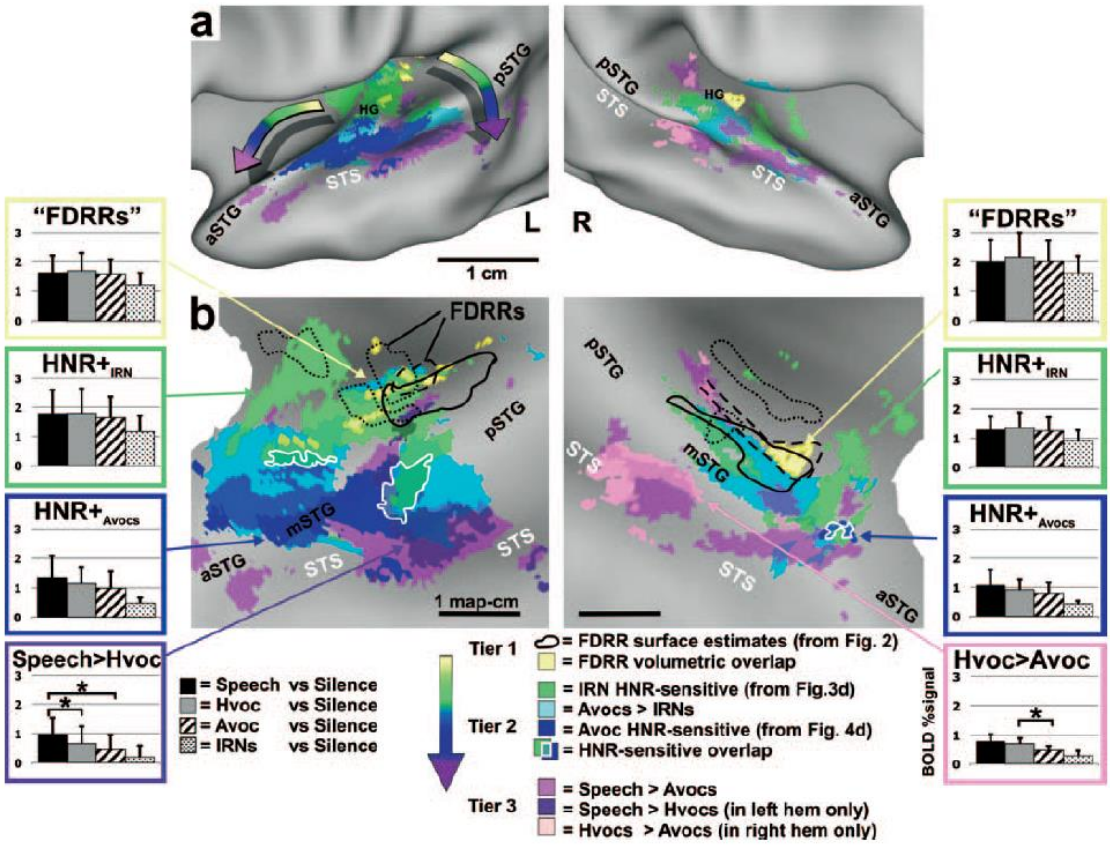


FIGURE 2-5. Location of HNR-sensitive cortices relative to human vocalization processing pathways and FDRRs. Data are illustrated on slightly inflated (**a**) and "flat map" (**b**) renderings of our averaged cortical surface models. Volumetric averages of FDRR (yellow) and volumetrically-aligned FDRR boundary outlines (black) were derived from data in Fig. 2-2. HNR-sensitive data are from Fig. 2-4d. Data from paradigm #5 (Speech, Hvoc, Avoc, IRN) are all at $\alpha<0.01$, corrected. Refer to key for color codes. Intermediate colors depict regions of overlap. The "rainbow" arrows in panel **a** depict two prominent progressions of processing tiers showing increasing specificity for the acoustic features present in human vocalizations. Overlap of IRN and animal vocalization HNR-sensitivity are indicated (white outlines). Histograms from several ROIs show group-averaged response magnitudes (mean plus s.d.) to each of the four sound categories used in paradigm #5 (refer to text for other details).

FIGURE 2-6

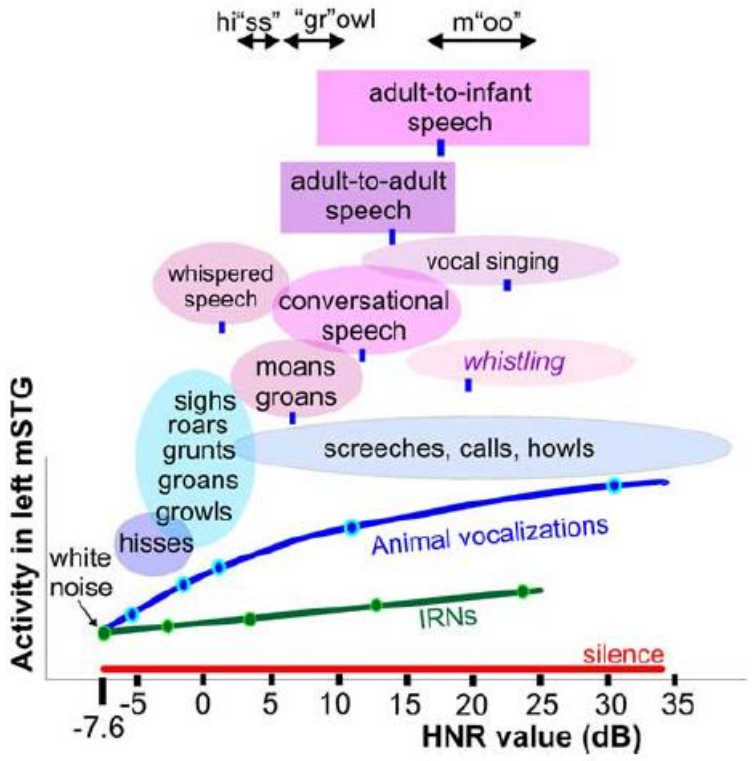
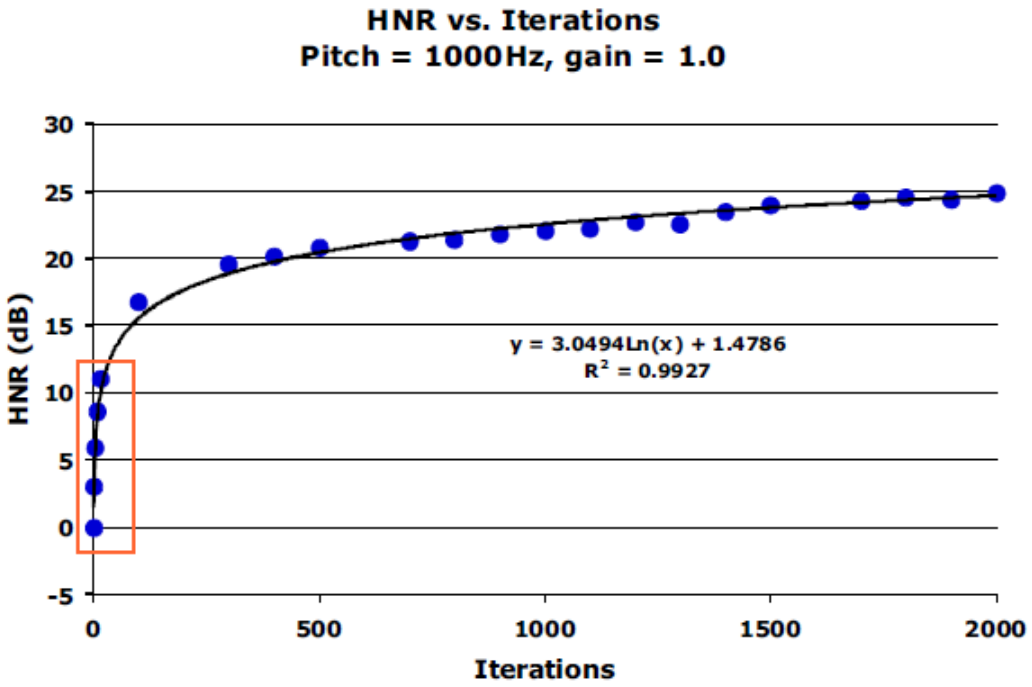


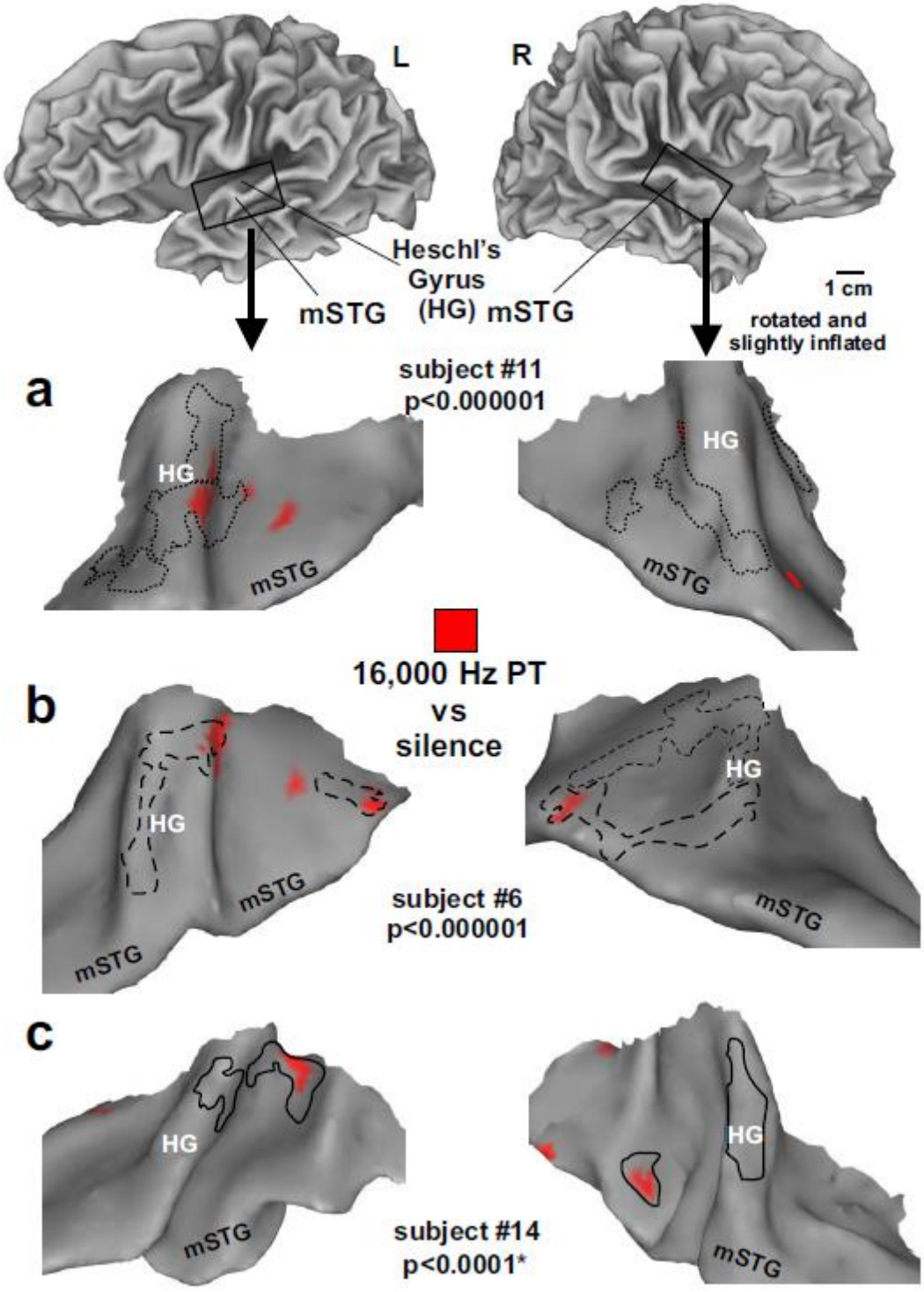
FIGURE 2-6. Typical HNR value ranges for various sub-categories of vocalizations. Oval and box widths depict the minimum to maximum HNR values of the sounds we sampled, charted relative to the group-averaged HNR-sensitive response profile of the left mSTG (from Fig. 2-4d). Green and blue dots correspond to sound stimuli illustrated in Fig. 2-1a-b. Blue ovals depict sub-categories of animal vocalizations explicitly tested in paradigm #4. Ovals and boxes with violet hues depict sub-categories of human vocalizations (12-18 samples per category), and blue tick marks indicate the mean HNR value. For instance, conversational speech, including phrases explicitly tested in paradigm #5, had a mean of +12 dB HNR, within a range from roughly +5 to +20 dB HNR. Adult-to-adult speech (purple box; mean = +17.2 dB HNR) and adult-to-infant speech (violet box; mean = +14.0 dB HNR) produced by the same individual speakers were significantly different (*t-test* $p<10^{-5}$). Stress phonemes of three spoken onomatopoetic words depicting different classes of vocalizations are also indicated. Refer to Methods for other details.

SUPPLEMENTARY FIGURE 2-1



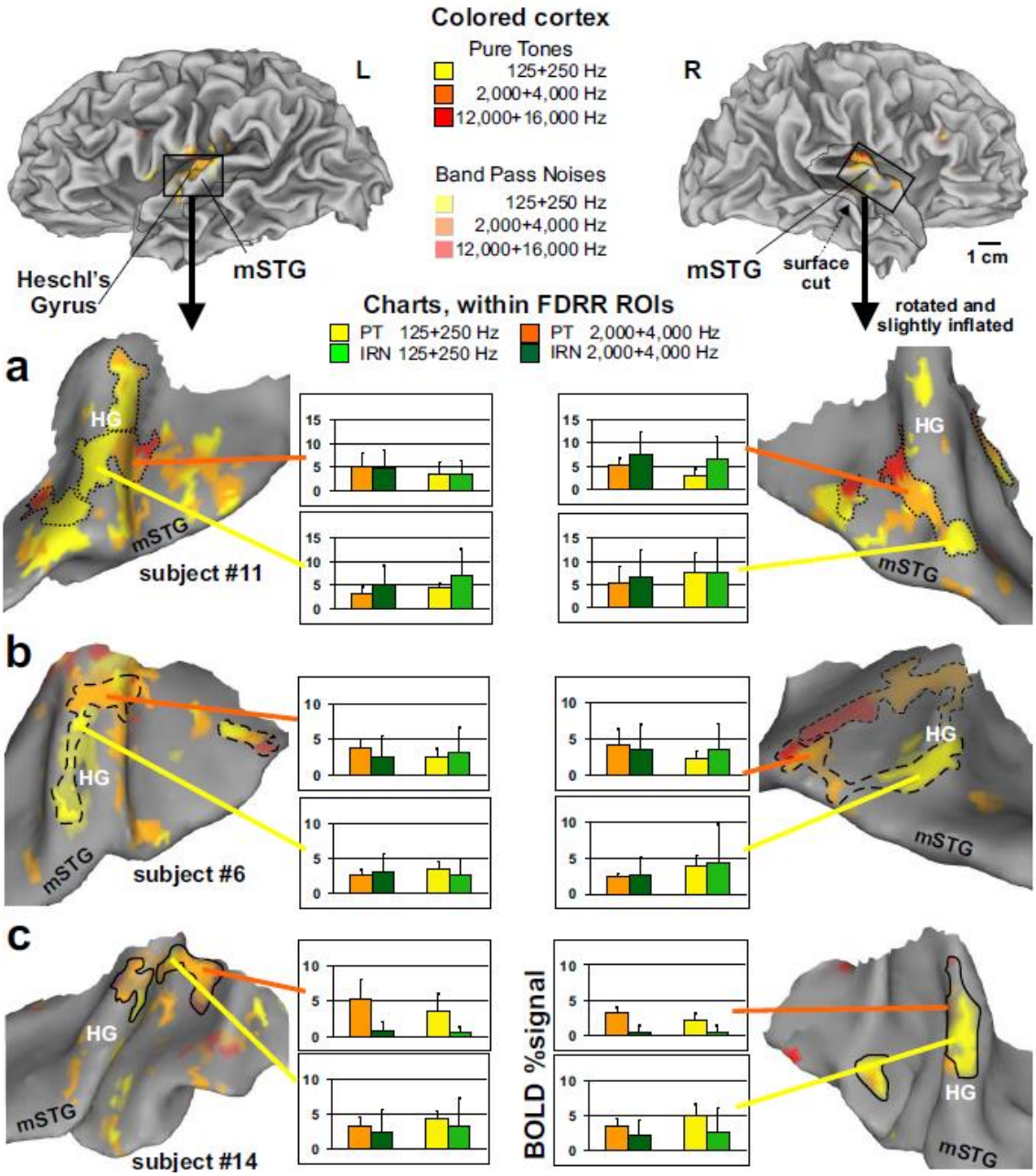
SUPPLEMENTARY FIGURE 2-1. Representative sample of constructed iterated rippled noises (IRN). HNR values of IRNs as a function of gain and number of iterations. IRNs were constructed so as to span HNR values from -6 to +25 dB HNR, being comparable to the observed range for animal vocalizations. Different gains had to be included (not shown) to construct a more complete set of IRNs with low HNR values. For comparison, the orange rectangles indicate the approximate range of IRNs reported by Griffiths et al., (1998), based on number of iterations.

SUPPLEMENTARY FIGURE 2-2



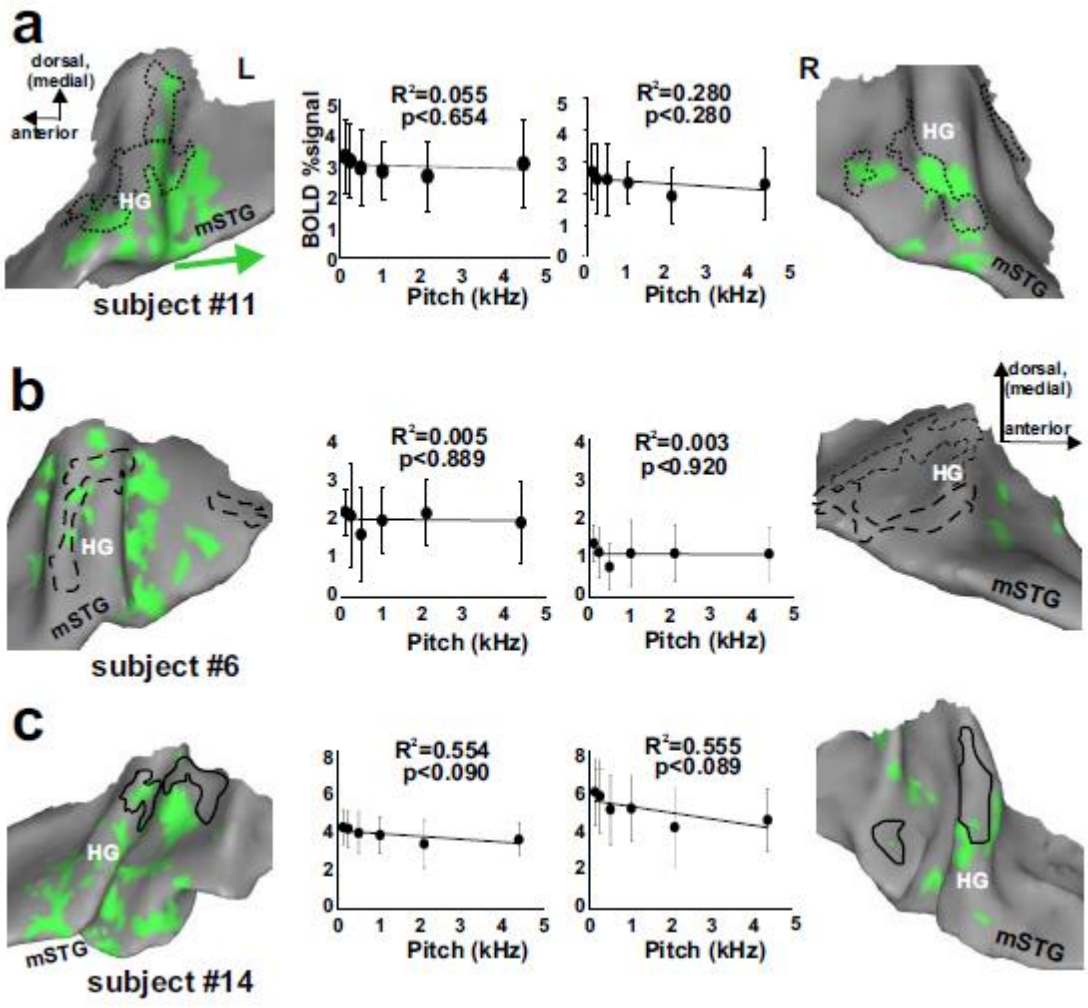
SUPPLEMENTARY FIGURE 2-2. Cortical activation to 16,000 Hz pure tone stimuli versus silent events (red) was present in both hemispheres in three participants (a-c) tested with this high frequency. Subject #14 (panel **c**) had difficulty hearing the 16 kHz pure tone, especially in one ear, as assessed by audiometry prior to scanning. Interestingly, for this individual we had to lower the threshold setting to $p<0.0001$ to reveal activation for this PT frequency, yet activation was still present in both hemispheres. Most of the 16 kHz sensitive regions overlapped, or were in close proximity to the high frequency ranges within the outlined FDRRs from Figure 2-2 (black outlines). Note that the differences in activation in Figure 2-2 (red) were due to the presence of the 12 kHz stimulus not represented here.

SUPPLEMENTARY FIGURE 2-3



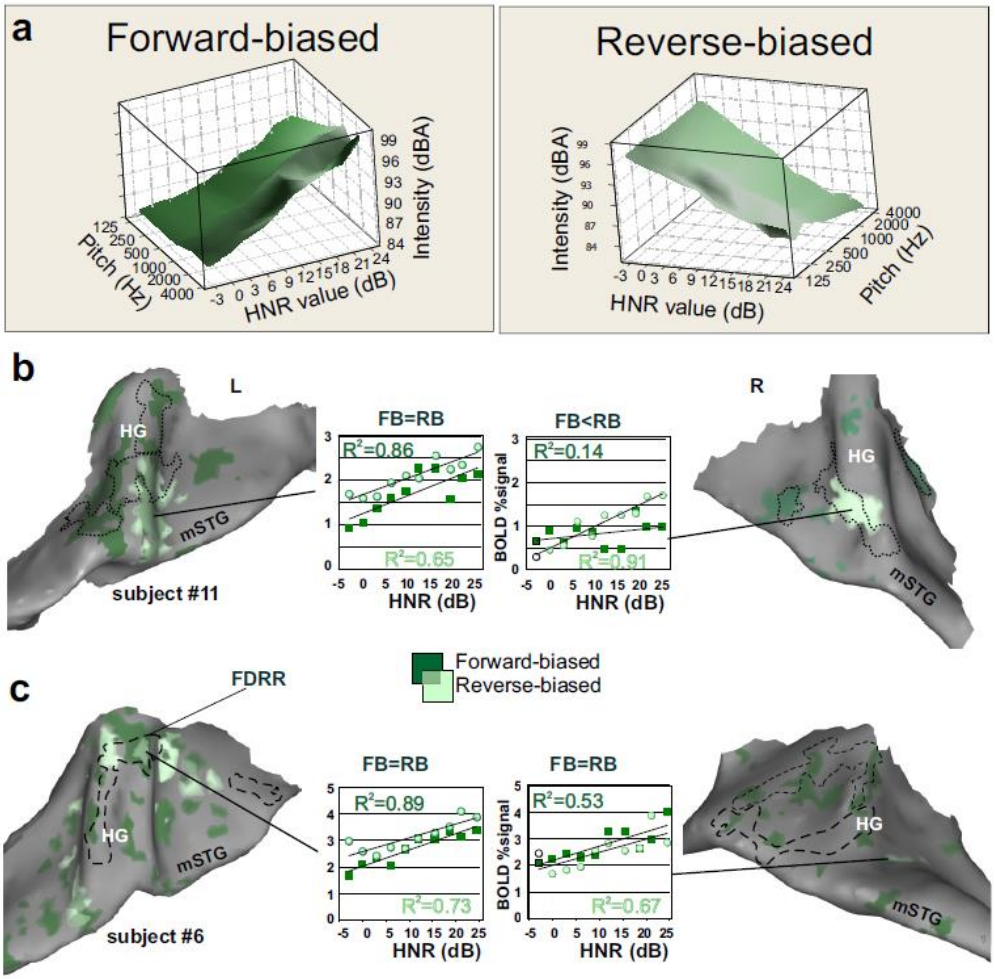
SUPPLEMENTARY FIGURE 2-3. Pure tone tonotopy versus IRN-tonotopy. Colored cortex and FDRR outlines from Figure 2-2. Charts depict activation to pure tones (yellow and orange) and IRNs of the corresponding periodicity pitch (light and dark green). Note that many, but not all, of the FDRR ROIs showed the same frequency-preference trend (e.g. in charts: orange > yellow and dark green > light green). The IRN pitch maps were not as evident as the tonotopic maps, presumably due to the fact that our IRNs contain power at all harmonics. This may also be due to not being able to use IRN pitches above roughly 5 kHz, they surpassed the psychophysical upper limit of musical pitch perception (Langner, 1992, Rosen, 1992), and they became distorted when presented through our sound delivery system.

SUPPLEMENTARY FIGURE 2-4



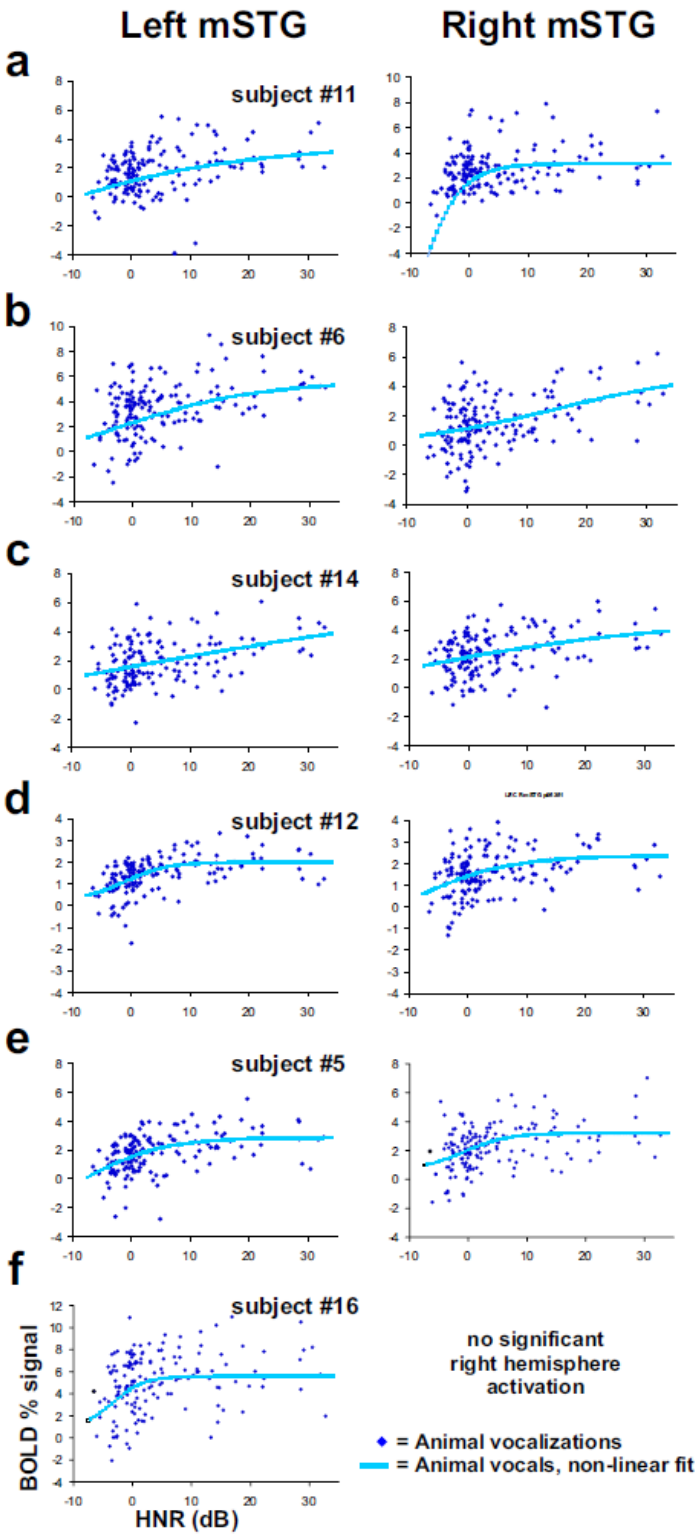
SUPPLEMENTARY FIGURE 2-4. Sensitivity to IRN pitch. Within the IRN HNR-sensitive regions (green) there were no significant correlations in activation with the periodicity pitch of the IRN (charts). Also see Figure 2-1 for IRN pitch range.

SUPPLEMENTARY FIGURE 2-5



SUPPLEMENTARY FIGURE 2-5. Intensity vs HNR-sensitivity to IRNs. In contrast to the PT and BPN stimuli, which could be equated for perceived loudness across individuals (e.g. Fig. 2-1e), the IRN stimuli proved to be more difficult to match perceptually. Thus, participants (n=4) were tested to directly assess the effects of parametric increases or decreases in IRN stimulus intensity. **a,** Graphical depiction of the 60 IRN stimuli showing a forward- or reversed-bias with stimulus intensity (total RMS power). **b-c**, Individual data sets (2 of 4 shown) illustrating activity under the two separate conditions performed during the same scanning session: dark green = HNR biased with loudness, light green = HNR biased against loudness. Both conditions revealed cortex sensitive to the HNR value of the IRNs.

SUPPLEMENTARY FIGURE 2-6



SUPPLEMENTARY FIGURE 2-6. Response profiles (BOLD percent signal change) for regions showing HNR-sensitivity to animal vocalizations in 6 participants. Charts depict the sigmoid-response fit for the mSTG ROIs.

CHAPTER 3:

Late auditory evoked potentials exhibiting sensitivity to harmonic signal content

William J. Talkington, Brandon D. Smith, Stephanie K. Khoo,
Christopher A. Frum, James W. Lewis

Center for Neuroscience,
Center for Advanced Imaging in the Department of Radiology,
Department of Neurobiology and Anatomy
West Virginia University, Morgantown, WV 26506

This manuscript was submitted for review to the European Journal of Neuroscience on December 19th, 2012.

ABSTRACT

Communicative vocalizations of most mammals are typically characterized by strong harmonic content. Using functional magnetic resonance imaging (fMRI) we previously reported that the harmonics-to-noise ratio (HNR) value of a sound, whether naturally produced or artificially constructed, represents an acoustic signal attribute to which early cortical stages of the human auditory system show parametric sensitivity. However, the temporal processing dynamics of HNR as an acoustic signal attribute in a range that is typical of ethologically-relevant sounds remained unknown. In the present study we recorded cortical auditory evoked potentials (AEP) in response to artificially constructed iterated ripple noise (IRN) sounds that parametrically spanned an ethologically-relevant range of HNR values. The N1-P2 AEP complex, shown to be sensitive to speech and speech-like sounds, generally demonstrated a positive and monotonically increasing response to HNR value (-3 to +24 dB HNR). Somewhat surprisingly, however, low HNR value ranges showed a decrease in AEP responses (from white noise (-7.6dB HNR) to -3 dB HNR). Moreover, this biphasic response profile persisted even when testing IRN sounds that were reverse biased with intensity (perceived loudness). Together with our previous fMRI findings, these results provide converging neuroimaging evidence that early auditory cortices in humans contain a processing stage involved in signal feature detection of harmonic content – a characteristic attribute of many communicative vocalizations and utterances.

INTRODUCTION

Rapid segregation of vocalizations from complex and noisy auditory scenes is critical for effective communication. This skill likely relies upon neuronal pathways optimized for extracting and analyzing acoustic signal attributes characteristic of vocalizations (Bregman, 1990, Billings et al., 2011). Vocal cords and articulatory structures produce sounds predominantly by vibrating columns of air – a physical arrangement that generates strong harmonic content as well as other idiosyncratic combinations of non-linear acoustic components (Fitch et al., 2002, Lewis et al., 2005). Harmonic signal content represents a prominent low-level spectral feature of natural vocalization sounds across numerous species (Riede et al., 2001, Lewis et al., 2009) that is quantifiable with a harmonics-to-noise ratio (HNR) (Boersma, 1993).

We have shown that different categories of animal and human vocalizations are separable along a continuum of HNR values (Lewis et al., 2009). Using functional magnetic resonance imaging (fMRI), we further reported regions of auditory cortex along the middle superior temporal gyri (mSTG) that were parametrically sensitive to the HNR values of animal vocalizations. Those findings suggested that harmonic attributes represent bottom-up signal features that may be critical to rapidly process communicative utterances. Studying the neuronal representations of vocalization acoustics is often hindered by their spectrotemporal complexity. Thus, we and others have used acoustically simpler sounds, such as iterated rippled noise (IRN), to elucidate the functions of auditory circuits.

IRN is artificially-produced sound created by subjecting a sample of broadband Gaussian-distributed noise to an iterative delay-and-add process (Yost, 1996a). The perceived pitch strength or depth of IRN has been shown to vary with the number of iterations and the gain applied during each iterative cycle (Yost, 1996b). The IRN pitch percept has been used extensively to test and refine pitch processing models in psychophysical and neuroimaging settings (fMRI, electroencephalography (EEG), and magnetoencephalography (MEG)) (Patterson et al., 1996, Yost, 1996b, Griffiths et al., 2001, Patterson et al., 2002, Krumbholz et al., 2003, Hall et al., 2005, Soeta et al., 2005,

Jones, 2006, Hall and Plack, 2007). Our earlier fMRI study also utilized IRNs (Lewis et al., 2009); however, these IRN stimuli were not used to study pitch-processing, but rather were used to systematically explore HNR as a quantifiable, cortically-represented spectral signal attribute that may be crucial to vocalization processing.

No studies, to our knowledge, have used electrophysiological measures to systematically examine cortical responses to IRNs that span the spectrum of HNR values found in *behaviorally-relevant* communicative vocalizations and utterances (Lewis et al., 2009). We chose to examine the effects of IRN HNR values on the auditory N1-P2 complex, a pair of late auditory evoked potential (LAEP) components thought to be generated near primary auditory cortices (PACs) (Näätänen and Picton, 1987, Tremblay et al., 2001, Jaaskelainen et al., 2004, Martin et al., 2008, Picton, 2011). The magnitude of the N1-P2 complex is thought to reflect the expansiveness and synchrony of cortex responding to a stimulus. We hypothesized that maximal N1-P2 amplitudes would be centered on IRN HNR values characteristic of conversational human speech (approximately 6-15dB HNR) (Lewis et al., 2009).

MATERIALS AND METHODS

This study was comprised of two separate experiments designed to examine electrophysiological auditory evoked potentials, specifically the N1-P2 complex (Tremblay et al., 2001, Martin et al., 2008), to parametric changes of the harmonic content (measured with Harmonics-to-Noise Ratio, HNR) in artificial iterated rippled noise (IRN) stimuli (Expt. 1) and whether these responses were insensitive to deliberate intensity-biasing (Expt. 2).

Participants

Native English-speaking and right-handed healthy adults (n=16) participated in the following two experiments. Sixteen subjects participated in Experiment 1; however, one subject in Expt. 1 was eliminated from group averaging and analyses for excessive artifacts (Expt. 1: n = 15/16, mean age = 26.2 years, SD = 4.97 years, 7 female). Fifteen of the subjects from Expt. 1 participated in Expt. 2 (n = 15, mean age = 26.2 years, SD = 5.06 years, 8 female); one subject failed to return and participate in Expt. 2. They all self-reported normal hearing and no history of audiological or neurological disorders. Research protocols were approved by the West Virginia University Institutional Review Board and in conformance with the Declaration of Helsinki. Each subject provided informed consent after receiving explanations of all experimental procedures and received a stipend.

Stimuli

Iterated Rippled Noise (IRN) stimuli were created with unfiltered Gaussian noise and a custom MatLab (MathWorks, Natick, MA; William Shofner, personal communication) script through the "IRNO" process with iteration, gain, and delay (n, g, d) as design variables(Yost, 1996a). The delay value was set to 2ms (producing a 500Hz pitch) and each IRN stimulus was adjusted to 200ms in duration with 2ms linear on/off ramps. The harmonic content of IRN stimuli was quantified with a Harmonics-to-Noise Ratio (HNR)

calculation using the freely available software Praat (http://www.fon.hum.uva.nl/praat/; Time step: 0.01s; Minimum pitch: 75Hz; Silence threshold: 0.1; Periods per window: 1.0)(Boersma, 1993). HNR values allow one to contrast the relative strengths of harmonic versus noisy signal components in a wide variety of acoustic stimuli, including complex sounds like vocalizations (Lewis et al., 2009). The number of iterations (i) and gain (g) values were varied to generate IRNs that had approximate HNR values of -7.6 (white noise; i=0, g=N/A), -3 (i=1, g=0.4), +3 (i=2, g=1.0), +9 (i=0.9, g=16), +15 (i=64, g=1.0), and +24dB (i=1100, g=1.0). White noise (-7.6dB) represents the lower limit of the HNR value calculation; the other values were chosen to span an HNR range that effectively encompassed HNR-values derived from behaviorally-relevant vocal communication categories (cf. Fig. 2-6 of Chapter 2).

Electrophysiology procedures common to all experiments

Electroencephalographic (EEG) recordings were collected with Neuroscan SynAmps hardware and Scan 4.3 Acquire software using 21-channel Quik-Caps (Ag/Ag-Cl sintered electrodes; 10-10 system). Data from each channel were sampled at 1 kHz and filtered on-line from 0.1-100Hz. All experimental sessions consisted of six EEG recording runs that lasted approximately seven minutes. Each run contained 408 IRN stimulus trials (68 stimuli per HNR value); stimulus onsets were separated by a random uniformly-distributed inter-stimulus interval (ISI) from 900-1100ms to minimize timing-based habituation effects. During each run, participants watched silent and subtitled films to divert their attention from the IRN stimuli (Pettigrew et al., 2004). IRN stimuli were delivered to the right ear of each subject using electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax LTD., Gardena, CA) with a Windows PC installed with Presentation software (version 11.1, Neurobehavioral Systems, Inc.) and a CDX01 Digital Audio sound card interface. Audio output loudness was adjusted to the comfort level of each subject (70-80dB) while still retaining clear sound percepts of all stimuli, regardless of HNR value or loudness bias.

Experiments 1 and 2: HNR-dependent auditory evoked potentials

Experiment 1 investigated HNR-dependent N1-P2 amplitude responses using the IRN stimuli described above that were equated to one another for overall stimulus intensity (dB RMS power, Root Mean Square). Experiment 1 will be interchangeably referred to as the "iso-intensity" condition.

Experiment 2 was designed as a critical control paradigm for Experiment 1 to assess any potential effects of IRN intensity upon N1-P2 amplitudes or latencies. IRNs with greater iteration numbers are perceived to be louder (Soeta et al., 2007); thus, this experiment aimed to control for any inadvertent perceptual salience biases related to stimuli loudness. This experiment was designed to mimic one of our previous fMRI control paradigms (reproduced here in Fig. 3-1) (Lewis et al., 2009). For Experiment 2, the IRN stimuli were modified to linearly decrease in overall intensity with increasing HNR values. This modification ensured that IRNs with progressively higher HNR values were physically and perceptually quieter than those with lower HNR values. Specifically, the relative intensity of the 3dB HNR IRN was held constant and a -0.75dB RMS/dB HNR intensity ramping-function was applied to the other stimuli. In this paradigm, white noise stimuli (-7.6dB HNR) were physically the most intense and perceived to be the loudest, +24dB HNR IRNs were perceived to be the quietest. A cohort of five individuals not participating in the EEG portion of the study listened to the intensity-biased IRN stimuli in all pair-wise combinations and made loudness judgments. Each subject appropriately rated the sounds in a manner that reflected our biasing function.

Data analysis

All analyses were performed in the MatLab environment with open-source EEG-analysis software EEGLAB (ver. 10.2.5.8) (Delorme and Makeig, 2004) and an associated plugin ERPLAB (www.erpinfo.org). Continuous EEG data for each subject were initially combined across runs and high-pass filtered (0.1 Hz). Data were low-pass filtered at 30Hz for plotting purposes only. Individual event epochs were defined by 700ms windows with 200ms pre-stimulus and 500msec post-stimulus periods. Epochs

were extracted from the EEG data for each IRN event type (-7.6 (white noise), -3, +3, +9, +15, +24dB) and baseline corrected with responses from the 200ms pre-stimulus periods. Epochs were rejected with a moving window peak-to-peak artifact detection function in ERPLAB; the entire epoch was included in artifact rejection, the voltage threshold was 100 μV, the moving window width was 200ms, and the window step size was 50ms. N1 and P2 amplitudes for each subject were measured and calculated in responses to different HNR-valued IRNs; the AEPs were averaged for each subject across the fronto-central electrodes Fz, F3, and F4 (Fig. 3-2). N1 responses were defined as the mean potential value between 85 and 135ms and P2 response between 150 and 200ms (Picton, 2011); N1-P2 amplitudes were the differences between these two values for each condition. N1-P2 amplitudes for each condition (all HNR values) were entered into a one-way repeated measured ANOVA that included a within-subjects factor of HNR (six aforementioned values); a separate ANOVA was performed for each experiment (sections 3.1 and 3.2). Post-hoc comparisons between N1-P2 amplitude means for different HNR values were corrected for multiple comparisons with Bonferroni corrections. Both ANOVAs had an alpha level of 0.05 and the Greenhouse-Geisser epsilon correction was used when sphericity could not be assumed (Jennings and Wood, 1976).

RESULTS

Two experimental paradigms, utilizing AEPs, were conducted by presenting listeners with IRNs that systematically varied in harmonic content. Experiment 1 was designed to determine if there is an HNR-dependent response profile present in the auditory N1-P2 AEP complex. In conjunction with the first experiment, Experiment 2 was designed to control for possible systematic biases produced by overall signal intensities (perceivable loudness differences) of IRNs that might have influenced HNR-dependent response profiles in auditory cortex.

Experiment 1: HNR-dependence of the auditory N1-P2 complex

Figure 3-2 shows the group-averaged waveform morphology and scalp topography for all of the HNR conditions in Experiment 1 (iso-intensity). This topography was similar to those of all other conditions in both experiments and is consistent with a stereotypical auditory N1-P2 complex response, with the greatest amplitudes occurring at fronto-central electrodes (Vaughan and Ritter, 1970, Tremblay et al., 2001). Figure 3-3A displays group-averaged (n=15) evoked potentials to IRN stimuli that were equally intense in RMS power but differed in their harmonic content (HNR). A main effect of HNR was seen on N1-P2 amplitudes (Table 3-1; $F_{5,70}$ = 63.574, P = 0.000), supporting our hypothesis that IRN stimuli with greater HNR values would generally evoke stronger N1-P2 responses. More specifically, however, the profile did not show the greatest amplitudes in the HNR range reflective of most human vocalizations (approximately 6-15dB HNR) as we had predicted. Rather, a monotonically increasing amplitude trend was seen between HNR values between approximately -3 and +24 dB; an inverse relationship was seen between -7.6 and -3dB. Pairwise comparisons (Table 3-2) between HNR conditions indeed revealed that the N1-P2 values produced by white noise (-7.6dB HNR) were significantly higher than those produced by -3dB HNR IRNs (P = 0.024). AEP responses to white noise were indistinguishable from those produced by +3dB HNR IRNs (P=1.0). We reasoned that the overall monotonically-increasing N1-P2 amplitude

trend could have been related to the perceived loudness of the IRN sounds at different HNR ranges, which was addressed with Experiment 2.

Experiment 2: Loudness bias control

Our earlier fMRI study examined the effects of forward- and reverse-biasing IRN intensities as functions of HNR values upon blood oxygen level dependent (BOLD) responses (Lewis et al., 2009). We found that neighboring or overlapping regions of auditory cortex (near or within HG) on the order of 2-4mm along the cortical surface were parametrically sensitive to HNR values regardless of intensity biases (Fig. 3-1, dark and light green hues). Experiment 2 of the present study involved a similar modification of the IRN stimuli used in Experiment 1 by reverse-biasing their intensities with increasing HNR values. The intensity of the +3dB HNR IRN stimulus was held constant and the other five IRN stimuli were intensity adjusted by a linear -0.75dB RMS/dB HNR function, thus making the white noise and +24 dB stimuli the most and least intense stimuli, respectively. A cohort of participants (n=5) not involved with the electrophysiological experiments perceptually rated each intensity-biased sound in pair-wise combinations between all stimuli. All of these subjects perceived noticeable loudness decreases with each increasing HNR valued stimulus, thereby corroborating our intensity modifications.

The results of Experiment 2 revealed some effects of the intensity biasing procedure, noticeable especially at the most extreme HNR values. Specifically, the more negative end of the HNR range produced proportionally greater N1-P2 responses than in Experiment 1 (cf. Fig. 3-3B and 3-3A); conversely the more positive end of the HNR scale produced smaller responses. Nonetheless, a main effect for HNR values still persisted in the results of this experiment (Table 3-1; $F_{5,70}$ = 21.482, P = .000). This finding is not surprising because these responses are known to be sensitive to the intensity of auditory stimuli (Näätänen and Picton, 1987, Picton, 2011). The general biphasic response profile seen in Experiment 1 was still apparent, strengthening its validity.

DISCUSSION

Summary of findings

We demonstrated that the amplitude of the N1-P2 AEP complex shows a nonlinear parametric sensitivity profile to harmonic content (HNR) in artificially-constructed IRN stimuli. Specifically, we 1) described this effect across an HNR range that reflects ethologically-relevant acoustic signal features, 2) showed that it persists even when compensating for possible intensity effects, and 3) revealed that its profile is biphasic in low HNR ranges. These results are compared and contrasted with similar findings from pitch-depth processing models and with respect to human cortical pathways that may be optimized to process the subtleties of vocalizations.

HNR findings and their relation to perception-based pitch-processing models

Other human electro- and magnetoencephalographic studies incorporating IRNs or similar stimuli have focused more on pitch and/or pitch-depth processing mechanisms that rely heavily upon perceptual features of sound signals. (Griffiths et al., 1998, Patterson et al., 2002, Krumbholz et al., 2003, Jones, 2006, Hall and Plack, 2009, Barker et al., 2012). EEG and MEG studies using IRNs or similar stimuli have investigated both pitch-onset responses (POR) and sound-onset responses (SOR). PORs recorded via MEG or EEG occur approximately 130-300ms after pitch onset (Krumbholz et al., 2003, Jones, 2006); in these studies, PORs were generated in response to the transition between white noise and an IRN that produces a perceived pitch that is inversely dependent upon the time delay applied during its creation (Yost, 1996a). Krumbholz et al. demonstrated that PORs reliably increase in amplitude with greater IRN iteration number – a parameter that is correlated with overall pitch strength (Yost, 1996b). EEG has produced similar results to those found with MEG and additionally showed that IRN-evoked PORs are reliably produced by high- or low-passed stimuli (Jones, 2006). Collectively, the authors of these

studies suggested that these responses were largely driven by the temporal regularity of the stimuli (periodicity pitch).

The above POR studies were designed to avoid purported confounds produced by SORs or energy-onset responses (EOR), the phenomena traditionally viewed to generate the auditory N1 (Näätänen and Picton, 1987). However, an MEG study investigating SORs to IRNs produced very similar results to the aforementioned POR studies (Soeta et al., 2005); N1m peak amplitudes increased as a function of IRN iteration number. Additionally, others have suggested that the SOR and POR are produced by very similar cortical generators (Seither-Preisler et al., 2004). Our current results, or at least a subset of our findings, are largely consistent with the above mentioned SOR and POR studies. However, our stimuli encompassed a much broader spectrum of acoustic characteristics, especially along the dimension of harmonic content (HNR). We specifically manipulated IRN iteration numbers as well as the gain applied during each delay-and-add process; gains for our stimuli ranged between 0.1 to 1.0. Doing so allowed us to achieve HNR values much closer to those exhibited by white noise samples (-7.6 dB). The "least structured" IRNs in many earlier studies were created with one iteration and a unity gain (although see Soeta et al., 2005that also used a white noise "IRN" of zero iterations). According to our HNR estimates, the IRNs with the lowest HNR values in many studies would be approximately 0dB (IRNs where iteration and gain values were equal to one (n=g=1)), which may have precluded earlier studies from observing the biphasic response reported here. Hence, our use of a broader set of IRN stimuli (with respect to HNR values), through greater variations of the iteration and gain values, allowed us to more fully encompass the quantifiable HNR range exhibited by natural real-world vocalizations (Lewis et al., 2009).

IRNs or similar stimuli have also been used in numerous studies that have investigated pitch-processing with psychophysical assessments and/or functional neuroimaging (Griffiths et al., 1998, Patterson et al., 2002, Hall et al., 2006, Hall and Plack, 2009). However, Hall et al. (2009) questioned the use of IRNs in pitch processing studies; their criticism stemmed from a perceived lack of appropriate control stimuli to IRN sounds. Specifically, they suggested that previous findings of IRN-specific activity may be dominated not by IRN pitch per se but by slow spectrotemporal modulations that

are randomly produced during IRN creation (Hall and Plack, 2009, Barker et al., 2012, Steinmann and Gutschalk, 2012). They addressed these concerns by producing a new class of IRN-derived stimuli that do not have perceptible pitches (IRNo, "no pitch" IRN). Contrasting BOLD activity produced by IRN and IRNo stimuli revealed very minimal differences, suggesting that perceived pitch may not be the driving factor of IRN-produced cortical activity. Thus, the goal of the present study was to advance a model of acoustic signal feature processing based on quantifiable measures (HNR) that encompassed a biologically relevant parameter range.

IRNs within an ethologically-relevant range of HNR

We designed our study to examine cortical responses to harmonic auditory content – a hallmark signal attribute of vocalization sounds(Riede et al., 2001). We have previously shown that different classes of vocalizations could be partially distinguished along the HNR continuum, including hisses, growls, groans, whispers, calls and speech (Lewis et al., 2009). For the current AEP study, we created a set of IRN stimuli that spanned a wide array of HNR values, notably encompassing the ethologically-relevant range found in communicative vocalizations described in our previous work (cf. Fig. 2-6 of Chapter 2). Our current results demonstrated that HNR is robustly represented in the N1-P2 complex (Fig. 3-3). The N1-P2amplitude response profile to IRNs was generally monotonically increasing with HNR values except for the range between -7.6dB (white noise) and -3dB HNR (see below). This finding contradicted our initial hypothesis that predicted the greatest amplitudes in response to IRNs with HNR values near those found in conversational speech (approximately 6-15 dB HNR).

The vertex scalp components that comprise the N1-P2 are thought to be generated near primary and secondary auditory cortices along and near Heschl's gyrus (Näätänen and Picton, 1987, Martin et al., 2008, Picton, 2011), suggesting that HNR-sensitivity occurs in early cortical auditory stages. Concordantly, our previous fMRI results (Fig. 3-1) showing similar BOLD effects along left and right Heschl's gyri and STG using the same basic IRN stimuli further corroborate this notion (Lewis et al., 2009). HNR-sensitive regions revealed with fMRI were spatially situated between tonotopically-

sensitive regions (along HG) and areas preferentially activated by human vocalizations (STG/STS). We interpreted IRN HNR-sensitive regions in the context of cortical template theories (Griffiths and Warren, 2002, Kumar et al., 2007) incorporating combination-sensitive neurons (Suga et al., 1983, Misawa and Suga, 2001, Medvedev et al., 2002) that show a preference to harmonic acoustic signals. These harmonic templates possibly represent foundational elements of early auditory cortical circuits and may crucially aid in the fine level processing of vocalization sounds.

HNR feature detection models

In contrast to perception-based pitch models, we are interpreting our results within the theoretical contexts of species-invariant feature extraction models. These models posit that combination-sensitive neurons, spectrotemporal templates, or other acoustic information filters act as bottom-up neuronal mechanisms for segregating and streaming auditory event information into distinct processing pathways (Suga et al., 1983, Margoliash and Fortune, 1992, Kanwal et al., 1999, Medvedev et al., 2002). These functional networks are posited to emerge in early auditory networks, become increasingly complex, and eventually combine in a hierarchical manner (Näätänen et al., 2001, Griffiths and Warren, 2002, Warren et al., 2005, Obleser et al., 2007). Our findings support a model that includes dedicated stages for harmonic processing.

The processing of concurrent or specific combinations of harmonics is thought to represent one means of auditory streaming (Rauschecker et al., 1995, Medvedev et al., 2002, Carlyon, 2004). For instance, presenting a listener with a sound containing mistuned harmonics can lead to the perception of two "distinct" sounds being presented simultaneously (Alain et al., 2001). Within our experimental paradigm, IRN stimuli can be viewed as "activating" or "matching" templates that are sensitive to integer-multiple harmonics spaced at intervals of 500Hz; increased N1-P2 amplitudes represent greater synchrony in these activated templates. These or similar templates would likely aid in the processing the strong harmonic content of vocalization sounds. Increasing the gain and iterations used during each IRN delay-and-add process would result in sound stimuli that more effectively engage or match these neuronal templates (i.e. the statistics of more

harmonic IRN noise approaches the optimal input for the described receptive fields). HNR as a signal attribute aids in quantifying this quality; IRNs with higher HNR values increasingly reflect better template matches. However, such a simple spectral template-matching representation appears to only be found in earlier cortical stages, precluding IRNs from activating more complex templates found in higher-order auditory regions such as the STS (Lewis et al., 2009, Talkington et al., 2012).

Similar to the current study, we previously identified cortical foci with fMRI that were parametrically sensitive to the HNR values of IRNs (Fig. 3-1) (Lewis et al., 2009). Those foci were anatomically near and overlapping with primary auditory cortices (PAC) along Heschl's gyrus (HG) and extended partially onto the mSTG; N1-P2 responses are generally thought to originate in similar cortical areas (Näätänen and Picton, 1987, Picton, 2011). Regions showing BOLD sensitivity to the HNR values of animal vocalizations showed partial overlap with those exhibiting IRN HNR-sensitivity, but generally occurred more laterally along the mSTG closer to human voice-sensitive areas near the STS (Belin et al., 2000, Belin et al., 2002). Higher-order cortices may be composed of templates for more behaviorally-relevant and familiar sounds, such as conspecific vocalizations (Talkington et al., 2012) and other categories of sound with more complicated "naturalistic" spectrotemporal characteristics (Belin et al., 2000, Belin et al., 2002, Fecteau et al., 2004, Leaver and Rauschecker, 2010, Lewis et al., 2012, Talkington et al., 2013). Collectively with our previous fMRI findings, our current results provide converging multi-modal neuroimaging evidence supporting harmonic content processing as distinct stages in the early auditory cortical networks of humans with typical hearing – individuals who rely on processing the complex subtleties of communicative vocalizations and speech signals on a daily basis.

HNR-sensitivity and perceived loudness

Previous studies have reported that the perceived loudness of IRNs or similar stimuli increases proportionally with the number of iterative delay-and-add cycles used to generate the sounds; this effect persists even if all of the sounds are equally intense (identical RMS power) (Soeta et al., 2007). IRNs are often perceived to be increasingly louder as a function of iteration number and gain, even after equalizing sound intensities (Soeta et al., 2007, Lewis et al., 2009). This phenomenon has been attributed to greater sensations of pitch strength of IRN stimuli, a perceptual attribute that is generally correlated with our HNR measures (Yost, 1997). More synchronously activated cortical regions could lead to a louder percept, confounding the interpretations of a template-based model. Thus, similar to our earlier fMRI study, we created intensity-biased IRN stimuli that decreased in overall intensity with respect to their HNR values to minimize possible confounds related to signal intensity or perceived loudness. The results of Experiment 2 corroborated the results of Experiment 1, showing a similar and robust biphasic N1-P2 amplitude response profile to the HNR signal attribute. There were slight intensity-related effects for very high and low HNR values (e.g. -7.6db and +24dB); however, this result was not surprising given the steep slope of our intensity-biasing function and previous findings that show larger AEPs as a function of stimulus intensity (Näätänen and Picton, 1987). Combined with our intensity-biased fMRI experiment, the current data provide critical support for the stable cortical representation of this acoustic attribute and the biphasic response profile seen in Experiment 1.

Biphasic HNR-dependent N1-P2 amplitude response profile

The most surprising finding of the present study was the shape of the HNR-sensitivity profile represented in the auditory N1-P2. As mentioned earlier, we unexpectedly revealed a biphasic N1-P2 amplitude response profile with respect to increasing IRN HNR values. These results contradict those from the aforementioned studies that have shown simpler monotonically-increasing amplitude functions as IRN iterations are increased. In particular, minimal N1-P2 amplitudes were produced by IRNs in an HNR

range approximately around -3dB HNR resulting in a “dip” in the otherwise increasing response profile. Individual subjects usually produced the smallest N1-P2 amplitudes in response to either the -3dB or the +3dB IRNs. Group-level averaged data (Fig. 3-3) showed the smallest amplitude responses at the -3dB HNR value.

One possibility is that the biphasic response findings reflect an example of a stochastic resonance or facilitation phenomenon (Wiesenfeld and Moss, 1995, McDonnell and Ward, 2011). A broad definition of stochastic resonance in neuronal systems describes it as beneficial noise within the context of signal detection and signal processing (McDonnell and Abbott, 2009). Specifically, stimuli that are sub- or near-threshold can become more easily detectable with the addition of noise to the signal(Douglass et al., 1993, Levin and Miller, 1996, Russell et al., 1999, Moss et al., 2004); the greatest detection probabilities for perithreshold stimuli occur with intermediate added signal noise. Confirmation of stochastic facilitation would require modulation of IRN HNR values with respect to other stimuli in simple perception and/or discrimination experiments (Srebro and Malladi, 1999, Moss et al., 2004). Alternatively, the current data seems to have the appearance of an inverse stochastic resonance curve. Models and real-world analysis of Hodgkin-Huxley systems suggest that intermediate amounts of noise can minimize neuronal activity (Paydarfar et al., 2006, Gutkin et al., 2009, Tuckwell and Jost, 2012). The N1-P2 amplitudes in response to very low HNR IRNs (around -3dB) could reflect a similar activity minimum in the human auditory system.

A related possibility is that the biphasic response profile represents a functional overlap of two (or more) potential cortical mechanisms (linear amplifiers/filters) useful for auditory scene analysis. Such a system could be formed by two acoustic filters: one specialized for the streaming of harmonic information in a range commonly found in mammalian vocalizations (typically 0dB to +24dB HNR) and another filter that may act to suppress or accommodate to noisy acoustic elements of a scene or background (-7.6dB to approximately -3dB or 0dB HNR) that are not as likely to represent distinct vocalization sources. These filters may be represented in distinct neuronal populations that could show partial to complete spatial overlap; nonetheless, our current AEP and

former fMRI results together suggest that these HNR-sensitive regions are located near the confluence of HG (near PACs) and the STG.

Given the combined strength of our previous fMRI results with the current electrophysiological data, we believe that harmonic signal processing is instantiated as a distinct intermediate cortical processing stage in the human auditory system, perhaps including our two proposed filtering functions that work in concert to simultaneously optimize harmonic signal enhancement and noise suppression. This putative signal processing principle could be tested and included in future models of the human auditory system or in prosthetic device algorithms designed to mimic its biological operations. Additionally, this cortical response pattern could be used to complement traditional audiometric measures when fitting a patient for hearing prosthetics that are designed specifically for the enhancement of vocalizations and speech.

ACKNOWLEDGMENTS

This work was supported by the NIGMS NIH COBRE grant P30 GM103503 (to the Center for Neuroscience of West Virginia University) and an individual pre-doctoral award to WJT funded by the Air Force Office of Scientific Research (AFOSR; American Society of Engineering Education (ASEE) National Defense Science and Engineering Graduate (NDSEG) Fellowship).

FIGURES AND TABLES

FIGURE 3-1

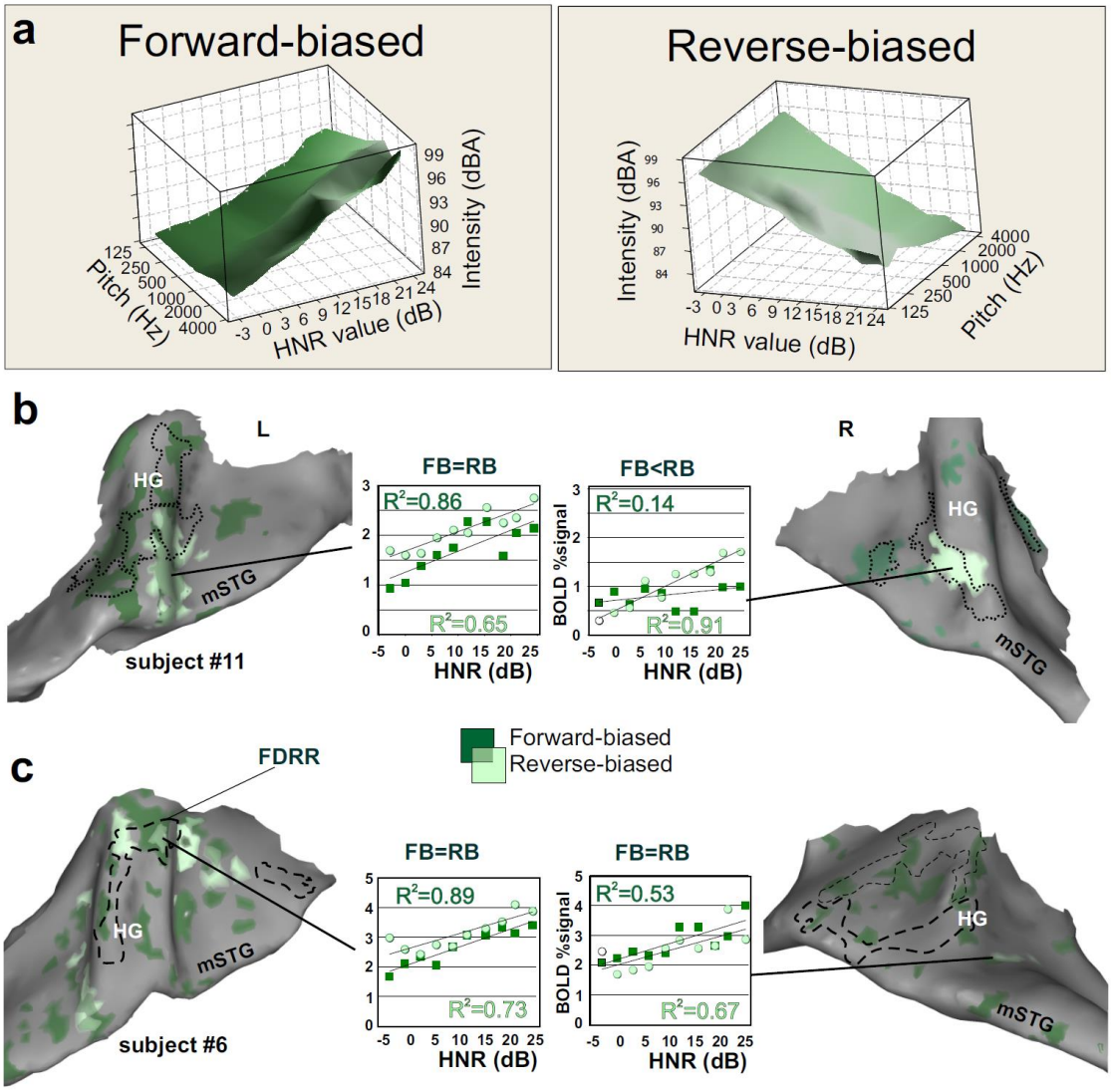


FIGURE 3-1. Intensity vs HNR-sensitivity to IRNs from our earlier fMRI study using IRNs that varied parametrically with HNR value. (**a**) Graphical depiction of 60 IRN stimuli showing a forward- or reversed-bias with stimulus intensity (total RMS power). (**b-c**) Cortical models of the left and right hemisphere auditory cortex for two participants illustrating functionally-defined regions of interest. Brain models were slightly inflated and smoothed to facilitate viewing of Heschl's gyrus (HG) and surrounding cortex. The green whole brain green mesh inset approximates the location of cortex cortical patches illustrated. Black dashed and dotted outlines depict frequency-dependent response regions (FDRRs) in auditory cortex, derived from a separate tonotopy mapping paradigm to estimate the locations of primary auditory cortices (PACs) for each participant. Individual data sets (2 of 4 shown) illustrating IRN HNR-sensitivity (green) under two separate conditions performed during the same fMRI scanning session: dark green = HNR forward-biased (FB) with loudness, light green = HNR reverse-biased (RB) against loudness. Both conditions revealed cortex sensitive to the HNR value of the IRNs located along and immediately surrounding PACs ($\alpha<0.05$, corrected for multiple comparisons). Charts show the linear correlation between HNR value and blood-oxygen level dependent (BOLD) activity (percent signal change relative to silent events; mean plus s.d.). The IRN data points were binned at 3 dB HNR intervals for clarity. L=left hemisphere, mSTG = middle superior temporal gyrus. Modified from Lewis et al., (2009) with permission from the Journal of Neuroscience.

FIGURE 3-2

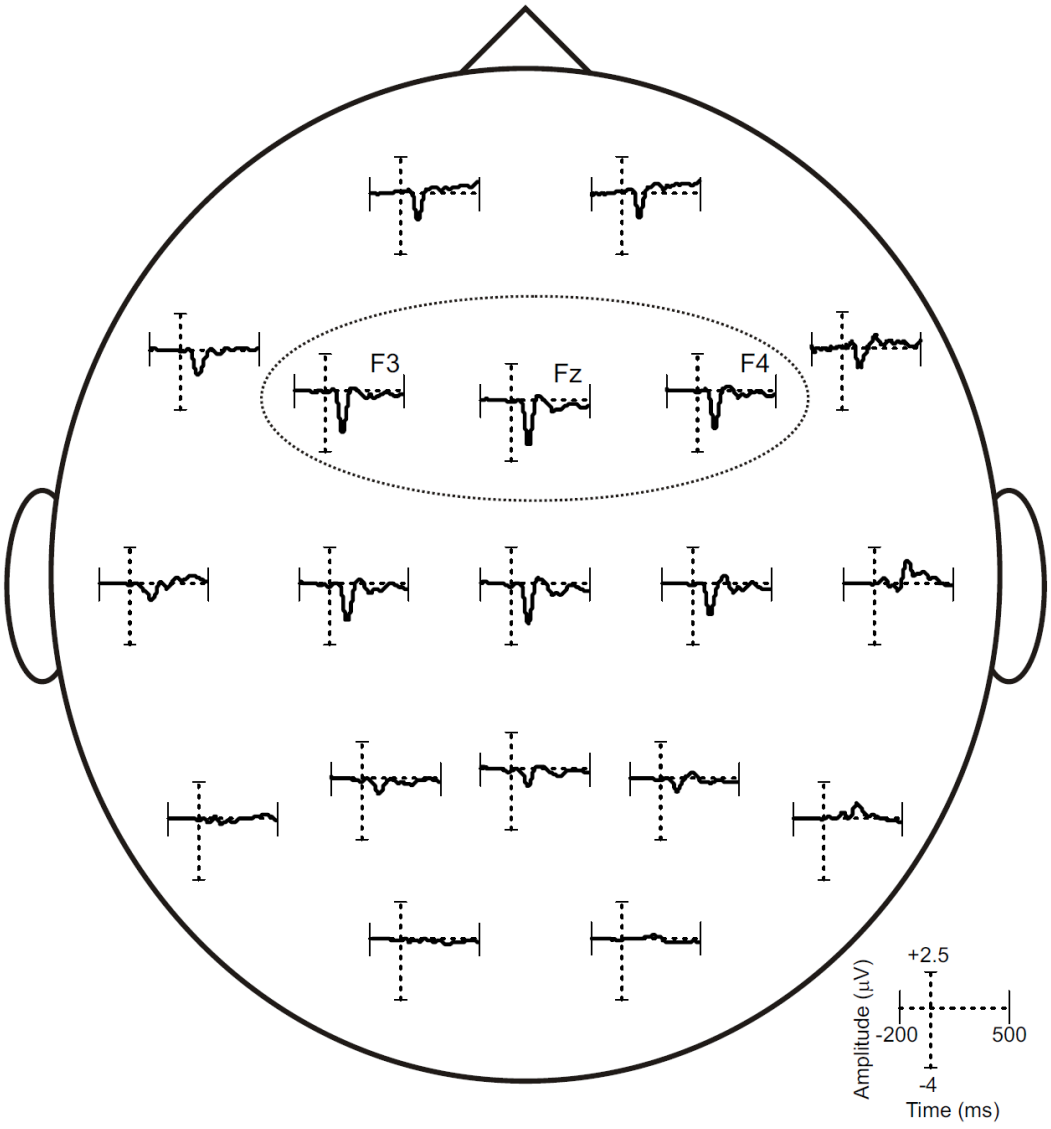


FIGURE 3-2. Group averaged (n=15) scalp topography and waveform morphology across all IRN HNR conditions from Expt. 1 (Iso-Intensity condition). Electrodes F3, Fz, and F4 (circled) were used in all analyses.

FIGURE 3-3

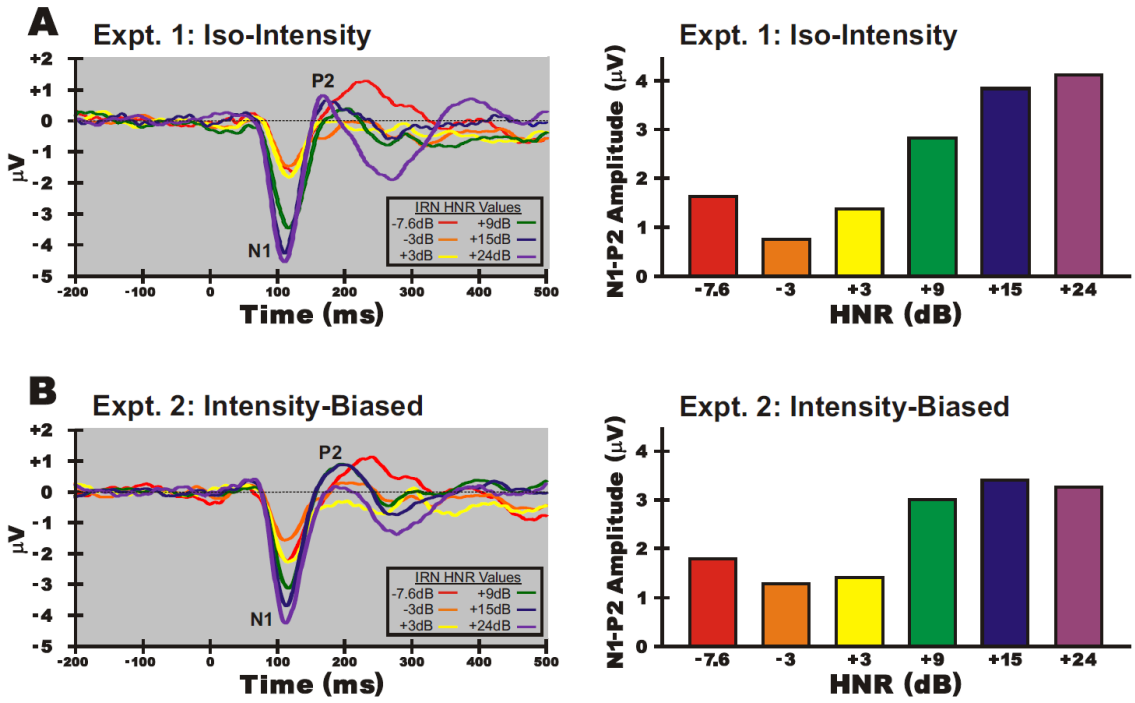


FIGURE 3-3. Average waveforms for electrodesF3, Fz, and F4 for white noise and the five IRN stimuli with parametrically varying HNR values. (**A**) N1-P2 complex responses to IRNs that had equal RMS intensity (Expt. 1; Iso-Intensity). (**B**) Response profile to IRNs that were reversed-biased with intensity (Expt. 2; Intensity-Biased), similar to the fMRI paradigm illustrated in Fig. 1.

TABLE 3-1. Group-averaged N1-P2 amplitudes and statistics for both experiments. Expt. 1: Iso-Intensity; Expt. 2: Intensity-Biased. All results reported are the average responses from electrodes F3, Fz, and F4. Standard deviations are in parentheses.

	HNR (dB)						Statistics
	-7.6	-3	+3	+9	+15	+24	df = 5,70
Iso-Intentisy	1.636	0.752	1.38	2.832	3.84	4.107	F = 63.574
Amplitude (μV)	(1.234)	(1.225)	(1.579)	(1.711)	(1.435)	(1.702)	p = .000
							ε = .657
Intensity-Biased	1.795	1.291	1.408	3.012	3.401	3.264	F = 21.482
	(1.515)	(1.428)	(1.292)	(1.454)	(1.242)	(1.686)	p = .000
							ε = .579

TABLE 3-2. Pairwise HNR condition comparisons for both experiments. Expt. 1: Iso-Intensity; Expt. 2: Intensity-Biased. Significance values in parentheses; the alpha values for both experiments were set at 0.05 and probabilities under this threshold are in boldface type.

HNR (dB)		Iso-Intensity	Intensity-Biased
A	B	A – B (Sig.)	A – B (Sig.)
-7.6	-3	**0.884 (0.024)**	0.505 (0.246)
	+3	0.257 (1.000)	0.387 (0.740)
	+9	**-1.196 (0.014)**	**-1.216 (0.004)**
	+15	**-2.204 (0.000)**	**-1.605 (0.002)**
	+24	**-2.471 (0.000)**	**-1.469 (0.012)**
-3	+3	-0.628 (0.308)	-0.117 (1.000)
	+9	**-2.080 (0.000)**	**-1.721 (0.001)**
	+15	**-3.088 (0.000)**	**-2.110 (0.000)**
	+24	**-3.355 (0.000)**	**-1.974 (0.004)**
+3	+9	**-1.452 (0.000)**	**-1.604 (0.001)**
	+15	**-2.460 (0.000)**	**-1.992 (0.000)**
	+24	**-2.727 (0.000)**	**-1.856 (0.004)**
+9	+15	**-1.008 (0.001)**	-0.389 (1.000)
	+24	**-1.275 (0.000)**	-0.252 (1.000)
+15	+24	-0.267 (1.000)	0.136 (1.000)

CHAPTER 4:

Late auditory evoked potentials in native Mandarin speakers exhibiting sensitivity to harmonic signal content

William J. Talkington, Brandon D. Smith, Stephanie K. Khoo,
Christopher A. Frum, James W. Lewis

Center for Neuroscience,
Center for Advanced Imaging in the Department of Radiology,
Department of Neurobiology and Anatomy
West Virginia University, Morgantown, WV 26506

ABSTRACT

In the present study, we recorded cortical auditory evoked potentials (AEP) from native Mandarin speakers in response to artificially constructed iterated ripple noise (IRN) sounds that parametrically spanned an ethologically-relevant range of HNR values. Similarly to native English speakers, the N1-P2 AEP complex demonstrated a positive and monotonically increasing amplitude response to HNR values between -3dB to +24 dB; low HNR value ranges showed a decrease in AEP amplitude responses (from white noise (-7.6dB HNR) to -3 dB HNR). The results from native Mandarin speakers were quantitatively indistinguishable from those produce by native English speakers. Together with our AEP findings from Chapter 3, these results provide converging evidence of a stable representation of HNR as a cortically represented acoustic signal attribute, regardless of individual language experiences.

INTRODUCTION

Long-term experiences with tonal versus non-tonal spoken languages has been shown to shape auditory brainstem encoding of acoustic features such as pitch (Krishnan et al., 2005, Krishnan and Gandour, 2009, Krishnan et al., 2010a, b). Generally, when compared to speakers of non-tonal languages such as English, the auditory brainstem responses (ABR) produced by Mandarin (or Thai) speakers more efficiently encode or "track" the pitch of linguistically relevant pitch contours. Pitch tracking is often measured with an auditory brainstem phenomenon referred to as the frequency following response (FFR) that has been shown to be sensitive to the intensity and frequency of tonal stimuli (Stillman et al., 1978) and can represent portions of speech signals (Krishnan et al., 2004, Johnson et al., 2005).

Thus, we questioned if and how robust the biphasic N1-P2 amplitude response pattern to the harmonic content (HNR) of IRN stimuli (described in Chapter 3) would appear in native Mandarin speakers. As individuals with life-long experience distinguishing tones and tonal variations in complex harmonic vocalizations, we hypothesized that Mandarin speakers would produce an equivalent or stronger response pattern (more defined biphasic "dip" response) to IRNs of varying HNR values. We proposed this notion due to their reliance on accurate harmonic signal processing for successful language comprehension and production.

MATERIALS AND METHODS

Participants

Using a paradigm identical to Experiment 1 of Chapter 3, the current experiment studied native Mandarin-speaking right-handed participants (n=5, two female, average age = 27.6 years). Each subject used Mandarin as their primary language at home; the subjects received between 7.5-15 years of formal English instruction, but no subject had used English on a daily basis for more than 3 consecutive years. Thus, in contrast to our monolingual English subjects from Chapter 3, the current cohort of subjects was highly proficient at both hearing and producing tonal Mandarin speech sounds.

Stimuli, Electrophysiology Procedures, and Data Analyses

Refer to the Chapter 3 Materials and Methods section for details on the IRN stimuli, and data collection parameters implemented in this experiment. Other than the subject population and subsequent analyses performed in this chapter, the current experiment mirrors Experiment 1 from Chapter 3. Additional analyses included a two-way repeated measures ANOVA to compare the two subject groups, native speakers of English and Mandarin, respectively.

RESULTS

Figure 4-1A displays group-averaged (n=5) evoked potentials in native Mandarin speakers to IRN stimuli that were equally intense in RMS power but differed in their harmonic content (HNR). Similar to Expt. 1 in Chapter 3, a main effect of HNR was seen on N1-P2 amplitudes ($F_{5,20}$ = 14.695, P= 0.003). Our hypothesis that Mandarin speakers would produce a biphasic N1-P2 response profile was confirmed. A monotonically increasing amplitude trend was seen between HNR values between approximately -3 and +24 dB; an inverse relationship was seen between -7.6 and -3dB. Pairwise comparisons between HNR conditions indeed revealed that the N1-P2 values produced by white noise (-7.6dB HNR) were significantly higher than those produced by -3dB HNR IRNs (P = 0.048). AEP responses to white noise were indistinguishable from those produced by +3dB HNR IRNs (P=1.0).

In addition to confirming the biphasic response profile in native Mandarin speakers, we also aimed to compare their responses to native English speakers to determine whether lifelong language experience modulates this cortical response. Figure 4-1B displays the results from Experiment 1 in Chapter 3 from native English speaking subjects (equivalent to Figure 3-3A). Comparing the two groups demonstrated no differences between their respective HNR-dependent AEP trends ($F_{5,90}$ = 0.678, P=0.577).

DISCUSSION

Extensive acoustic training or experience with behaviorally relevant pitches or sounds is thought to modify function in the relevant structures or networks subserving the auditory mechanisms described above. Concordantly, previous studies have investigated the effects of expert musical skill and language experience on auditory processes (Wong et al., 2007, Chen et al., 2008, Krishnan et al., 2009b). The auditory abilities and responses of Mandarin speakers, and other tonal language speakers, are often compared to those of speakers from non-tonal languages with the rationale that one's listening experiences and the behavioral need to discriminate tones and tonal changes may influence cortical, and even subcortical, network processing. For instance, the frequency-follower response (FFR), likely generated in the inferior colliculi or lateral lemnisci (Stillman et al., 1978), has been shown to produce stronger pitch-following responses in native Mandarin speakers relative to English speakers (Krishnan et al., 2004). In particular, dynamic pitch-varying IRN-derived stimuli homologous to Mandarin Tone 2 have shown greater pitch-tracking to behaviorally (i.e. linguistically) relevant sounds in the brainstems of native Mandarin speakers (Krishnan et al., 2009a).

We recorded HNR-dependent AEP amplitudes from a cohort of native Mandarin speakers. No quantitative differences in their biphasic response trend were found when compared to the results from their English-speaking counterparts. However, a larger cohort of Mandarin speaking subjects (n=5 currently) may reveal subtle differences that are currently lacking statistical strength. A lack of differences between these two groups may be consistent with the notion that tonal language experience imparts processing advantages useful for *dynamic* frequency tracking in stimuli (Krishnan et al., 2010c). Nonetheless, these results support the robust representations of harmonic content in human auditory cortices regardless of language experience; this suggests that harmonic signal encoding is a fundamental processing feature that is common to all hearing individuals.

FIGURES

FIGURE 4-1

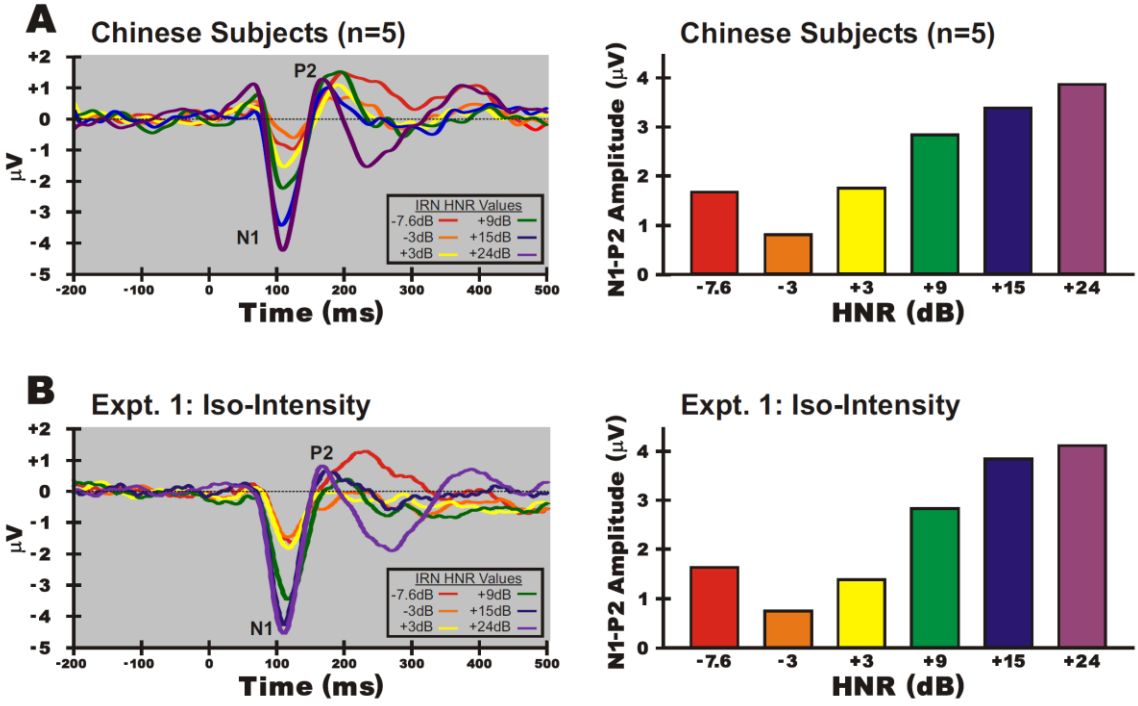


FIGURE 4-1. Average waveforms for electrodes F3, Fz, and F4 for white noise and the five IRN stimuli with parametrically varying HNR values. (**A**) N1-P2 complex responses in native Mandarin speakers to IRNs that had equal RMS intensity. (**B**) N1-P2 complex responses in native English (USA) speakers to IRNs that had equal RMS intensity (Chapter 3 - Expt. 1; Iso-Intensity).

CHAPTER 5:

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

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

[1]Center for Neuroscience, Center for Advanced Imaging, Departments of Physiology and Pharmacology, West Virginia University, Morgantown, WV, 26505

[2]Division of Theatre and Dance, College of Creative Arts
West Virginia University, Morgantown, WV 26506 USA

This manuscript was published in the Journal of Neuroscience on June 6th, 2012.

ABSTRACT

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

INTRODUCTION

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

Within human auditory cortices, we predicted that there should be a categorical hierarchy reflecting an increasing sensitivity to one's conspecific vocalizations and utterances. Previous studies investigating cortical voice-sensitivity in humans have compared responses to *stereotypical* speech and non-speech vocalizations with responses to other sound categories, including animal vocalizations and environmental sounds (Belin et al., 2000, Belin et al., 2002, Fecteau et al., 2004). However, these comparisons did not always represent gradual categorical differences, especially when using broadly defined samples of "environmental sounds". Thus, in the current study, we utilized animal vocalizations together with naturally-produced human-mimicked versions (Lass et al., 1983). Human-mimicked animal vocalizations acted as a crucial intermediate vocalization category of human-produced stimuli, acoustically and conceptually bridging

between animal vocalizations and stereotypical human vocalizations. We therefore avoided confounds associated with using over-learned acoustic stimuli when characterizing these early vocalization processing networks (e.g. activation of acoustic schemata (Alain, 2007)). Using high-resolution functional magnetic resonance imaging (fMRI), our findings suggest that the cortical networks mediating vocalization processing are not only organized by verbal and prosodic non-verbal information processing (left and right hemispheres, respectively), but also that the left hemisphere processing hierarchy becomes organized along an acoustic dimension that reflects increasingly meaningful conspecific communication content.

MATERIALS AND METHODS

Participants

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

Vocalization sound stimulus creation and acoustic attributes

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

Six non-imaging volunteers recorded human-mimicked versions of corresponding animal vocalization stimuli. Each mimicker attempted to match the spectrotemporal qualities of the real-world animal vocalizations. A group of four listeners then assessed the acoustic similarity of each animal-mimic pair until reaching a consensus for the optimal mimicked recordings. A subset of our fMRI subjects (n=18/22) psychophysically rated all of the animal vocalization and human-mimics after their respective scanning

sessions. Subjects were asked to rate each stimulus (button response) along a 5-point Likert-scale continuum to assess the "animal-ness" (low-score, 1 or 2) or "human-ness" (high score, 4 or 5) quality of the recording. Stimuli rated ambiguously along this dimension were given a score of three (3). The number of subjects who correctly categorized each animal or human-mimicked vocalization is displayed in Table 5-1.

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

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

All of the sound stimuli were quantitatively analyzed; the primary motivation for these analyses was to acoustically compare the stimuli in each animal-mimic pair (Table 5-1). The harmonic content in each stimulus was quantified with a harmonics-to-noise ratio (HNR) using Praat software (http://www.fon.hum.uva.nl/praat/) (Boersma, 1993). HNR algorithm parameters were the default settings in Praat (Time step (s): 0.01; Min. Pitch (Hz): 75; Silence threshold: 0.1; Periods per window: 1.0). Weiner entropy and spectral structure variation (SSV) were also calculated for each sound stimulus (Reddy et al., 2009, Lewis et al., 2012). We used a freely-available custom Praat script to calculate Weiner entropy values (http://www.gbeckers.nl/; Gabriel J.L. Beckers, Ph.D.); the script

was modified to additionally calculate SSV values which are derived from Weiner entropy values.

Scanning paradigms

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

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

Using techniques described previously from our laboratory, a subset of participants (n=5) participated in an fMRI paradigm designed to tonotopically map auditory cortices (Lewis et al., 2009). Briefly, tonotopic gradients were delineated in each subject's hemispheres using a "Winner-Take-All" (WTA) algorithm for calculating preferential blood-oxygenated level dependent (BOLD) responses to three different frequencies of pure-tones (PT) and one-octave band-pass noises (BPN) relative to "silent" events: 250Hz (Low), 2000Hz (Medium), and 12,000Hz (High). An uncorrected node-wise statistical threshold of $p<0.001$ was applied to each subject's WTA cortical maps; tonotopic gradients were then spatially-defined in regions that exhibited contiguous Low-

Medium-High progressions of preferential frequency responses along the cortical mantle. The tonotopic gradients of all subjects were then spatially averaged, irrespective of gradient direction, on the common group cortical surface model (created by averaging the surface coordinates of all 22 fMRI participants, see below). This effectively created a probabilistic estimate of primary auditory cortices (PAC) for our group of participants to be used as a functional landmark. These results were in agreement with anatomical studies that implicate the likely location of human PAC to be along or near the medial two-thirds of Heschl's gyrus (HG) (Morosan et al., 2001, Rademacher et al., 2001).

Magnetic resonance imaging data collection and pre-processing

Stimuli were presented during relative silent periods without functional scanner noise by utilizing a clustered-acquisition fMRI design (Edmister et al., 1999, Hall et al., 1999). Whole-head, spiral in-and-out images (Glover and Law, 2001) of the BOLD signals were acquired on all trials during functional sessions including silent events as a control condition using a 3T GE Signa MRI scanner. A stimulus or silent event was presented every 9.3 seconds, and 6.8 seconds after event onset BOLD signals were collected as 28 axial brain slices approximately centered on the posterior superior temporal gyrus (STG) with 1.875 x 1.875 x 2.00 mm^3 spatial resolution (TE = 36 msec, OPTR = 2.3 sec volume acquisition, FOV = 24 mm). The presentation of each stimulus event was triggered by the MRI scanner via a TTL pulse. At the end of functional scanning, whole brain T1-weighted anatomical MR images were acquired with a spoiled GRASS pulse sequence (SPGR, 1.2 mm slices with 0.9375 x 0.9375 mm^2 in plane resolution). Both paradigms utilized identical functional and structural scanning sequences.

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

Individual subject analysis

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

Group-level analyses

Regression coefficients for relevant functional contrasts generated with AFNI/SUMA were grouped across the entire subject pool and entered into two-tailed *t*-tests. These results were then corrected for multiple comparisons in the following manner using Caret6 (Van Essen et al., 2001, Hill et al., 2010): (1) permutation-based corrections were initiated by creating 5000 random permutations of each contrast's t-score map; (2) t-maps were smoothed by an average neighbors algorithm with four iterations (0.5 strength per iteration); (3) Threshold-free cluster enhancement (TFCE) was applied to each permutation map (Smith and Nichols, 2009), optimized for use on cortical surface models

with parameters: E=1.0, H=2.0 (Hill et al., 2010); (4) a distribution ranking maximum TFCE scores was created to find the 95th percentile statistical cutoff value; (5) this value was then applied to the original t-score map to produce the dataset in Figure 5-1.

Lateralization indices were calculated for each of the functional contrasts described within this manuscript (Fig. 5-1, M>A, F>M, E>M, and M>E). We accomplished this using a threshold- and whole-brain region of interest (ROI)-free method (Jones et al., 2011). For each functional contrast, we created distributions of non-thresholded *t*-test scores within each hemisphere. After log-transforming these distributions, the centers of each ($-4 \leq t \leq 4$) were fit with parabolic equations to approximate noise in the distributions. Subtracting these noise-approximations from the original score distributions and integrating the results provided a quantitative measure for an individual contrast's strength of activation within a hemisphere. Left and right hemisphere scores were then plotted against one another; the absolute distances of these points from the zero-difference "bilateral" line (slope = 1) represented the relative lateralization of a given function (Fig. 5-1 illustrates these scores graphically.).

Psychophysical affective assessments of sound stimuli

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

RESULTS

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

To assess our hypothesis that the use of non-stereotypical human vocalizations might reveal earlier stages of species-specific vocalization processing, we sought to identify cortical regions preferentially activated by *human-mimicked* animal vocalizations. Preferential group-averaged BOLD activity to the human-mimicked stimuli relative to their corresponding animal vocalizations was strongly left-lateralized and confined to a large focus in the group-averaged dataset. This activation encompassed regions from the lateral-most aspects of HG, further extending onto the STG, and marginally entering the STS (M>A; Fig. 5-1, yellow, $p<0.05$ Threshold-Free Cluster Enhancement (TFCE) and permutation-corrected (Smith and Nichols, 2009)). BOLD values in regions defined by this contrast and others discussed below are highlighted in Figure 5-2. This mimic-sensitive focus (yellow) was located near and partially overlapping functional estimates of PAC. Even within some individuals, the activation foci for human mimic sounds

bordered or partially overlapped their functional PAC estimates (Fig. 5-3; yellow near or within black dotted outlines). Right-hemisphere mimic-sensitive activity in the group-averaged data set was confined to a small focus along the upper bank of the STS (Fig. 5-1, yellow). We also calculated a lateralization index (LI) (Jones et al., 2011) with whole brain threshold- and ROI-independent methods (Fig. 5-1; $LI_{M>A}$=-2.68) that strongly supported this robust left-lateralization at the group level.

When contrasted with the animal vocalizations, the corresponding human mimic vocalizations were generally well matched for low-level acoustic features such as rhythm, cadence, loudness, and duration. Acoustic and psychophysical attributes were also derived to quantify some of the differences between the mimic-animal vocalizations at sound-pair and categorical levels. One acoustic attribute we measured is related to harmonic content, a signal quality that is significantly represented in vocalizations (Riede et al., 2001, Lewis et al., 2005); this was accomplished by quantifying a harmonics-to-noise ratio (HNR) for each stimulus (see Methods). We previously reported harmonic processing as a distinct intermediate stage in human auditory cortices by showing cortical regions that were parametrically sensitive to the harmonic content of artificial iterated rippled noise (IRN) stimuli and real-world animal vocalizations (Lewis et al., 2009). In the present study, HNR values for human-mimicked vocalizations were typically greater than their corresponding animal vocalizations; these differences persisted at the categorical level (*t*-test, $p<0.05$) (Table 5-1).

Two other acoustic attributes we calculated were related to signal entropy measures. Also known as the spectral flatness measure (SFM), Weiner entropy quantifies the spectral density in an acoustic signal in the form of resolvable spectral bands (Reddy et al., 2009). Consequently, white noise ("simple" diffuse spectrum) and pure tones (infinite spectral power or density at one frequency) lie at the extreme ends of this attribute's range (white noise: 0, pure tone: -∞). This attribute has been used previously to characterize environmental sounds (Reddy et al., 2009, Lewis et al., 2012). Generally, vocalizations produce the most negative values since they usually contain very specifically structured spectral content (often a fundamental frequency and a few formants). Human-mimicked animal vocalizations from the current study typically had less negative entropy values than their animal vocalization counterparts (Table 5-1),

implying that they possessed relatively less ordered acoustic structure. Group-level analysis confirmed the Weiner entropy differences between these two categories (*t*-test, $p<0.0005$).

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

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

were collected under different conditions (naïve and in the scanner vs. non-naïve and outside of the scanner) also precluded a rigorous "error trials" analysis.

To further identify where these human vocal tract-sensitive regions (M>A) were located in the auditory processing hierarchy, we also compared the responses to mimic stimuli with responses to foreign and English speech samples. Preferential activation to unfamiliar foreign speech, which is incomprehensible with respect to locutionary (semantic) content (Austin, 1975), relative to human-mimicked animal vocalizations (F>M) should reflect the general processing of dynamic spectrotemporal acoustic features typical of spoken languages and utterances at later auditory stages. Cortical regions preferentially responding to foreign speech from six different non-Romance and non-Germanic languages (Akan, Farsi, Hebrew, Hindi, Mandarin, and Yoruban) predominantly radiated posterolaterally out from the left-hemisphere mimic-sensitive focus and into the left STS, as well as medially onto HG (Figure 5-1, red).

In addition to basic speech sensitivity, contrasting the BOLD activity between English speech vocalizations and the mimic stimuli (E>M) specifically highlighted cortical sensitivity to the subtle differences in acoustic cues that convey *comprehensible* locutionary communication in one's native language relative to the more fundamental vocal tract sound signals. This condition revealed a strongly left-lateralized expanse of activity (Fig. 5-1, dark blue) situated further into the STS than foreign-vs-mimic sensitive regions. Responses to the spoken verbal stimuli in our paradigm, whether foreign or native, produced the most strongly left-lateralized networks along the STG and STS (Fig. 5-1; $LI_{F>M}$=-4.47; $LI_{E>M}$=-5.48). Importantly, our experimental design emphasized conspecific vocal-tract sensitivity and thus did not incorporate any overt phonological, syntactic, or semantic "language tasks" that may have produced more bilateral activation (Hickok and Poeppel, 2007). Collectively, these results form the basis of a *left-hemisphere* auditory processing hierarchy that is organized by increasingly complex and precise statistical representations of conspecific communication sounds. Directed laterally and anterolaterally along the cortical ribbon (Chevillet et al., 2011), this hierarchy ultimately culminated in the cortical representations of conspecific utterances that express locutionary (semantic) information, similar to models of intelligible speech processing (Scott et al., 2000, Davis and Johnsrude, 2003, Friederici et al., 2010).

Individual subjects revealed similar left hemisphere hierarchies that seemed to emerge from around PACs (e.g. Fig. 5-3, yellow to red to blue progressions emanating near HG, extending onto the STG, and into the STS).

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

DISCUSSION

Lateralized Cortical Sensitivity to the Human Vocal-Tract

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

The human-mimicked animal vocalizations used in this study varied qualitatively in their overall imitation "accuracy" when compared to their corresponding animal vocalizations. Attempting to match the *pitches* of the animal vocalization, the mimickers likely relied more upon the use of their vocal folds (cords). Furthermore, straining the limits of their vocal abilities and the physical limitations of their vocal structures likely emphasized additional *nonlinear* acoustic elements that are unique to or characteristic of the human vocal tract (Fitch et al., 2002). While a definitive analysis of human-specific vocal acoustics was beyond the objectives of the current study, we nonetheless aimed to acoustically describe our animal vocalization and human-mimic stimuli in part by using three quantitative measures: HNR, Weiner entropy, and SSV. Each of these attributes has been used previously to describe various categories of sound including vocalizations, tool sounds, and other environmental sounds (Riede et al., 2001, Lewis et al., 2009, Reddy et al., 2009, Lewis et al., 2012). The increased harmonic content (greater HNR values) seen

for human-mimics versus animal vocalizations may reflect a greater reliance on the vocal folds when attempting to match pitches. This effect may further be paralleled by the relative increases in the signal entropy of the mimics. Nonlinear acoustic phenomena emphasized during vocal strain would effectively spread the overall spectral density of human-mimicked vocalizations (Fitch et al., 2002). In addition to the harmonic and spectral entropy changes we observed, human-mimicked versions of animal vocalizations further revealed a decrease in their spectral dynamicity (SSV measures). By straining themselves to go beyond their typical vocal repertoires, mimickers may have been less likely to implement highly-learned and complicated articulatory routines that are typically used during language production. Overall, the quantitative changes observed within these three acoustic signal attributes (Table 5-1) are consistent with the notion that the human-mimicked animal vocalizations generally represented acoustically "simplified" versions of their real-world counterparts, simplified in a manner that emphasized acoustic phenomena that are unique to the human vocal tract.

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

The results of our experiment newly suggest that "voice-sensitivity" for humans predominantly emerges along the boundary between the left HG and STG in close

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

The Evolution of Conspecific "Voice-Sensitivity"

The evolution of cortical networks that mediate vocal communication and language functions, and more specifically the lateralization of these functions and supporting anatomical structures, is a burgeoning area of research (for review see Wilson and Petkov, 2011). Specific anatomical differences between primate species point to left-hemisphere biases for the structures that would putatively support the emergence of language functions. For instance, diffusion tensor imaging (DTI) tractography results demonstrate a striking expansion and increasing connectivity of the left arcuate fasciculus between macaques, chimpanzees and humans (Rilling et al., 2008). Additionally, posterior temporal lobe regions such as the planum temporale in chimpanzees display asymmetries in gross anatomical structure, similar to those seen in humans (Gannon et al., 1998, Hopkins et al., 1998). Neuroimaging techniques are increasingly being used to describe whole-brain networks for vocalization processing in lower primates (Gil-da-Costa et al., 2006). Functional neuroimaging (fMRI) in Old World monkeys (macaques) has demonstrated bilateral foci showing preference for species-specific vocalizations with a more selective focus occurring along the right anterior superior temporal plane (Petkov et al., 2008). Another macaque study using positron emission tomography (PET) revealed preferential left anterior temporal lobe activity to species-specific vocalizations (Poremba et al., 2004). PET neuroimaging in great apes (chimpanzees) has revealed a right hemisphere preference for certain conspecific vocalizations and utterances (Taglialatela et al., 2009). However, the responses to conspecific vocalizations in chimpanzees were

not directly compared to those produced by other species thereby precluding interpretations regarding species-specificity *per se*. While the auditory pathways in lower and higher non-human primates that process vocalization sounds require further study, findings hitherto support the presence of at least some lateralized functions in networks for processing conspecific vocalizations.

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

Vocalization Processing in a Neurodevelopmental Context

The present findings also have significant implications for language development in children. Early stages in human auditory processing pathways may be, or develop through experience to become, optimized to process the statistically representative qualities unique to the human vocal tract. This arrangement would promote maximal extraction of conspecific communicative content from complex auditory scenes (i.e. socially-relevant vocal communication from other humans). Seminal steps of this process would likely involve encoding the fundamental acoustic signatures of personally significant vocal tracts, initially including the voices of one's caretakers', one's own voice and, for social animals, the voices of other conspecifics. For example, human infants generally produce more positive and preferential responses to 'motherese' and other baby-directed vocalizations (Cooper and Aslin, 1990, Mastropieri and Turkewitz, 1999). Those responses may be driven heavily by the relatively stable statistical structure of basic "simplified" vocalizations (Fernald, 1989), notably vowels and other utterances possessing relatively simple amplitude envelopes, elevated pitch and strong harmonic

content, the latter being a hallmark acoustic attribute of vocal communication sounds across most species (Riede et al., 2001, Lewis et al., 2005, Lewis et al., 2009). While it remains unclear whether sensitivity to intrinsic human vocal tract sounds reflects domain-specific functions (nature) or auditory experience (nurture), the auditory experiences that initiate or influence these functions may begin *in utero* (DeCasper et al., 1994b) while a fetus experiences harmonically-structured vocal sounds. Nonetheless, a longer developmental timeframe would ostensibly follow this initial sensitivity, during which an emergent sensitivity to more subtle and complex socially-relevant acoustic signal cues appears as more advanced communicative and language abilities develop (Wang, 2000).

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

language development including specific language impairments (SLI) and autism (Gervais et al., 2004, Shafer and Sussman, 2011).

FIGURES AND TABLES

FIGURE 5-1

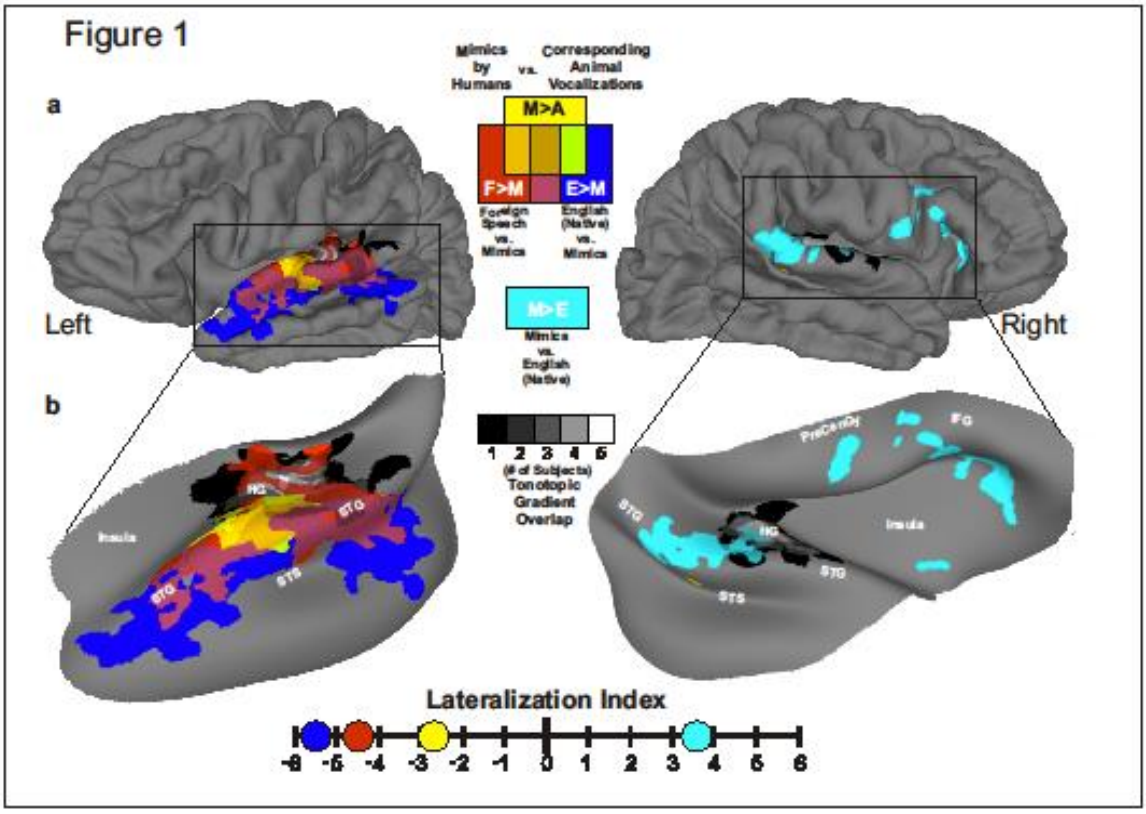


FIGURE 5-1. Conspecific vocalization processing hierarchy in human auditory cortex. a, Group-averaged (n=22) functional activation maps displayed on composite hemispheric surface reconstructions derived from all subjects. **b,** To better visualize the data, we inflated and rotated cortical projections within the dotted-outlines in **(a)**. The spatial locations of tonotopic gradients from five subjects were averaged (black-to-white gradients) and located along HG. Mimic-sensitive regions (M>A) are depicted by yellow hues, sensitivity to foreign speech samples versus mimic vocalizations (F>M) are depicted by red hues, and sensitivity to native English speech versus mimic vocalizations (E>M) is depicted by dark blue. Regions preferentially responsive to mimic vocalizations versus English speech samples (M>E) are depicted by cyan hues. Corresponding colors indicating functional overlaps are shown in the figure key. All data are TFCE-enhanced and permutation-corrected for multiple comparisons to $p<0.05$. To quantify the laterality of these functions, we calculated and plotted lateralization indices using threshold- and whole-brain region of interest (ROI)-free methods. Lateralization indices showed increasingly left-lateralized function (negative values indicate a leftward bias) for processing conspecific vocalization with increasing amounts of locutionary information; $LI_{M>A}$=-2.68, $LI_{F>M}$=-4.47, $LI_{E>M}$=-5.48, $LI_{M>E}$=+3.59. Additional anatomy: pre-central gyrus (PreCenGy), and inferior frontal gyrus (IFG).

FIGURE 5-2

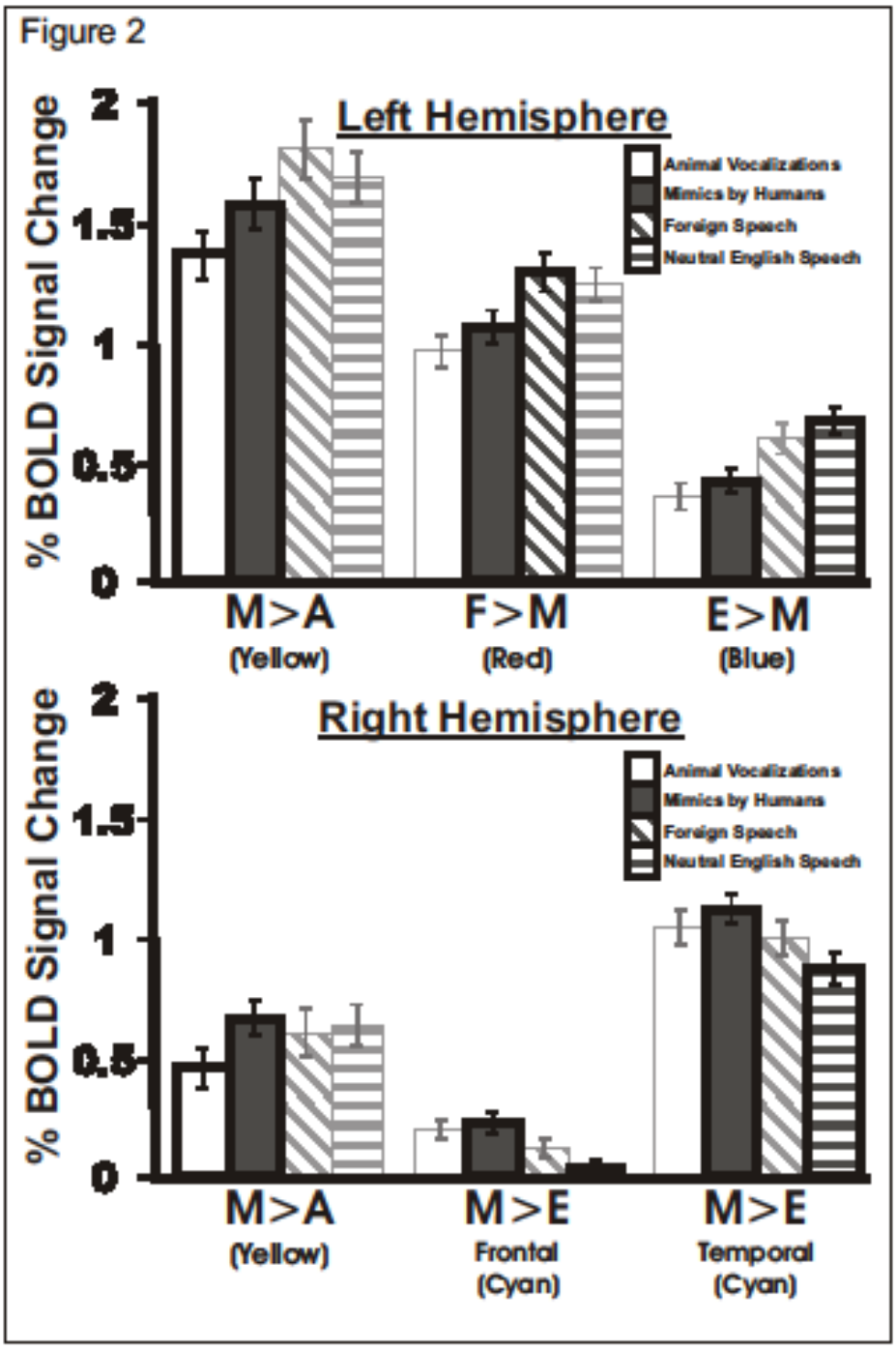


Figure 2
Left Hemisphere
% BOLD Signal Change
2
1.5
1
0.5
0
Animal Vocalizations
Mimics by Humans
Foreign Speech
Neutral English Speech
M>A
(Yellow)
F>M
(Red)
E>M
(Blue)
Right Hemisphere
% BOLD Signal Change
2
1.5
1
0.5
0
Animal Vocalizations
Mimics by Humans
Foreign Speech
Neutral English Speech
M>A
(Yellow)
M>E
Frontal
(Cyan)
M>E
Temporal
(Cyan)

FIGURE 5-2. Quantitative representation of BOLD fMRI activation. Mean BOLD signal responses (n=22 subjects) to the four vocalization categories were quantified for each focus or region identified in Fig. 5-1. Data correspond to the means ± s.e.m. The functional regions identified in Fig. 5-1 are indicated under each four-bar cluster. Left hemisphere regions from Fig. 5-1: M>A (yellow), F>M (red), and E>M (dark blue); right hemisphere regions from Figure 5-1: M>A (yellow), M>E-Temporal and M>E-Frontal (cyan).

FIGURE 5-3

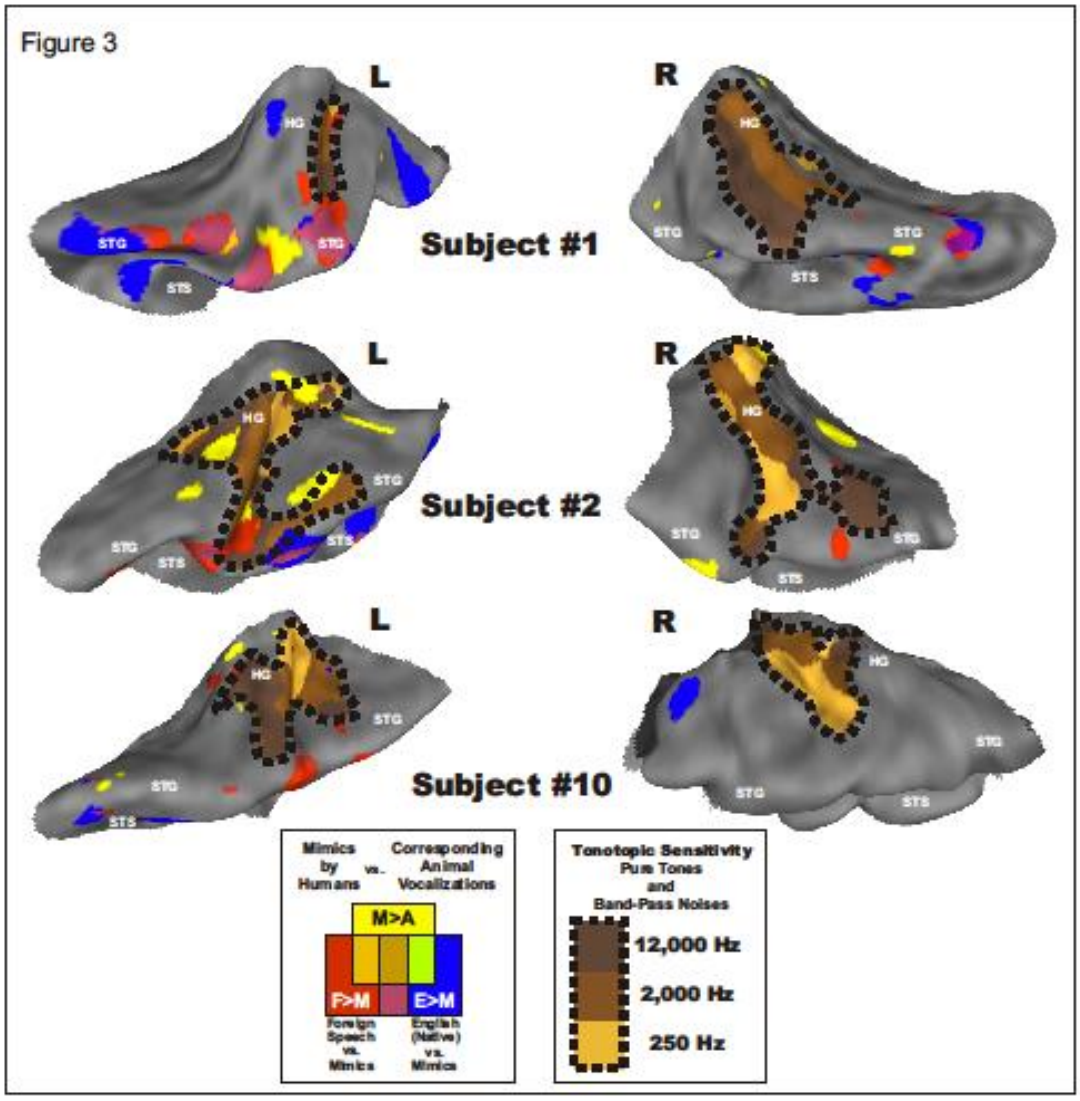


Figure 3
L
R
HG
STG
STS
Subject #1
Subject #2
Subject #10
Mimics by Humans vs. Corresponding Animal Vocalizations
M>A
F>M
E>M
Foreign Speech vs. Mimics
English (Native) vs. Mimics
Tonotopic Sensitivity
Pure Tones and Band-Pass Noises
12,000 Hz
2,000 Hz
250 Hz

FIGURE 5-3. Vocalization-sensitive regions near primary auditory cortices in individual participants. Individual cortical maps showing the locations of vocalization-sensitive cortices with respect to tonotopically-organized regions (Primary Auditory Cortices, PAC). Tonotopic organization (dotted outlines) occurred primarily along regions within and surrounding Heschl's gyrus (HG). Areas activated by the M>A, F>M, and E>M functional contrasts from Fig. 5-1 are highlighted. Cortex that was preferentially sensitive to mimicked versions of animal vocalizations versus corresponding real-world animal vocalizations (M>A, yellow) often occurred near or in some instances overlapped an individual's PAC. More lateral regions along the superior temporal gyrus (STG) and within the superior temporal sulcus (STS) often showed preference to neutral foreign (F>M, red) and English phrases (E>M, blue) over non-stereotypical human-mimicked animal vocalizations.

TABLE 5-1

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

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

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

CHAPTER 6:

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

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

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

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

This invited review was submitted for peer review in Hearing Research on November 8th, 2012.

ABSTRACT

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

INTRODUCTION

Vocalizations represent some of the most complex sounds of the natural world. The acoustic signals of even very short utterances can be rapidly processed to extract distinct meaning. This can include alerting the listener to danger, a mate, or food: More specific socially relevant information such as the identity of the source (e.g. species, gender, or specific individual), its intent, health status, or emotional state in some instances also be quickly surmised. The auditory systems of vocalizing mammals, notably including humans and non-human primates, develop to rapidly decompose incoming vocal communication signals (on second or sub-second timescales), utilizing multiple hierarchical processing stages from the brainstem to higher order auditory cortices.

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

Much of the early work that described mammalian auditory systems has incorporated the use of acoustically "simple" stimuli, including pure tones, band-pass noise, amplitude-modulated tones, and harmonic complexes. The major benefits of using simpler stimuli to elucidate the signal processing architecture and function of these networks are obvious; simple sound stimuli permit the design of very exact and controlled experimental manipulations that produce physiological responses with greater interpretable power. While these experiments have garnered much information about the function of hierarchically organized auditory networks, they generally do not reflect the nature of sounds that are experienced in real-world situations. Naturalistic sounds like vocalizations can be composed of extremely nuanced combinations of acoustic

phenomenon. This quality of vocalization signals becomes especially problematic when attempting to design precise experiments that can produce generalizable results. By probing the neuronal/auditory system with natural (often behaviorally relevant) stimuli, which it has arguably become optimized to process not only give insight into its respective operation, but are likely to reveal more cross-modal "whole-brain" physiological and behavioral responses.

While canonical auditory regions (i.e. primary auditory cortices (PAC), belt and parabelt regions) obviously play fundamental roles in vocalization processing, the influence of other "non-auditory" cortical and cognitive systems are increasingly becoming necessary to consider and model when examining how the auditory system extracts or derives meaningful representations for naturalistic vocalization stimuli. For instance, vocalizations often lead to assessments of the emotional and intentional states of other animals (conspecific or otherwise). This implies that the perception of vocal communication sounds tap into motor systems, including mirror neuron systems (Rizzolatti et al., 1996), and into introspective systems, including posterior and anterior insular systems (Craig, 2009), thereby entailing rather widespread cortical networks and pathways.

This prospective review considers two major themes: First, the more advanced communication abilities that humans have with non-speech utterances should presumably be present in some capacity in non-human primates (Fitch, 2011). Thus, one focus is to examine a comparison of cortical processing of natural vocalizations across primate species, especially great apes. Second, a listener's ability to attain a sense of "meaning" behind communicative utterances, and subtleties therein, requires extensive periods (years) of learning. Since much of the foundation of processing pathways and networks that are ultimately recruited for sound perception may develop in early stages of life, another focus of this review will be an analysis of fetal to early childhood human neurodevelopment in vocal signal processing. Based on these two themes, we address future research directions that we feel will facilitate significant advances in our understanding of basic hearing perception mechanisms. These advances in turn will contribute to a better understanding of vocal communication impairments, leading to more targeted therapeutic treatments, and to methods for developing more intelligent

biologically-inspired hearing prosthetics. In the following sections we address non-speech utterances as a gross-level category of vocalizations, cortices that are sensitive to human (conspecific) voice, and low-level acoustic signal processing of vocalization features. This is followed by a prospective review of non-human primate studies and by human studies involving the development of the auditory system at very early stages of life.

Utterances, paralinguistic signals, and non-speech

Throughout human evolution, spoken language perception has arguably become the most important function for the human auditory system. According to one set of linguistic theories, speech acts and utterances can be divided into four main categories (Austin, 1975). This includes *locutionary* utterances (speech), which convey semantic information and are expressed through use of many different acoustic signal forms ranging from simple phonemes, to words, to grammatically complex word combinations, as evidenced by the 6000 or so language systems currently spoken on our planet. However, humans commonly produce and hear other forms of non-locutionary communicative utterances on a daily basis, including many that transcend language boundaries. This includes *phatic* expressions to convey social information such as emotional status (e.g. a wince revealing pain, or a grunt of obeisance), *perlocutionary* expressions that are intended to cause a desired psychological consequence to a listener (e.g. tone in voice to persuade, convince, scare, or inspire), and *illocutionary* expressions wherein the utterance itself conveys the idea that the speaker will undertake an obligation (e.g. promising, ordering, greeting, warning, congratulating). Given that the auditory system must develop to accommodate processing of these types of acoustic cues, social interactions are likely to be critical for normal development. Germane to this idea is the hypothesis that social interaction is also essential for natural speech learning (Kuhl, 2007),

The use of spoken language to examine acoustic signal processing and general mechanisms mediating hearing perception has encountered a number of problems. For instance, congenitally deaf individuals can also easily acquire locutionary communication skills in the form of sign language and written language. Thus, there are likely to be a myriad of processing stages and hierarchies that link the complexities of language processing networks with those more central to hearing perception. In other words, vocal communication and language systems need not be mutually dependent on one another. While humans are the only species known to fully process, extract, and comprehend locutionary acoustic information (language), other vocalizing social species, such as monkeys and great apes, presumably have evolved to utilize some of these other non-

locutionary classes of utterances, which may be essential for effective social communication. Thus, to understand more rudimentary cortical mechanisms for processing communicative vocalizations, researchers have increasingly been investigating the processing of non-verbal vocalizations and paralinguistic signals (e.g. calls, grunts, coughs, sighs, etc.) not only in humans, but also in non-human primates. The use of non-locutionary utterances as behaviorally relevant, naturalistic stimuli permits more direct comparisons across species, wherein identical sets of sound stimuli can be used, as addressed in a later section. As a result of this basic approach, considerable interest remains in the pursuit of characterizing cortical regions or pathways that may show sensitivity or selectivity to the voice or vocal tract signal attributes that may be inherent to a given species—that is, sensitivity to "conspecific" vocalizations.

Cortical sensitivity to human voice

Cortical regions sensitive to human (conspecific) voice (or speech) have been traditionally identified bilaterally within the superior temporal sulci (STS) when examining responses to stereotypical sounds, including speech, animal vocalization, environmental sounds, and non-verbal sounds such as coughs, sighs, and moans (Belin et al., 2000, Belin et al., 2002, Fecteau et al., 2004, Lewis et al., 2009). However, more recent studies consider these STS regions to represent "higher-order" auditory cortices that function more generally as substrates for auditory experience, showing activity to artificially constructed non-vocal sounds after perceptual training (Leech et al., 2009, Liebenthal et al., 2010). The STS regions may not function in a domain-specific manner solely for vocalization processing. Rather, humans may typically develop to become "experts" at processing voice and speech signals, which consequently leads to recruitment and development of circuits in the STS that compete to process those signals in a domain-general manner (Pascual-Leone and Hamilton, 2001). Additionally, some of the samples of vocalizations and control natural sounds used to test for voice sensitivity have included broadly defined "categories", which may contain more subtle sub-categories upon which the auditory system is organized.

Using a novel class of *non-stereotypical* vocalization sounds, human-mimicked animal vocalizations, Talkington et al. reported a left hemisphere dominant activation to naturally produced sounds unique to the human *conspecific* vocal tract (Talkington et al., 2012). This was contrary to previous findings that described human-voice or species-specific sensitivity as a right hemisphere dominant or bilateral function (Belin et al., 2000, Fecteau et al., 2004). This "category" of vocalization sounds allowed them to probe intermediate cortical networks that show fine-grained sensitivities to the acoustic subtleties of the human vocal tract. Moreover, their results suggest that the cortical pathways supporting vocalization perception are initially organized by sensitivity to the human vocal tract in stages prior to the left STS. This and other studies have started examining different categories of calls, either across or within primate species, as addressed in section 5. These types of studies have consequently led to a resurgence of

the search for auditory cortices showing sensitivity to specific bottom-up "lower-level" acoustic signal attributes that may be inherent to communicative utterances and sub-classes therein. Such signal attributes may serve as primitives, which early and intermediate stages of the auditory system use to rapidly process sound, as addressed next.

Acoustic signal processing of vocalization

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

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

In humans, sensitivity to harmonic content (defined by a harmonics-to-noise ratio; HNR) as been interpreted in the context of such models. For instance, Lewis et al reported cortical regions showing parametric sensitivity to the HNR values of artificially constructed iterated rippled noises (IRNs; Figure 6-2, green) and of animal vocalizations (blue) (Lewis et al., 2009). These stages of HNR-sensitivity in humans were juxtaposed between tonotopically-defined regions (yellow) and STS regions sensitive to speech (purple). Conceivably, specific combinations of tonotopic outputs could converge to form cortical networks that are sensitive to harmonic qualities, representing intermediate stages of vocalization processing. Interestingly, different categories of animal calls and utterances could in part be grouped based on harmonicity (HNR values) of different types of vocalizations (Figure 6-3). The search for regions "selective" for processing conspecific vocalizations may critically depend on the specific category or sub-category of vocalization sound(s) under consideration. Additionally, revealing voice-sensitive regions may further depend on specific task factors, reflecting how the vocal information is to be used. This may also help to reveal the signal features that lead to difference in processing between the left and right hemispheres. Collectively, the results from identifying bottom-up signal processing should impact the design of intelligent hearing aids and implants, which may enhance or retain such features relative to background acoustic noise (Coath et al., 2005, Coath and Denham, 2005, Coath et al., 2008). The issue of categories of non-speech vocalizations will also apply to the study of non-human primate auditory systems, which are considered next.

Non-human primate cortical vocalization processing

Given their evolutionary proximity to humans, data concerning the structures and pathways that are involved in the perception and processing of naturalistic vocalizations are particularly relevant to discussions of human language origins. Typically, functional neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have been employed to visualize activity during passive listening to conspecific vocalizations in macaque monkeys (Gil-da-Costa et al., 2004, Poremba et al., 2004, Gil-da-Costa et al., 2006, Petkov et al., 2008). However, a relatively early study with Japanese macaques, *Macaca fuscata*, is relevant to this discussion as well (Heffner and Heffner, 1984). Here, the researchers evaluated the ability of their subjects to discriminate between two variants of a species-specific call type before and after unilateral or bilateral ablation of the superior temporal gyrus. Whereas the discriminative performance of monkeys who sustained unilateral ablation of the right superior temporal gyrus was unaffected, those subjects with unilateral ablation of the left superior temporal gyrus were temporarily unable to complete the auditory discrimination task. Those individuals that received subsequent bilateral ablations never recovered their discriminative ability. The authors concluded that perception of species-typical vocalizations is mediated in the superior temporal gyrus with the left hemisphere playing a predominant role (Heffner and Heffner, 1984). It is noteworthy that monkeys who received an ablation to the left superior temporal gyrus (but not the right), did subsequently regain their discriminative abilities.

More recently, Poremba, et al., (2004) utilized positron emission tomography (PET) to determine whether or not increased neuronal metabolic activity is observed in the superior temporal gyrus of rhesus monkeys during passive listening to a variety of auditory stimuli including conspecific vocalizations. The authors reported that only rhesus monkey vocalizations (and not phase-scrambled conspecific vocalizations, human vocalizations, ambient background noise, or environmental sounds) resulted in significantly greater metabolic activity in the left dorsal temporal pole of the superior temporal gyrus. In a second study, Gil-da-Costa and colleagues (2006) similarly utilized

PET to visualize cerebral metabolic activity in the rhesus monkey brain during the presentation of conspecific vocalizations. In contrast to the results reported by Poremba et al., (2004), the authors report significant activation in the posterior region of the temporal lobe (their 'temporoparietal' (Tpt) area) in response to passive listening to conspecific vocalizations when compared to non-biological sounds. Significant activation was also observed in the monkey ventral premotor cortex. No differences between the two conspecific call types (coos vs. screams) were observed, but both vocalizations evoked greater activity in the monkey temporoparietal area, ventral premotor cortex, and the posterior parietal cortex as compared to nonbiological sounds.

Romanski and colleagues (2004) examined the response properties and selectivity of neurons in the rhesus macaque ventrolateral prefrontal cortex (vlPFC) to the presentation of species-specific vocalizations (Romanski et al., 2005). The authors were interested, among other things, in whether or not certain neurons in the vlPFC would respond similarly to morphologically distinct calls that have similar functional referents. They report that of the cells they recorded from, most were selective for more than one vocalization type (average of 3). However, they found that the neurons were likely responding to calls with similar acoustic characteristics and signal features, as opposed to similar functional referents.

While a considerable number of studies have utilized macaque monkeys to examine processing pathways for conspecific vocalizations, surprising little work has been done in other primate species, notably great apes. However, Taglialatela et al., (2009) recently used PET to visualize cerebral metabolic activity in chimpanzees in response to passive listening to two broad categories of conspecific vocalizations, proximal vocalizations (PRV) and broadcast vocalizations (BCV) (Taglialatela et al., 2009). PRV are relatively low intensity vocalizations typically produced by individuals in direct proximity of one another, and are seemingly directed towards these individuals. BCV are much higher amplitude calls as compared to the PRV, are also produced by individuals in the presence of conspecifics, but appear to be directed to distant individuals. Two important findings emerge from this study. First, right-lateralized activity was observed in the posterior temporal lobe, including the planum temporale, when chimpanzees were presented with PRV (but not time-reversed conspecific calls). However, similar lateralized activity was

not observed during passive listening to BCV. These results suggested that a functional distinction may exist between calls classified broadly as BCV and PRV that corresponds to differences in their processing in the chimpanzee brain. Thus, these findings complement the human literature addressed in earlier sections, suggesting that the primate auditory system may develop distinct pathways for processing different categories of socially relevant vocalizations.

Previous behavioral work has found evidence of group-level structural variation in the pant hoot vocalizations (Taglialatela's BCV category) produced by both wild and captive chimpanzees (Arcadi, 1996, Marshall et al., 1999, Crockford et al., 2004). For example, Crockford and colleagues (2004) report structural differences in the pant hoots of male chimpanzees living in neighboring communities, but not between groups from a distant community. These results could not be accounted for by genetic or habitat differences suggesting that the male chimpanzees may be actively modifying the structure of their calls to facilitate group identification (Crockford et al., 2004). Therefore, chimpanzees may be using pant hoots as a means for discriminating among familiar and unfamiliar individuals.

Secondly, although some consistencies are evident between the results reported in the single chimpanzee study (Taglialatela et al, 2009), and those published previously from monkey species (Gil-da-Costa et al. 2006; Poremba et al. 2004, Petkov et al., 2008), important differences are apparent. Consistent with what has been reported for monkeys, right-lateralized activity is observed in the chimpanzee superior temporal gyrus in response to conspecific vocalizations. However, Poremba et al., (2004) reported right-lateralized activity in posterior regions of the superior temporal gyrus to all auditory stimuli, and left-lateralized activity in the temporal pole only in response to conspecific vocalizations (Poremba et al., 2004). Furthermore, Gil-da-Costa et al. (2006) reported significant activation in response to conspecific vocalizations in monkey temporoparietal, posterior parietal, and ventral premotor cortex, but did not observe any lateralized activation, even in the left temporal pole as reported previously by Poremba et al. (2004). Petkov and colleagues (2008) identified a region of auditory cortex in the macaque brain that is selectively active during the perception of species-specific vocalizations (Petkov et al., 2008), and this region was located in the anterior temporal lobe. When compared to

the data from chimpanzees (Figure 6-4), significant activation was observed in the anterior portions of the right superior temporal gyrus following the presentation of both BCV and PRV (compared to time-reversed conspecific vocalizations). Taglialatela and colleagues did not specifically aim to identify a conspecific-sensitive region in chimpanzees in their study. Therefore future studies should seek to specifically determine if these anterior portions of the chimpanzee superior temporal gyrus are selective for conspecific vocalizations and/or different categories of natural calls. In addition, Petkov et al., (2008) reported that this anterior monkey "voice" region was specifically sensitive to the vocalizations produced by *familiar* conspecifics. Such investigations have yet to be carried out in chimpanzees, but would be important for a) reconciling the human and monkey findings (addressed above), and b) obtaining a clearer picture of the more recent phylogenetic changes associated with the evolution of spoken language processing in the human brain.

Of course, the communicative behaviors of most primate species – including human language - typically span more than one sensory modality, and therefore include signals that go beyond the auditory stream. To this end, a number of researchers have aimed to examine auditory/visual processing in response to the presentation of conspecific vocalizations and their concomitant facial expressions. The results of these studies primarily indicate that multimodal communicative information (i.e. monkey vocalizations and the corresponding facial expressions) appear to be integrated in the rhesus monkey VLPFC as well as in auditory cortex (Sugihara et al., 2006; Ghazanfar et al. 2005). For example, Sugihara et al., (2006) presented conspecific vocalizations with or without accompanying video/still images of the face of a vocalizing rhesus macaque (Sugihara et al., 2006). They found multisensory neurons in the rhesus macaque VLPFC that exhibit enhancement or suppression in response to the presentation of face/vocalization stimuli. Romanski (2012) has proposed that the integration of vocalizations and faces that occurs in the macaque prefrontal cortex may represent an evolutionary precursor to the processing of multisensory linguistic input in the frontal lobe of the human brain(Romanski, 2012). This is an intriguing hypothesis, particularly when considering recent data indicating that both spoken language and symbolic gestures are processed by a common network in humans that includes the inferior frontal gyrus and posterior

temporal lobe (Xu et al, 2009). Thus, the picture that emerges is that inferior frontal regions as well as temporal cortex in non-human primates is involved in constructing meaning from incoming signals in multiple modalities. However, future studies with chimpanzees will be critical to evaluate this hypothesis.

Along these lines, another potentially fruitful area of research will likely be the increased incorporation of more than one species in a single experimental paradigm. In addition to developmental studies (addressed in the following section), the roots of human auditory structures, function, and skill can be investigated best by examining homology that exists in closely related species. This is not to imply that the anatomy and physiology of closely related extant species represent the exact conditions that were necessary to form the bases of human function; these species have also continued to evolve away from their precursors concomitantly with humans. Recently, Joly et al. performed near identical experiments in two species, humans and rhesus monkeys (Joly et al., 2012). During their respective experiments, members of both species heard speech sounds (French), non-verbal emotional human vocalizations, monkey vocalizations of different emotional valence, and spectrally "scrambled" versions of all stimuli.

Studies such as these are important for they provide a) an opportunity to directly compare the processing of auditory signals by different species, and b) they consider the fact that just as all human language utterances may not be functionally equivalent, the same may be true of nonhuman primate vocalizations. Therefore, researchers will be challenged to closely examine the actual vocal communicative behavior of the species under study and move beyond a "species-specific" vocalization model to one that examines different categories or classes of calls in a meaningful and ecologically relevant way.

Vocalization processing in human infant auditory circuits

Most auditory cortical mapping and other physiological studies in both human and non-human primates thus far have largely been performed with adult subjects. The operation of very efficient and streamlined mature systems may act to "conceal" or mask critical intermediate auditory processes that help lead to coherent percepts. Given technical advances in human neuroimaging (i.e. fMRI, EEG/ERP, infrared (fNIRS)), the immature auditory systems of developing humans (and non-human primates) represent an increasingly valuable context for advancing our understanding of vocalization processing; investigating these networks as they are forming will garner greater understanding of their eventual mature forms and processes. Behavioral and physiological responsiveness and preferences for specific acoustic information, namely human vocalizations and speech, is evident in human fetuses. In particular, the strongest behavioral responses in fetuses (e.g. fetal heart rate change) and physiological changes in infants (e.g. infant sucking) were to maternal vocalizations; they also exhibit *preferential* responses to their mother tongue early in life (DeCasper and Fifer, 1980, Moon et al., 1993, DeCasper et al., 1994a, Kisilevsky et al., 2003, Kisilevsky et al., 2009, Beauchemin et al., 2011, Sato et al., 2012). These preferential responses to maternal voices and other vocal signals argue for the presence of cortical networks that are (or become) optimized for processing these acoustic signals (see below). Whether these early preferences in auditory circuits are genetically ore epigenetically predetermined (Werker and Tees, 1999) to some degree (domain-specific) or are mostly experience-dependent remains largely unknown, and represents an exciting topic of future study. Parsimony, however, argues for a combination of both. Auditory (and other sensory systems) may begin with "experience-expectant" network structures (proto-networks) and processes that eventually give way to more "experience-dependent" organizational activity. Regardless, neuroimaging methods have begun to reveal the structure and function of early auditory communication processing networks during infant development.

Neuroimaging and other neurophysiological methods are increasingly revealing the early network structures and processing stages that emerge in the developing auditory

system. Developmental neuroscience within the context of vocalization processing has thus far begun to contribute greatly towards hemispheric specialization. Previous findings in fully developed adult subjects generally have favored models that posit left/right hemisphere differences defined by different temporal processing timescales (Zatorre et al., 1992, Poeppel, 2003). The processing of rapid acoustic signal changes is thought to be a left hemisphere dominant or even bilateral function (e.g. consonant sounds)while the right hemisphere shows the greatest sensitivity to spectrally-stable envelope-level structure (e.g. vowel-like speech sounds or sounds containing strong prosodic cues), though findings in this research are sometimes conflicting (Boemio et al., 2005, Hickok and Poeppel, 2007, Obleser et al., 2008, Overath et al., 2008, Zatorre and Gandour, 2008). One group has tested this theory in infants ranging from three days old to three and six months old with temporally modulated noise samples using various methods including EEG and functional near-infrared spectroscopy (fNIRS) (Telkemeyer et al., 2009, Telkemeyer et al., 2011). FNIRS, sensitive to changes in hemoglobin and deoxyhemoglobin concentrations indirectly evoked by local neural activity, is increasingly being used to measure cortical responses in infants due to its relative non-invasiveness and ability to be used in more natural settings (Quaresima et al., 2012). Responses across all of these age groups remained relatively stable; sounds with rapid modulations produced fairly bilateral response patterns whereas the strongest responses to slowly modulating stimuli were dominant in the right hemisphere.

Evidence for hemispheric specialization for certain types of vocalizations in infants and young children is growing and is arguably critical for fully understanding the operations (and potential dysfunctions) of these circuits. When specifically investigating speech sounds, it appears that there may already be a left hemisphere processing dominance even in newborn babies (Dehaene-Lambertz et al., 2002, Pena et al., 2003, Sato et al., 2012). Future studies, especially those including infants, will require taking gestational versus post-natal ages into account. Gestational age has been shown to alter auditory responses to speech sounds, perhaps reflecting “critical periods” of auditory development (Caskey et al., 2011, Key et al., 2012, Pena et al., 2012). The left hemisphere preference for intelligible speech is also reliably shown in four year olds;

segmental (phonological) and suprasegmental (prosodic) speech information processing appears to be clearly defined along hemispheric boundaries (Wartenburger et al., 2007).

Auditory neurophysiological evidence from infants concerning the processing of prosodic cues in utterances thus far points to a right hemisphere dominance for this function. The right hemisphere in four year olds seems to present a clear dominance for processing prosodic envelope-level information in vocalization signals (Wartenburger et al., 2007), consistent with dual-pathway models of language function (Friederici and Alter, 2004). Other studies have shown that this prosodic preference in the right hemisphere exists at numerous developmental time points. Grossman et al. showed a right hemisphere preference for human voice sounds; this effect was amplified when considering network modulations caused by different categories of prosodic cues (Grossmann et al., 2010). Even earlier in the developmental timeframe, Cheng et al. has reported right-lateralized mismatch ERP responses in newborns (less than five days old) between speech samples of varying emotional valence (Cheng et al., 2012); responses to negative valence stimuli were especially strongest, perhaps reflecting an evolutionary processing bias for threatening stimuli (Vuilleumier, 2005). These finding generally corroborate the results from similar studies performed in adults showing stronger right hemisphere activation for emotionally evocative stimuli (Grandjean et al., 2005). Many studies investigating these functions often include linguistic content which may preclude stronger lateralization results; it is likely that shared overlapping functions exist across the hemispheres. Experimental design of prosodic cues studies also plays a large role in lateralization of results (Kotz et al., 2006). Additionally, functional and lesion studies suggest that the processing of emotional speech and emotional non-verbal stimuli are predominantly, though not exclusively, governed by the left and right hemispheres, respectively (Crosson et al., 2002, Ethofer et al., 2006b, Pell, 2006). Within this framework, preferential right temporo-parietal responses to the prosodic pitch contours of speech are seen in three month old infants and are thought to represent facilitation of burgeoning left hemisphere networks during the learning of syntactic speech structures (Homae et al., 2006). This idea forms the basis of prosodic bootstrapping theories of language acquisition (Gleitman and Wanner, 1982, Jusczyk, 1997) and may explain infant preferences for vocalizations with strong prosodic cues.

In addition to strong preferences for maternal voices and speech in general, infants generally show behavioral biases for many relatively simplistic harmonic vocalization sounds that contain strong prosodic cues, oftentimes occurring in the form of "motherese speech" in social settings. These utterances usually have elongated and exaggerated vowels or vowel-like sounds and often have no intelligible speech content. A preference for acoustically "simple" vocalizations is not only seen in the behavior and physiology of infants themselves, but also manifests reciprocally in the behavior of adults. Most people revert to producing motherese or similar vocalizations when in the presence of an infant or toddler, ostensibly for the purpose of pleasing them. Laughter, smiling, and other positive responses in the infant provide early non-verbal feedback to adults that likely encourages more bonding interactions (Caron, 2002, Mireault et al., 2012), a social phenomenon that may have behavioral and acoustic evolutionary origins (Gamble and Poggio, 1987, Knutson et al., 2002, Vettin and Todt, 2005, Davila Ross et al., 2009). These interactions not only promote bonding and other important social relationships but may also be useful for initiating sources of regular acoustic information useful for building and refining vocalization processing networks in the developing auditory system. Indeed, understanding the role of a developing infant's social environment during auditory development will crucial to understanding the vocal perception (Kuhl, 2007, 2010).

The functional architectures of these auditory communication networks are presumably formed and constrained in part by their respective physical architectures. Understanding the functional aspects of auditory development will be greatly enhanced by also describing concomitant changes in anatomical features. Leroy et al. performed an extensive cortical maturation study in an infant cohort over the first several months of life (Leroy et al., 2011). Calculating a maturation index (MI) derived from T2-weighted magnetic resonance signals, the authors demonstrated that portions of the STS (especially the ventral banks) are of the more slowly developing perisylvian cortical regions, especially when compared to frontal regions. The right STS showed earlier maturation when compared to the left STS however, consistent with other structural and genetic right-sided asymmetries found in the early developing brain (Sun et al., 2005, Hill et al., 2010). Conversely, the authors show also maturation correlations between white and gray

matter in regions that corresponding to the frontal and posterior territories of the left arcuate fasciculus. The left arcuate fasciculus that is thought to be myelinated more rapidly in the left hemisphere of infants (Dubois et al., 2009) and functional activity in corresponding left posterior STG/STS regions of infants often show the highest correlations with age (Grossmann et al., 2010, Blasi et al., 2011). This area may correspond to a postero-temporal region of cortex called Spt that is instrumental for sensory-motor integration, specifically with regard to speech processing (Hickok et al., 2009). The arcuate fasciculus is also proposed to be the structural foundation for the phonological loop and human language faculties at large (Aboitiz et al., 2010). A comparative anatomical study involving macaques, chimpanzees, and humans highlighted the increased cortical connectivity of "language-supporting" regions in humans that may have spurred the development of extensive language skills (Rilling et al., 2008).

Recently, it has been suggested that general voice-sensitivity emerges in the infant brain between four and seven months of age predominantly in posterior right temporal regions (Grossmann et al., 2010). The "voice" category in this study included speech and non-speech signals that were contrasted against "non-voice" stimuli generally used in adult studies (cars, airplanes, telephones, etc.) (Belin et al., 2000). The findings from Grossman et al. may have represented the infant homolog to adult Temporal Voice Areas (TVA); these areas have traditionally been defined with broadly defined vocalization and non-vocalization categories (ibid). Blasi et al. has also demonstrated a right hemisphere bias for processing neutral non-verbal vocalizations versus non-voice environmental sounds that would likely be familiar to infants in the right anterior superior temporal cortex (Blasi et al., 2011). However, another fNIRS study using similar stimuli from Blasi et al. described age-dependent preferential voice responses in bilateral temporal cortices (Lloyd-Fox et al., 2012). These studies investigated voice-sensitivity in a manner that was similar to experiments performed in adults. While all of the findings do show varying degrees of human voice sensitivity, contrasting activity to human vocalizations (verbal or non-verbal) with activity to sounds that are clearly not products of human vocal tracts (man-made mechanical objects, environmental sounds, etc.) likely produces results that only reflect their extreme categorical differences. Namely, these contrasts

cross numerous categorical boundaries both acoustically and conceptually (Engel et al., 2009). This would likely conceal sub-threshold activity in more intermediate vocalization-specific networks where more subtle acoustic and conceptual distinctions are realized. Additionally, within the voice categories, the representative stimuli also crossed many categorical boundaries (e.g. speech vs. non-speech, native vs. foreign speech, emotional vs. neutral non-verbal vocalizations, etc.) which likely lead to very broad and relatively non-specific cortical network activations. Future work should utilize distinct yet closely related categories of vocalization sounds when describing regions and activity profiles that show human voice-sensitivity. Contrary to previous findings (Belin et al., 2000, Fecteau et al., 2004), Talkington et al. described strongly left-lateralized conspecific vocalization sensitive regions near primary auditory cortices using non-stereotypical human-mimicked animal vocalizations (Talkington et al., 2012).

While the focus of this review is centered on auditory system processing, it would not be prudent to completely ignore the multisensory nature of communication. Recently, Grossman et al. recorded ERPs from 4 and 8-month old infants as they watched dynamic audio-visual pairings of monkey faces and vocalizations as well as human-mimicked versions of the same stimuli in order to measure their capacity for multisensory integration and perception of vocalization production (Grossmann et al., 2012, Talkington et al., 2012). Similar to Talkington et al., the authors reasoned that using non-stereotypical stimuli in unfamiliar contexts provided for stronger tests and interpretations of neuronal mechanisms.

Comprehensive examinations of infants and toddlers will provide fundamental details regarding the anatomical and functional principles that become fully instantiated in mature neuronal circuits. Critical neurobehavioral milestones during development can be paired with concomitant changes in anatomy and function as nascent auditory networks and "proto-networks" form. This will promote the formation of more direct and accurate models for describing the relationships between anatomical structures, physiology, perception, and higher-order cognitive functions. These improved models will greatly aid in determining the etiology of auditory-related communication disorders as well as provide critical information for evidence-based therapies that can be implemented during specific developmental periods.

FIGURES

FIGURE 6-1

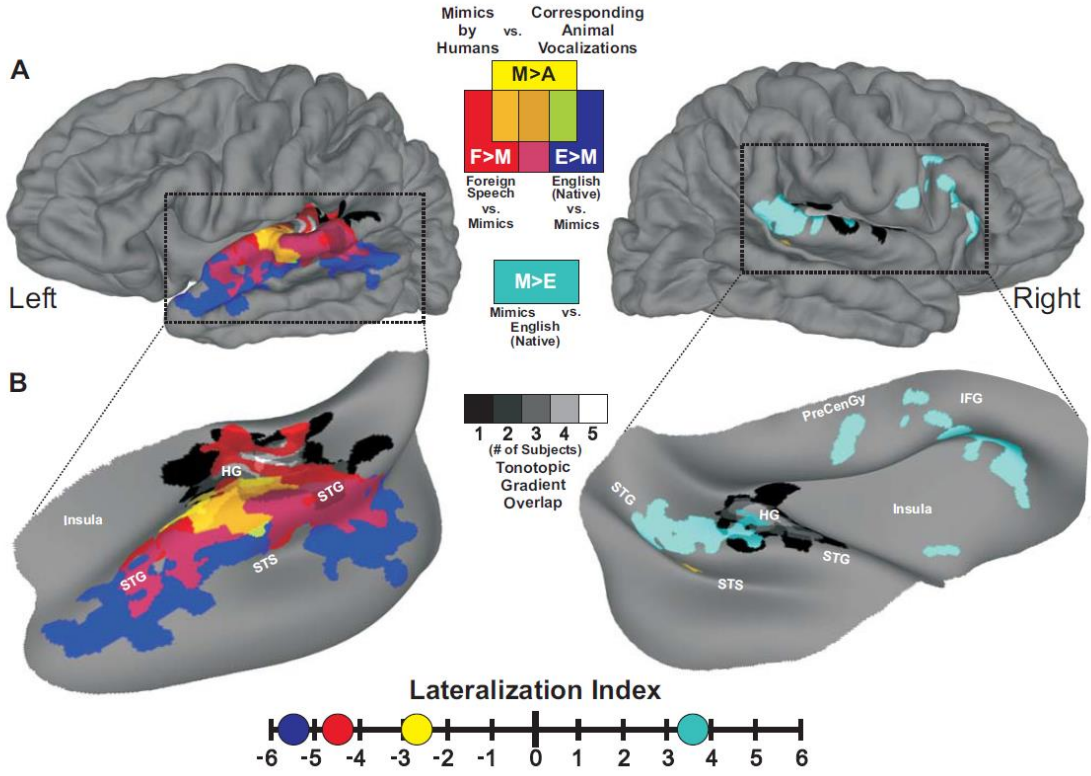


FIGURE 6-1. Conspecific vocalization processing hierarchy in human auditory cortex. A. Group-averaged (n=22) functional activation maps displayed on composite hemispheric surface reconstructions derived from the subjects. **B.** To better visualize the data, we inflated and rotated cortical projections within the dotted-outlines. The spatial locations of tonotopic gradients from five subjects were averaged (black-to-white gradients) and located along Heschl's gyrus (HG). Mimic-sensitive regions (M>A) are depicted by yellow hues, sensitivity to foreign speech samples versus mimic vocalizations (F>M) is depicted by red hues, and sensitivity to native English speech versus mimic vocalizations (E>M) is depicted by dark blue. Regions preferentially responsive to mimic vocalizations versus English speech samples (M>E) are depicted by cyan hues. Corresponding colors indicating functional overlaps are shown in the figure key. All data are corrected for multiple comparisons to $p<0.05$. This illustration has been adapted with permission from Talkington et al., (2012).

FIGURE 6-2

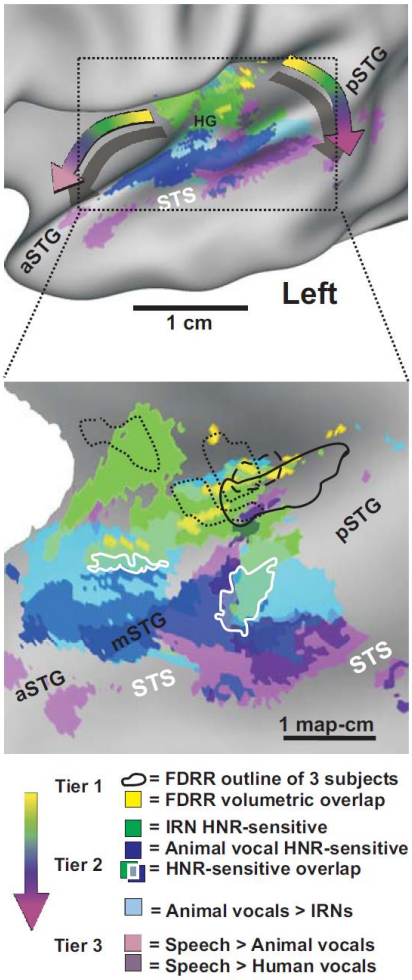


FIGURE 6-2. Location of cortices parametrically sensitive to harmonic content (HNR-sensitive) relative to human vocalization processing pathways and tonotopically-organized regions that estimate the location of primary auditory cortices. Data are illustrated on slightly inflated (upper panel) and "flat map" (lower panel) renderings of averaged human cortical surface models. Data are all at $\alpha<0.01$, corrected. Refer to key for color codes. Intermediate colors depict regions of overlap. The curved "rainbow" arrows depict two prominent progressions of processing tiers showing increasing specificity for the acoustic signal features present in human vocalizations. Overlap of IRN (green) and animal vocalization (blue) HNR-sensitivity are indicated (white outlines). This illustration and caption adapted from from Lewis et al., (2009).

FIGURE 6-3

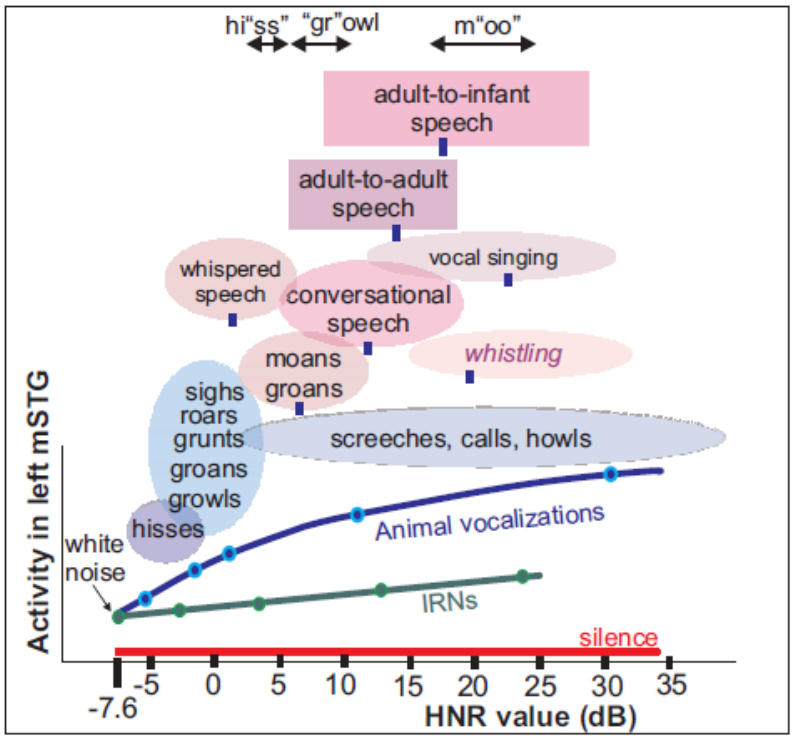


FIGURE 6-3. Typical HNR value ranges for various sub-categories of mammalian vocalizations. Oval and box widths depict the minimum to maximum harmonic content (HNR values) of the sounds sampled, charted relative to the group-averaged HNR-sensitive response profile of the left mSTG (e.g. from Fig. 6-2). Green and blue dots correspond to IRN and animal vocalization sound stimuli, respectively, from Fig. 6-2. Blue ovals depict sub-categories of animal vocalizations explicitly tested. Ovals and boxes with violet hues depict sub-categories of human vocalizations (12-18 samples per category), and blue tick marks indicate the mean HNR value. For instance, conversational speech had a mean of +12 dB HNR, within a range from roughly +5 to +20 dB HNR. Adult-to-adult speech (purple box; mean = +14.0 dB HNR) and adult-to-infant speech (violet box; mean = +17.2 dB HNR) produced by the same individual speakers were significantly different (t-test $p<10\text{-}5$). Stress phonemes of three spoken onomatopoetic words depicting different classes of vocalizations are also indicated. This illustration and caption adapted from Lewis et al., (2009).

FIGURE 6-4

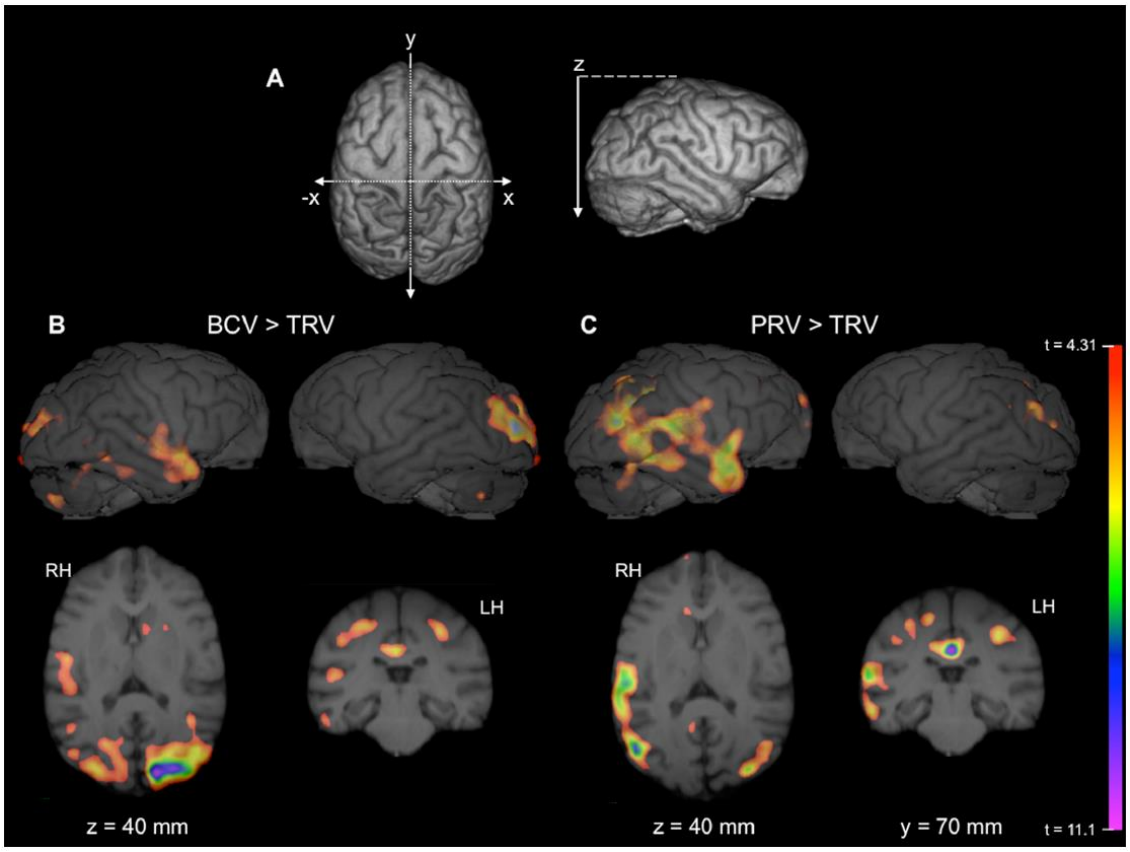


FIGURE 6-4. Significant areas of activation in chimpanzees for (**B**) broadcast vocalizations (BCV) relative to time-reversed vocalizations (TRV) and for (**C**) proximal conspecific vocalizations (PRV) relative to TRV. Top images are 3D rendered MR images of chimpanzee right (RH) and left hemispheres (LH) with significant ($t \geq 4.31$) PET activation overlaid. This illustration and caption adapted from Taglialatela et al., 2009.

CHAPTER 7:

The temporal dynamics of conspecific vocalization processing in human auditory cortices

ABSTRACT

The human auditory system is likely most sensitive to vocalizations produced by other humans (conspecifics). Presumably, the activity in early auditory cortical networks reflect processing preferences for these vocalizations. Previous studies have used stereotypical human-produced verbal and non-verbal vocalizations to investigate human voice sensitive cortical responses. By utilizing a novel category of non-stereotypical vocalizations, human-mimicked animal vocalizations, we have demonstrated early differential processing between human and animal-produced vocalizations using auditory evoked potentials. Specifically, the N1 responses to human-mimicked vocalizations are significantly greater than those produced by their corresponding animal vocalizations. This differential N1 response (approximately 75-135ms) precedes previous findings that claim species-specific vocalization processing in human auditory cortices occurs around 164ms. The current findings support previous fMRI findings using similar stimuli that revealed a left-lateralized conspecific vocalization sensitive region of auditory cortex near primary auditory cortices (PAC). Additionally, perceptual responses to these categories of sounds drive the amplitude of the later P300 response. Vocalizations perceived as human-produced generate greater amplitude P300 components than those sounds perceived as animal-produced. Collectively, these results suggest that preferential processing of conspecific vocalizations may occur as early as auditory cortical stages within or near PACs.

INTRODUCTION

The human auditory system is capable of extremely rapid sound decomposition and processing. Language faculties eloquently demonstrate these auditory skills; a plethora of information, emotion, and intent can be relayed from speaker to listener in a matter of a few seconds. Electrophysiological methods allow us to probe the underlying processes that subserve auditory communicative functions on biologically-relevant time scales. Complementary to the spatial processing hierarchies in auditory cortices that can be deduced using fMRI, EEG permits investigations of the temporal processing of complex vocalization sounds.

Face-sensitive and voice-sensitive regions have been identified in human cortex with fMRI (Kanwisher et al., 1997, Belin et al., 2000). Motivated by these studies, scientists have also investigated the temporal analogs of those findings. Face-sensitive ERP components have been found that produce the largest amplitude responses to face stimuli, occurring at approximately 170ms after stimulus onset (so called "N170 responses") (Bentin et al., 1996). Similar investigations in the auditory modality investigate the presence of "voice-sensitive" AEP responses. Early studies that compared AEPs between instrument-produced sounds described a "voice-specific response" (VSR) occurring at approximately 320ms after stimulus onset (Levy et al., 2001). However, this response seems to be very sensitive to the attentional states of participants (Levy et al., 2003); in this regard, the VSR may represent a neurophysiological marker for the allocation of attentional auditory resources.

Later studies that investigated responses to vocal adaptation effects and the processing of paralinguistic acoustic features of vocalizations described responses occurring earlier than the VSR (Schweinberger, 2001, Lattner et al., 2003, Beauchemin et al., 2006, Zaske et al., 2009). These findings were followed by those of Charest et al. that described a "Fronto-Temporal Positivity to Voices" (FTPV) at fronto-temporal electrode locations (e.g. FC5/6) occurring approximately 164ms after stimulus onset. The FTPV was identified by comparing AEP responses to voice sounds with responses to bird vocalizations and environmental sounds (Charest et al., 2009). Additional GFP findings from De Lucia et al. suggest a similar timeframe for species-specific vocalization

processing approximately 169-219ms after stimulus presentation (De Lucia et al., 2010). Similar to the rationale of Chapter 5, however, we reasoned that the stereotypical nature of the human vocalizations used in the aforementioned studies was not properly controlled.

This study incorporated short (180ms) animal vocalizations and human-mimicked versions of those stimuli to investigate the temporal processing dynamics of conspecific vocalization sensitivity in early auditory cortical circuits. We hypothesized that differential AEP responses between these two categories of sound would be reflected in the N1 component. This AEP component is thought to be generated near PACs (Näätänen and Picton, 1987) and would potentially reflect activity in the mimic-sensitive regions described in Chapter 5. Confirmation of our hypothesis would provide converging neurophysiological evidence (fMRI and EEG/AEP) of early auditory cortical networks that are optimized to process the acoustic qualities of conspecific vocalizations.

MATERIALS AND METHODS

Participants

We recorded EEG signals from eleven adult native English speaking participants (mean age: 25.5 yrs. ± 4.4 s.d.; five female; ten right handed, one ambidextrous). All participants were free of neurological, audiological, or medical illness, and were paid for their participation. Informed consent was obtained following guidelines approved by the West Virginia University Institutional Review Board.

Stimuli

Animal vocalizations were sourced from various professionally recorded (sampled at 44.1 kHz) CD collections (Sound Ideas Inc. Richmond Hill, Ontario, Canada and The Hollywood Edge, Hollywood, CA). Human mimicked animal vocalization stimuli were recorded by vocal actors in a sound isolation booth using a Sony PCM-D1 recorder (sampled at 44.1kHz). Recording was performed in stereo, but all sounds were converted to mono and played binaurally to minimize spatial cues. Additionally, all sound stimuli were shortened to 180 ms and low pass filtered by 10kHz. The attacks of the recorded and sourced stimuli were left in their original states to preserve acoustic attributes that may be important for categorical processing. The attacks of the recorded and sourced stimuli were left in their recorded states to preserve acoustic attributes that may be important for categorical processing. A 1 ms $\cos^2$ ramp was applied to the end of each stimulus and the entire stimuli set was equated for root mean square (RMS) power. Stimuli were presented binaurally to subjects through electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax LTD., Gardena, CA) via Presentation software (version 11.0, Neurobehavioral Systems, Inc.) running on a Windows PC. Overall loudness of the stimuli was adjusted to a comfortable level for each subject.

Electrophysiology procedures

Sixty-four channel electroencephalographic (EEG) recordings were collected with NeuroscanSynAmps hardware, Scan 4.3 Acquire software, and Quik-Caps (Ag/Ag-Cl sintered electrodes; 10-10 system). Impedances were kept below 10kΩ at all electrodes. All scalp electrodes were referenced to the left-mastoid, as well as an electrode placed on the right mastoid. All data was re-referenced to the algebraic average of the left and right mastoid electrode recordings before any further processing or analyses (Luck, 2005). A 1 kHz sampling rate was applied to all channels and signals were filtered on-line from 0.05-200Hz.

Stimulus presentation procedures

All EEG recording occurred in a sound isolation booth to minimize acoustic and electrical interference. Each EEG session consisted of six total runs (two separate experiments; see below) lasting approximately six minutes apiece; each run contained 162 stimuli (81 animal vocalizations and 81 corresponding human-mimicked versions). Inter-stimulus intervals (ISI) were random and uniformly distributed between 2300-2700ms to minimize habituation and to allow enough time for subject responses during the latter half of the experiment.

During the first three runs, participants watched a muted subtitled movie of their choice. After these runs, the subjects were informed that they would be performing a discrimination task while fixating on the wall of the booth. Using a four-button Neuroscan response pad, subjects were asked to respond after the presentation of each stimulus to indicate whether it was produced by an animal or a human. Either the extreme left or right buttons corresponded to the respective vocalization categories; button designations were counterbalanced across the entire subject group and across genders. Subjects were instructed to attempt a high level of accuracy. Though rapid response latencies were not stressed, subjects were encouraged to respond before the next stimulus presentation.

Data Analysis

All analyses of EEG and ERP data were performed using the MatLab-based open-source software packages EEGLAB (version 10.2.5.8; (Delorme and Makeig, 2004)) and ERPLAB (www.erpinfo.org). Continuous EEG data from each subject were combined for each segment of the experiment (passive and task trials). High-pass filters (0.1 Hz) were applied to these concatenated EEG datasets to remove slow baseline fluctuations. Epochs were defined around event timestamps with 200ms pre-stimulus baseline periods and 1500ms post-stimulus periods; baseline correction was performed with the pre-stimulus periods. Epochs that exceeded ±100μV at any time point were rejected as artifact trials and not included in subsequent averaging.

N1, P2, and P300 amplitudes were defined as the average amplitudes in predefined timeframes (N1= 85-135ms; P2=160-200ms; P300=400-800ms) (Luck, 2005). N1 and P2 amplitudes were measured from frontal scalp electrodes Fz, F3, and F4. Global field power (GFP) measures were measured for the P300 (Lehmann and Skrandies, 1980). GFP represents a reference-independent field strength measurement across entire electrode montages; simply, they measure the strength of a given potential (Murray et al., 2008). N1, P2, and P300 amplitudes were entered into repeated measures ANOVAs to test for main effects of vocalization category (animal vocalizations or human mimics). Greenhouse-Geisser corrections were applied in cases where sphericity could not be assumed (Jennings and Wood, 1976). Also, all pairwise comparisons were corrected for multiple comparisons with Bonferroni correction.

RESULTS

Figure 7-1 shows the group-averaged (n = 11) N1-P2 AEP waveform complex for the passive task experiment. The N1 amplitudes in response to human-mimicked vocalizations was greater than those in response to the corresponding real world animal vocalizations (Avg. N1 values, animal: -2.09μV, SD = 1.66, human-mimics: -3.54μV, SD = 2.15). A main effect of vocalization category was seen for N1 amplitudes ($F_{1,10}$ = 18.992, P = 0.001), supporting our hypothesis that human-mimicked animal vocalizations would produce greater N1 average amplitudes. P2 amplitudes between the two categories of sound appeared to be comparable (Avg. P2 values, animal: 5.48μV, SD = 2.24, human-mimics: 5.23μV, SD = 2.13). Indeed, no main effect of category was seen in the P2 component ($F_{1,10}$ = 0.780, P = 0.398).

The second portion of the experimental sessions required that participants actively discriminate stimuli. After the presentation of each stimulus event, subjects responded in a 2AFC design whether they believed they had heard a vocalization produced by a human or an animal. Figure 7-2 shows the group-averaged GFP P300 responses, a response that is elicited when subjects identify "target" stimuli in an oddball paradigm or when they make cognitive discriminations (Polich, 2007). A main effect of vocalization categories and associated perceptual responses was revealed ($F_{3,30}$ = 4.966, P = 0.006). The averaged P300 response amplitudes from largest to smallest are as follows: human-mimicked vocalizations perceived as human (HH; 3.45μV, SD = 1.46), animal vocalizations perceived as human (AH; 3.28μV, SD = 1.39), human-mimicked vocalizations perceived as animal (HA; 2.89μV, SD = 1.16), and animal vocalizations perceived as animal (AA; 2.86μV, SD = 1.14). Pairwise comparisons revealed that the HH condition produced significantly greater P300 amplitudes than both conditions in which the stimuli were perceived as animal vocalizations (HH vs AA, P = 0.045, HH vs HA, P = 0.039). The HH condition did not significantly differ from the AH condition (P = 1.0). The two conditions in which subjects perceived animal vocalizations (AA vs HA) also did not differ in their respective amplitudes (P = 1.0). Thus, the perceived category

of incoming stimulus events seems to be the strongest determining factor of P300 amplitudes.

DISCUSSION

This chapter describes an EEG-based experiment to elucidate temporal dynamics of conspecific vocalization processing in human auditory cortices. Similar to Chapter 5, a novel category of vocalizations, human-mimicked animal vocalizations, was used to critically control for cortical activity to over-learned conspecific vocalizations (speech, other stereotypical non-verbal vocalizations). Doing so allowed us to reveal human voice preferring AEP responses in the time frame of the N1 component, which is generated near primary auditory cortices.

Previous studies investigating similar phenomena included speech and non-verbal vocalizations in their "voice" categories (Charest et al., 2009). Concerns about attentional modulation confounds and potential preferential responses to speech signals prompted De Lucia et al. to re-analyze previously reported data within the context of conspecific vocalization processing (De Lucia et al., 2010). The previous study by this group had investigated electrophysiological processing differences between sounds produced by living and "man-made" sources (Murray et al., 2006); this study included a discrimination task not related to vocalization sounds. Differences between EEG-based signals in response to animal and human vocalizations were found in the timeframe of 169-219ms after stimulus presentation. Note, however, that this finding is very similar to previous assertions that vocalization segregation occurs at approximately 164ms (Charest et al., 2009).

Nonetheless, De Lucia et al. proposed a four-tiered temporal cortical processing hierarchy for human audition: (1) "general" sound processing (low-level spectrotemporal processing) occurs before approximately 70ms, (2) the differentiation between man-made and living sound sources occurs in a window near 70-119ms, (3) human versus animal vocalization discrimination occurs between approximately 169-219ms, and (4) music versus non-music discrimination occurs around 291-357ms. The latter two tiers support the original findings of Charest et al. and Levy et al., respectively (Levy et al., 2001, 2003, Charest et al., 2009). The findings of the current study suggest that the brain's ability to discriminate between human-produced and animal-produced vocalizations

occurs much earlier than their proposed third tier. Specifically, our results suggest that these processing differences exist in the N1 component, perhaps as early as 75-135ms. Additionally, the P300 results seen here likely reflect perceptual-driven processes that may affect the semantic organization of vocalization categories (Talkington et al., 2012).

These findings support our assumption that the human brain is optimized – intrinsically, through development, or a combination of both – for processing human vocalizations. If true, early cortical networks near PACs (or sooner) should show some preferential sensitivity to the human vocal tract. Future work will be able to examine the nature of this categorical boundary between these vocalization classes in neurophysiological responses such as auditory evoked potentials. Understanding these mechanisms will aid in the design of new hearing prosthetics that perform biologically inspired auditory signal processing. Additionally, understanding these intermediate processing stages will assist in the development of new neurologically-based rehabilitative therapies targeting communication disorders.

FIGURES

FIGURE 7-1

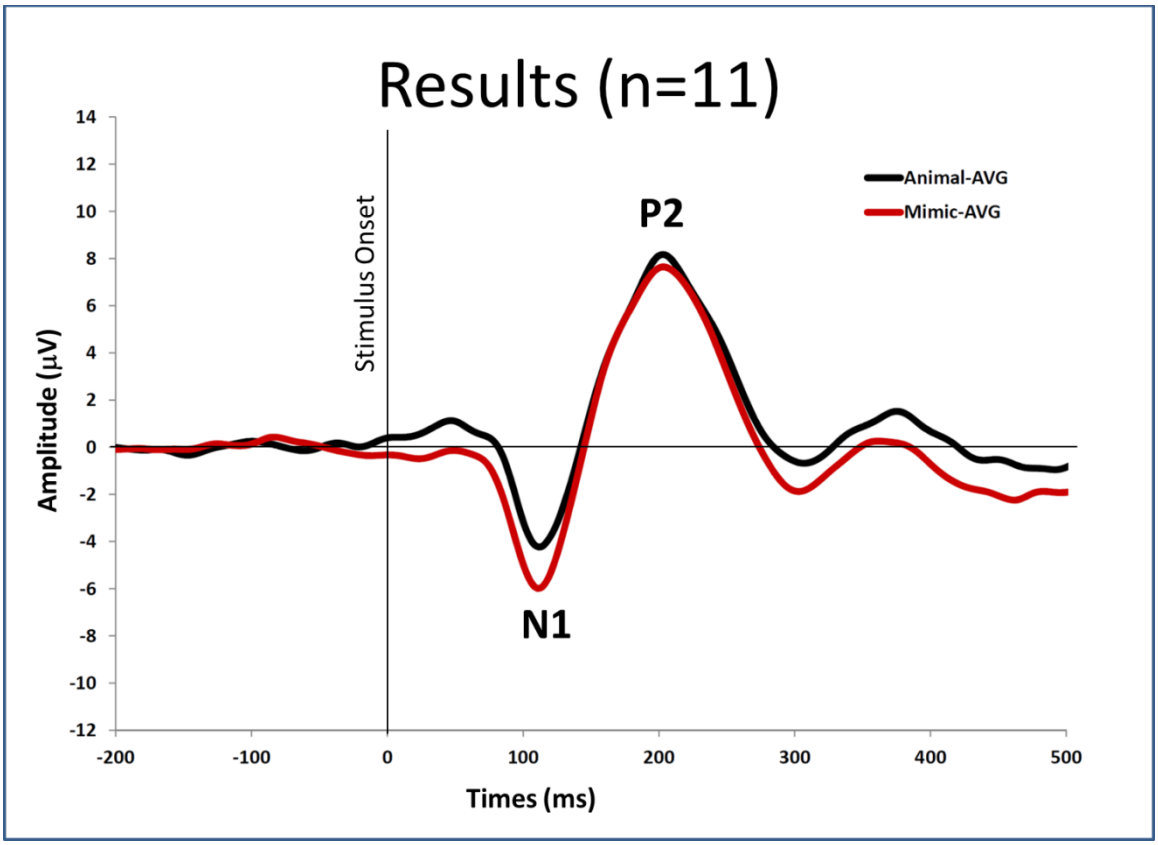


FIGURE 7-1. Averaged AEP waveforms for electrodes F3, Fz, and F4 in response to human-mimicked animal vocalizations and their corresponding animal vocalizations. Significant N1 amplitude differences were seen between the two categories; the respective P2 amplitudes are not significantly different. See text for amplitude values and results of statistical analyses.

FIGURE 7-2

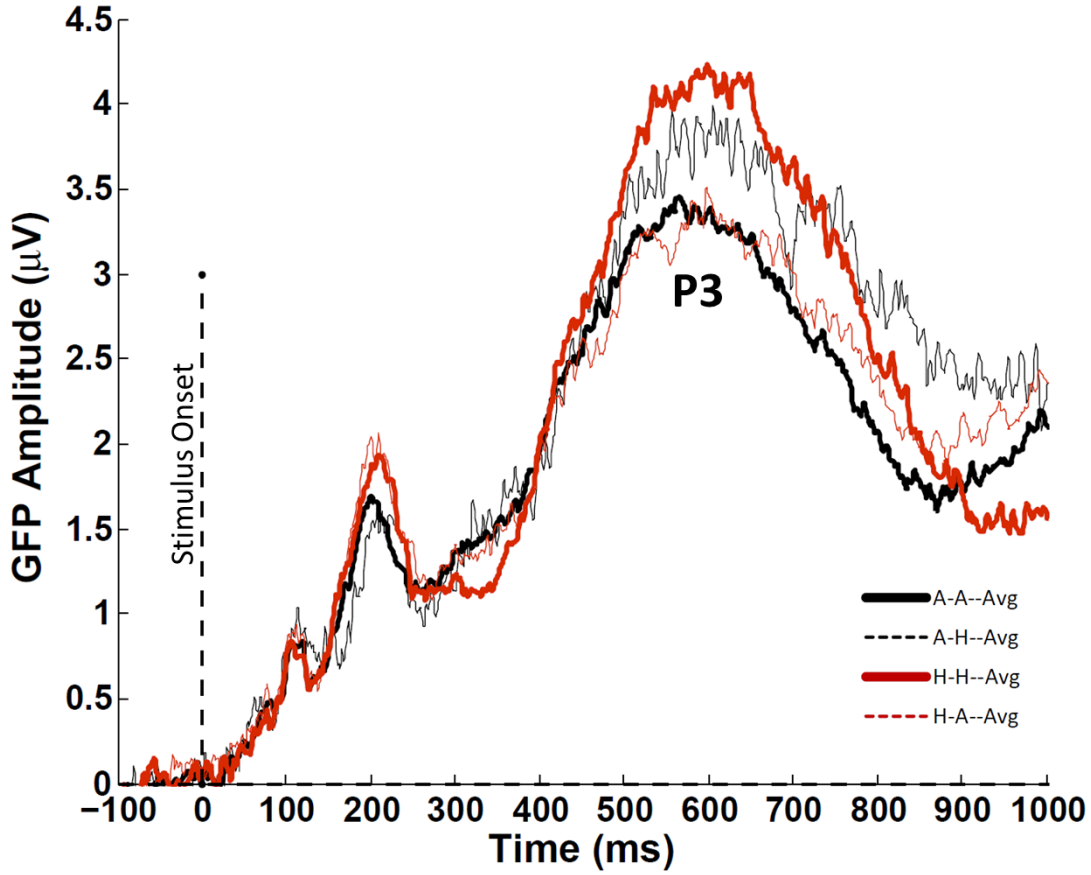


FIGURE 7-2. Averaged GFP waveforms in response to animal vocalizations and human-mimicked versions, separated by perceptual responses. The P300 (or P3) response potential is labeled and defined in the timeframe of 400-800ms. Animal vocalization perceived as animal-produced (A-A); animal perceived as human (A-H); human perceived as human (H-H); human perceived as animal (H-A). See text for amplitude values and results of statistical analyses.

CHAPTER 8:

General discussion and suggestions for future studies

SUMMARY

The goals of this dissertation were to identify and describe the cortical networks and mechanisms that subserve conspecific vocalization processing in humans. These networks form the foundational elements of the human language faculties – skills representing thousands of years of refinement and arguably the richest medium for information exchange. Compromised neuronal representations of vocalization sounds are the hallmark of numerous neurological diseases and conditions. Collectively, the findings presented herein suggest that human auditory cortical networks are most optimized to process the vocal patterns of their conspecifics. These preferences likely have an evolutionary origin, as other species show similar species-specific hearing and communicative phenomena. Additionally, these networks may be partially instantiated from birth and subsequently undergo further developmental refinement through lifelong experiences. More specifically, the experiments described in the previous chapters conclude that (1) early auditory networks show sensitivity to basic acoustic attributes that are characteristic of vocalizations (i.e. strong harmonic content) and (2) that a cortical preference exists for conspecific vocalizations versus non-species vocalizations in predominantly left-lateralized cortical regions near primary auditory cortices (PACs), even when those utterances are outside the typical repertoire of human-produced vocalizations. Ultimately, the present findings add critical information to the wider body of knowledge concerning the structure and function of the cortical networks that allow humans to meet the dynamic demands of their respective auditory environments.

DISCUSSION

HNR sensitivity in human auditory cortices

Vocalizations represent some of the most complex sound events in the natural world. Very subtle acoustic changes have the power to impart radically different meanings to a given utterance. Generally, due to the structure of most vocal apparatuses (e.g. vocal cords), harmonic content represents one of the more predominant acoustic dimensions of vocalizations (Riede et al., 2001, Fitch et al., 2002, Miller and Engstrom, 2010). Additionally, harmonic sounds tend to form statistically distinguishable auditory objects when compared to typical acoustic backgrounds in the environment that are more chaotic or noisy. The desire and need to be heard may have been a driving evolutionary force; more salient sounds that were easily separable from other, less ethologically relevant sounds, would have proven more useful as a communication medium. Harmonics, combinations of mathematically related frequencies, form the bases of numerous models to explain the nature of auditory circuits in animals (Lewicki and Konishi, 1995, Rauschecker et al., 1995, Medvedev et al., 2002, Medvedev and Kanwal, 2004, Kumar et al., 2007). Additionally, specific harmonic arrangements can impart categorical differences between different classes of stimuli (Le Prell et al., 2002).

Chapters 2 and 3 of this dissertation investigated the cortical representation of acoustic harmonic content of artificial IRN stimuli and animal vocalizations (Chapter 2 only). Both chapters identify and describe parametric sensitivities to the HNR of these stimuli and support the representation of this acoustic attribute in early auditory cortical networks. These findings support theories of auditory cortical organization that posit the existence of spectrotemporal templates that are most sensitive to specific combinations of acoustic attributes (Terhardt, 1974, Griffiths and Warren, 2002, Kumar et al., 2007). Spectrotemporal templates for vocalizations are likely built upon combinations of specific harmonic acoustic components that reflect conspecific vocalizations and other biologically relevant vocalizations (Suga et al., 1983, Lewicki and Konishi, 1995, Medvedev et al., 2002). Sensitivity to harmonics likely only represents one acoustic

dimension of vocalizations to which auditory cortices show sensitivity. Extremely complicated multi-dimensional spectrotemporal templates that integrate across the entire acoustic landscape probably form the networks necessary for adequate vocalization comprehension (Rauschecker et al., 1995, Griffiths and Warren, 2002). Our aforementioned studies incorporating IRN stimuli were restricted to very simple integer-harmonic spectral patterns. Future studies that utilize HNR as a quantitative measure of spectral template "matching" will benefit from the creation of artificial stimuli that more accurately reflect the acoustic characteristics of natural vocalizations. Extending the rationale of this dissertation's IRN stimuli, more advanced stimuli (conceptually and acoustically) will permit increasingly detailed investigations of the hierarchical structures and relationships in the auditory networks that support high-level auditory skills such as source identification, emotional prosodic cue processing, and eventually language comprehension.

Conspecific vocalization processing in human auditory cortices

Overlapping networks that show sensitivity to various acoustic attributes or combinations thereof eventually materialize as cortical regions that show preferential activity to a specific category of sound. Similar to face sensitive regions (e.g. FFA) in the occipito-temporal regions of human cortices (Kanwisher et al., 1997), voice-selective (or preferential) cortical regions have traditionally been localized to the bilateral STS in so called temporal voice areas (TVA) (Belin et al., 2000). However, in Chapter 5, by utilizing a novel set of human vocalizations, human-mimicked versions of animal vocalizations, we identified a left-lateralized region of cortex near primary auditory cortices that showed preferential activity to conspecific vocalizations. Additionally, in Chapter 6 using electrophysiology, we also demonstrate an early preference for human produced vocalization in auditory evoked potential (AEP) components reflecting activity near PACs. Using a non-stereotypical category of human vocalizations minimized cortical activity in regions that are optimized to process regularly encountered human vocalizations such as language, yawning, coughing, crying, screaming, etc. The functional contrast between human mimic stimuli and animal vocalizations revealed

cortical regions that were most sensitive to the acoustic characteristics of the human vocal tract.

When compared to activity related to language, whether foreign or native, this conspecific vocalization preferring region seems to form the basis for a left-lateralized temporal hierarchy used for the extraction of locutionary (or semantic) information from perceived human vocalization sounds (Austin, 1975, Scott et al., 2000). Conversely, right hemisphere networks appeared to preferentially process the more emotional prosodic cues within vocalizations (Ethofer et al., 2006a). Findings from developmental neuroscience studies suggest that the right hemisphere is dominant during the refinement of auditory communication skills (Grossmann et al., 2010, Blasi et al., 2011), and may represent an evolutionary link for these faculties (Petkov et al., 2008). Continued experimentation in these two subject groups will likely be a fruitful area of research as they represent the anatomical and functional cortical bases for vocalization processing in the adult human brain. Additionally, the use of closely related, yet distinct stimulus categories (e.g. human-mimicked vs. real world animal vocalizations) will aid in revealing cortical regions that are most critical for the semantic organization of the human auditory system. Understanding the mechanisms of these cortical processing differences will provide the opportunity to more fully understand the compromised function exhibited in numerous auditory communicative disorders. Our results thus far clearly demonstrate the existence of specialized human cortical networks used for processing the vocalizations of other humans in a species-specific manner. Furthermore, these networks are situated near cortical regions that perform more "basic" auditory processing, highlighting the fundamental importance of this skill.

REFERENCES

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

Abrams DA, Nicol T, Zecker S, Kraus N (2009) Abnormal Cortical Processing of the Syllable Rate of Speech in Poor Readers. Journal of Neuroscience 29:7686-7693.

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

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

Alain C, McDonald KL, Ostroff JM, Schneider B (2001) Age-related changes in detecting a mistuned harmonic. J Acoust Soc Am 109:2211-2216.

Altmann CF, Doehrmann O, Kaiser J (2007) Selectivity for animal vocalizations in the human auditory cortex. Cereb Cortex 17:2601-2608.

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

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

Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. Neuroimage 22:401-408.

Austin JL (1975) How to Do Things with Words. Cambridge: Harvard University Press.

Barker D, Plack CJ, Hall DA (2012) Reexamining the evidence for a pitch-sensitive region: a human fMRI study using iterated ripple noise. Cereb Cortex 22:745-753.

Bass AH, Gilland EH, Baker R (2008) Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. Science 321:417-421.

Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M (2006) Electrophysiological markers of voice familiarity. Eur J Neurosci 23:3081-3086.

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

Belin P, Fecteau S, Bedard C (2004) Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129-135.

Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105-2109.

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

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

Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological studies of face perception in humans. Journal of Cognitive Neuroscience 8:551-565.

Bilecen D, Seifritz E, Scheffler K, Henning J, Schulte AC (2002) Amplitopicity of the human auditory cortex: an fMRI study. Neuroimage 17:710-718.

Billings CJ, Bennett KO, Molis MR, Leek MR (2011) Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. Ear Hear 32:53-60.

Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing E (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512-528.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. Journal of Neuroscience 17:353-362.

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

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389-395.

Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc Inst Phon Sci 15:97-110.

Bregman AS (1990) Auditory Scene Analysis: MIT Press.

Carlyon RP (2004) How the brain separates sounds. Trends Cogn Sci 8:465-471.

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

Caskey M, Stephens B, Tucker R, Vohr B (2011) Importance of Parent Talk on the Development of Preterm Infant Vocalizations. Pediatrics 128:910-916.

Charest I, Pernet CR, Rousselet GA, Quinones I, Latinus M, Fillion-Bilodeau S, Chartrand JP, Belin P (2009) Electrophysiological evidence for an early processing of human voices. BMC Neurosci 10:127.

Cheang HS, Pell MD (2008) The sound of sarcasm. Speech communication 50:366-381.

Chen JL, Penhune VB, Zatorre RJ (2008) Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. J Cogn Neurosci 20:226-239.

Cheng Y, Lee SY, Chen HY, Wang PY, Decety J (2012) Voice and emotion processing in the human neonatal brain. J Cogn Neurosci 24:1411-1419.

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

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

Cirillo J (2004) Communication by unvoiced speech: the role of whispering. An Acad Bras Cienc 76:413-423.

Coath M, Balaguer-Ballester E, Denham SL, Denham M (2008) The linearity of emergent spectro-temporal receptive fields in a model of auditory cortex. Biosystems 94:60-67.

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

Coath M, Denham SL (2005) Robust sound classification through the representation of similarity using response fields derived from stimuli during early experience. Biol Cybern 93:22-30.

Cooke M, Ellis DPW (2001) The auditory organization of speech and other sources in listeners and computational models. Speech communication 35:141-177.

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

Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers And Biomedical Research 29:162-173.

Craig AD (2009) How do you feel--now? The anterior insula and human awareness. Nat Rev Neurosci 10:59-70.

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

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

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

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

Darwin CJ (1984) Perceiving vowels in the presence of another sound: constraints on formant perception. J Acoust Soc Am 76:1636-1647.

Davila Ross M, Owren MJ, Zimmermann E (2009) Reconstructing the evolution of laughter in great apes and humans. Curr Biol 19:1106-1111.

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

De Lucia M, Camen C, Clarke S, Murray MM (2009) The role of actions in auditory object discrimination. Neuroimage 48:475-485.

De Lucia M, Clarke S, Murray MM (2010) A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci 30:11210-11221.

DeCasper AJ, Fifer WP (1980) Of human bonding: newborns prefer their mothers' voices. Science 208:1174-1176.

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

DeCasper AJ, Lecanuet JP, Busnel MC, Granierdeferre C, Maugeais R (1994b) Fetal Reactions To Recurrent Maternal Speech. Infant Behav Dev 17:159-164.

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L (2002) Functional neuroimaging of speech perception in infants. Science 298:2013-2015.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9-21.

Douglass JK, Wilkens L, Pantazelou E, Moss F (1993) Noise enhancement of information-transfer in crayfish mechanoreceptors by stochastic resonance. Nature 365:337-340.

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

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

Elhilali M, Shamma SA (2008) A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. J Acoust Soc Am 124:3751-3771.

Ellis DPW (1997) The weft: a representation for periodic sounds. In: ICASSP, vol. 2, pp 1307-1310.

Engel LR, Frum C, Puce A, Walker NA, Lewis JW (2009) Different categories of living and non-living sound-sources activate distinct cortical networks. Neuroimage 47:1778-1791.

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

Ethofer T, Pourtois G, Wildgruber D (2006b) Investigating audiovisual integration of emotional signals in the human brain. Prog Brain Res 156:345-361.

Fecteau S, Armony JL, Joanette Y, Belin P (2004) Is voice processing species-specific in human auditory cortex? - An fMRI study. Neuroimage 23:840-848.

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

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

Fitch WT (2011) Speech perception: a language-trained chimpanzee weighs in. Curr Biol 21:R543-546.

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

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

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

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

Fujioka T, Kakigi R, Gunji A, Takeshima Y (2002) The auditory evoked magnetic fields to very high frequency tones. Neuroscience 112:367-381.

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

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

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

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

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

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. J Neurophysiol 84:1588-1598.

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

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

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

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20-25.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 8:145-146.

Griffiths TD, Buchel C, Frackowiak RS, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. Nat Neurosci 1:422-427.

Griffiths TD, Kumar S, Warren JD, Stewart L, Stephan KE, Friston KJ (2007) Approaches to the cortical analysis of auditory objects. Hear Res 229:46-53.

Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the temporal regularity of sound in the human brainstem. Nat Neurosci 4:633-637.

Griffiths TD, Warren JD (2002) The planum temporale as a computational hub. Trends Neurosci 25:348-353.

Grossmann T, Missana M, Friederici AD, Ghazanfar AA (2012) Neural correlates of perceptual narrowing in cross-species face-voice matching. Dev Sci 15:830-839.

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

Gunji A, Koyama S, Ishii R, Levy D, Okamoto H, Kakigi R, Pantev C (2003) Magnetoencephalographic study of the cortical activity elicited by human voice. Neuroscience Letters 348:13-16.

Gutkin BS, Jost J, Tuckwell HC (2009) Inhibition of rhythmic neural spiking by noise: the occurrence of a minimum in activity with increasing noise. Naturwissenschaften 96:1091-1097.

Hall DA (2005) Representations of spectral coding in the human brain. Int Rev Neurobiol 70:331-369.

Hall DA, Barrett DJ, Akeroyd MA, Summerfield AQ (2005) Cortical representations of temporal structure in sound. J Neurophysiol 94:3181-3191.

Hall DA, Edmondson-Jones AM, Fridriksson J (2006) Periodicity and frequency coding in human auditory cortex. Eur J Neurosci 24:3601-3610.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Human Brain Mapping 7:213-223.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140-149.

Hall DA, Plack CJ (2007) The human 'pitch center' responds differently to iterated noise and Huggins pitch. Neuroreport 18:323-327.

Hall DA, Plack CJ (2009) Pitch processing sites in the human auditory brain. Cereb Cortex 19:576-585.

Hashimoto T, Usui N, Taira M, Nose I, Haji T, Kojima S (2006) The neural mechanism associated with the processing of onomatopoeic sounds. Neuroimage 31:1762-1770.

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

Heffner HE, Heffner RS (1984) Temporal lobe lesions and perception of species-specific vocalizations by macaques. Science 226:75-76.

Hickok G, Okada K, Serences JT (2009) Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol 101:2725-2732.

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

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

Homae F, Watanabe H, Nakano T, Asakawa K, Taga G (2006) The right hemisphere of sleeping infant perceives sentential prosody. Neurosci Res 54:276-280.

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

Jaaskelainen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levanen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci U S A 101:6809-6814.

Jancke L, Shah NJ, Posse S, Grosse-Ryuken M, Muller-Gartner HW (1998) Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia 36:875-883.

Jennings JR, Wood CC (1976) Letter: The epsilon-adjustment procedure for repeated-measures analyses of variance. Psychophysiology 13:277-278.

Jiang W, Tremblay F, Chapman CE (1997) Neuronal encoding of texture changes in the primary and the secondary somatosensory cortical areas of monkeys during passive texture discrimination. J Neurophysiol 77:1656-1662.

Johnson KL, Nicol TG, Kraus N (2005) Brain stem response to speech: a biological marker of auditory processing. Ear Hear 26:424-434.

Joly O, Pallier C, Ramus F, Pressnitzer D, Vanduffel W, Orban GA (2012) Processing of vocalizations in humans and monkeys: a comparative fMRI study. Neuroimage 62:1376-1389.

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

Jones SJ (2006) Cortical processing of quasi-periodic versus random noise sounds. Hear Res 221:65-72.

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

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proceedings Of The National Academy Of Sciences Of The United States Of America 97:11793-11799.

Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. Curr Opin Neurobiol 9:164-170.

Kanwal JS, Fitzpatrick DC, Suga N (1999) Facilitatory and inhibitory frequency tuning of combination-sensitive neurons in the primary auditory cortex of mustached bats. J Neurophysiol 82:2327-2345.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302-4311.

Kastner S, De Weerd P, Ungerleider LG (2000) Texture segregation in the human visual cortex: A functional MRI study. J Neurophysiol 83:2453-2457.

Key AP, Lambert EW, Aschner JL, Maitre NL (2012) Influence of gestational age and postnatal age on speech sound processing in NICU infants. Psychophysiology 49:720-731.

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 (2009) Fetal sensitivity to properties of maternal speech and language. Infant Behav Dev 32:59-71.

Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z (2003) Effects of experience on fetal voice recognition. Psychol Sci 14:220-224.

Knutson B, Burgdorf J, Panksepp J (2002) Ultrasonic vocalizations as indices of affective states in rats. Psychol Bull 128:961-977.

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

Krishnan A, Gandour JT (2009) The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang 110:135-148.

Krishnan A, Gandour JT, Bidelman GM (2010a) Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Res 1313:124-133.

Krishnan A, Gandour JT, Bidelman GM (2010b) The effects of tone language experience on pitch processing in the brainstem. J Neurolinguistics 23:81-95.

Krishnan A, Gandour JT, Bidelman GM, Swaminathan J (2009a) Experience-dependent neural representation of dynamic pitch in the brainstem. Neuroreport 20:408-413.

Krishnan A, Gandour JT, Smalt CJ, Bidelman GM (2010c) Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. Brain And Language 114:193-198.

Krishnan A, Swaminathan J, Gandour JT (2009b) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J Cogn Neurosci 21:1092-1105.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res Cogn Brain Res 25:161-168.

Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res 189:1-12.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lutkenhoner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb Cortex 13:765-772.

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

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

Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD (2007) Hierarchical processing of auditory objects in humans. PLoS Comput Biol 3:e100.

Langers DR, Backes WH, van Dijk P (2007) Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. Neuroimage 34:264-273.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115-142.

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

Lattner S, Maess B, Wang Y, Schauer M, Alter K, Friederici AD (2003) Dissociation of human and computer voices in the brain: evidence for a preattentive gestalt-like perception. Hum Brain Mapp 20:13-21.

Le Prell CG, Hauser MD, Moody DB (2002) Discrete or graded variation within rhesus monkey screams? Psychophysical experiments on classification. Anim Behav 63:47-62.

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30:7604-7612.

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

Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalogr Clin Neurophysiol 48:609-621.

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

Levin JE, Miller JP (1996) Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. Nature 380:165-168.

Levy DA, Granot R, Bentin S (2001) Processing specificity for human voice stimuli: electrophysiological evidence. Neuroreport 12:2653-2657.

Levy DA, Granot R, Bentin S (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40:291-305.

Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. Proc Natl Acad Sci U S A 92:5582-5586.

Lewis JW (2006) Cortical networks related to human use of tools. The Neuroscientist 12:211-231.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148-5158.

Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA (2006) Lefties get it "right" when hearing tool sounds. Journal of Cognitive Neuroscience 18(8):1314-1330.

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

Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA (2009) Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. J Neurosci 29:2283-2296.

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

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

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

Lomber SG, Malhotra S (2008) Double dissociation of 'what' and 'where' processing in auditory cortex. Nat Neurosci 11:609-616.

Luck SJ (2005) An introduction to the event-related potential technique. Cambridge, MA, USA: The MIT Press.

Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage 14:802-816.

Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch HVc. Journal of Neuroscience 12:4309-4326.

Marshall AJ, Wrangham RW, Arcadi AC (1999) Does learning affect the structure of vocalizations in chimpanzees? Anim Behav 58:825-830.

Martin BA, Tremblay KL, Korczak P (2008) Speech evoked potentials: from the laboratory to the clinic. Ear Hear 29:285-313.

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

McDonnell MD, Abbott D (2009) What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. PLoS Comput Biol 5.

McDonnell MD, Ward LM (2011) The benefits of noise in neural systems: bridging theory and experiment. Nat Rev Neurosci 12:415-426.

Medvedev AV, Chiao F, Kanwal JS (2002) Modeling complex tone perception: grouping harmonics with combination-sensitive neurons. Biol Cybern 86:497-505.

Medvedev AV, Kanwal JS (2004) Local field potentials and spiking activity in the primary auditory cortex in response to social calls. J Neurophysiol 92:52-65.

Miller JR, Engstrom MD (2010) Stereotypic vocalizations in harvest mice (Reithrodontomys): Harmonic structure contains prominent and distinctive audible, ultrasonic, and non-linear elements. J Acoust Soc Am 128:1501-1510.

Miller LM, Schreiner CE (2000) Stimulus-based state control in the thalamocortical system. Journal of Neuroscience 20:7011-7016.

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

Misawa H, Suga N (2001) Multiple combination-sensitive neurons in the auditory cortex of the mustached bat. Hearing Research 151:15-29.

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

Morest DK (1965) The Laminar Structure of the Medial Geniculate Body of the Cat. J Anat 99:143-160.

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

Morosan P, Schleicher A, Amunts K, Zilles K (2005) Multimodal architectonic mapping of human superior temporal gyrus. Anat Embryol (Berl) 210:401-406.

Moss F, Ward LM, Sannita WG (2004) Stochastic resonance and sensory information processing: a tutorial and review of application. Clin Neurophysiol 115:267-281.

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20:249-264.

Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006) Rapid brain discrimination of sounds of objects. J Neurosci 26:1293-1302.

Näätänen R, Picton T (1987) The N1 Wave Of The Human Electric And Magnetic Response To Sound - A Review And An Analysis Of The Component Structure. Psychophysiology 24:375-425.

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) "Primitive intelligence" in the auditory cortex. Trends Neurosci 24:283-288.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14:474-480.

Noback CR, Strominger NL, Demarest RJ, Ruggiero DA (2005) The Human Nervous System: Structure and Function. Totowa, New Jersey: Humana Press.

Nunez PL, Srinivasan R (2006) Electric fields of the brain: the neurophysics of EEG. New York: Oxford University Press.

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

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

Obleser J, Zimmermann J, Van Meter J, Rauschecker JP (2007) Multiple stages of auditory speech perception reflected in event-related FMRI. Cereb Cortex 17:2251-2257.

Oliver DL, Morest DK (1984) The central nucleus of the inferior colliculus in the cat. J Comp Neurol 222:237-264.

Overath T, Kumar S, Stewart L, von Kriegstein K, Cusack R, Rees A, Griffiths TD (2010) Cortical mechanisms for the segregation and representation of acoustic textures. J Neurosci 30:2070-2076.

Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. J Neurosci 28:13268-13273.

Parker GJM, Luzzi S, Alexander DC, Wheeler-Kingshott CAM, Clecarelli O, Ralph MAL (2005) Lateralization of ventral and dorsal auditory-language pathways in the human brain. Neuroimage 24:656-666.

Pascual-Leone A, Hamilton R (2001) The metamodal organization of the brain. Prog Brain Res 134:427-445.

Patterson R, Handel S, Yost WA, Datta AJ (1996) The relative strength of the tone and noise components in iterated rippled noise. J Acoust Soc Am 100:3286-3294.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767-776.

Paydarfar D, Forger DB, Clay JR (2006) Noisy inputs and the induction of on-off switching behavior in a neuronal pacemaker. J Neurophysiol 96:3338-3348.

Pell MD (2006) Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang 96:221-234.

Pena M, Maki A, Kovacic D, Dehaene-Lambertz G, Koizumi H, Bouquet F, Mehler J (2003) Sounds and silence: an optical topography study of language recognition at birth. Proc Natl Acad Sci U S A 100:11702-11705.

Pena M, Werker JF, Dehaene-Lambertz G (2012) Earlier speech exposure does not accelerate speech acquisition. J Neurosci 32:11159-11163.

Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810-6815.

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

Pettigrew CM, Murdoch BE, Ponton CW, Kei J, Chenery HJ, Alku P (2004) Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. Journal of the American Academy of Audiology 15:469-485.

Picton TW (2011) Human auditory evoked potentials. San Diego: Plural Pub.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech communication 41:245-255.

Polich J (2007) Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 118:2128-2148.

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

Preuss TM, Goldmanrakic PS (1991) Myeloarchitecture and cytorarchitecture of the granular frontal-cortex and surrounding regions in the strepsirhine primate *galao* and the anthropoid primate *macaca*. J Comp Neurol 310:429-474.

Price C, Thierry G, Griffiths T (2005) Speech-specific auditory processing: where is it? Trends Cogn Sci 9:271-276.

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

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

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718-724.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97:11800-11806.

Rauschecker JP, Tian B, Hauser M (1995) Processing of Complex Sounds in the Macaque Nonprimary Auditory Cortex. Science 268:111-114.

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

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

Riede T, Mitchell BR, Tokuda I, Owren MJ (2005) Characterizing noise in nonhuman vocalizations: Acoustic analysis and human perception of barks by coyotes and dogs. J Acoust Soc Am 118:514-522.

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

Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Brain Res Cogn Brain Res 3:131-141.

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

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

Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734-747.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367-373.

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

Russell DF, Wilkens LA, Moss F (1999) Use of behavioural stochastic resonance by paddle fish for feeding. Nature 402:291-294.

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

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

Schonwiesner M, von Cramon DY, Rubsamen R (2002) Is it tonotopy after all? Neuroimage 17:1144-1161.

Schweinberger SR (2001) Human brain potential correlates of voice priming and voice recognition. Neuropsychologia 39:921-936.

Scott S, Wise R (2003) PET and fMRI studies of the neural basis of speech perception. Speech communication 41:23-34.

Scott SK (2005) Auditory processing--speech, space and auditory objects. Curr Opin Neurobiol 15:197-201.

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

Seither-Preisler A, Krumbholz K, Patterson R, Seither S, Lutkenhoner B (2004) Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. Eur J Neurosci 19:3073-3080.

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

Shama K, Krishna A, Cholayya NU (2007) Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. EURASIP Journal on Advances in Signal Processing 1-9.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303-304.

Shofner WP (1999) Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: analysis of neural autocorrelograms. J Neurophysiol 81:2662-2674.

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

Soeta Y, Nakagawa S, Tonoike M (2005) Auditory evoked magnetic fields in relation to iterated rippled noise. Hear Res 205:256-261.

Soeta Y, Yanai K, Nakagawa S, Kotani K, Horii K (2007) Loudness in relation to iterated rippled noise. J Sound Vibrat 304:415-419.

Srebro R, Malladi P (1999) Stochastic resonance of the visually evoked potential. Phys Rev E 59:2566-2570.

Steinmann I, Gutschalk A (2012) Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. J Neurophysiol 107:3458-3467.

Stillman RD, Crow G, Moushegian G (1978) Components of the frequency following potential in man. Electroencephalogr Clin Neurophysiol 44:438-446.

Suga N, Oneill WE, Kujirai K, Manabe T (1983) Specificity of combination-sensitive neurons for processing of complex bisonar signals in auditory-cortex of the mustached bat. J Neurophysiol 49:1573-1626.

Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci 26:11138-11147.

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

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

Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J Neurophysiol 91:1282-1296.

Talkington WJ, Rapuano KM, Hitt LA, Frum CA, Lewis JW (2012) Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. J Neurosci 32:8084-8093.

Talkington WJ, Taglialatela JP, Lewis JW (2013) Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. Hearing Research *Submitted*.

Teki S, Chait M, Kumar S, von Kriegstein K, Griffiths TD (2011) Brain bases for auditory stimulus-driven figure-ground segregation. J Neurosci 31:164-171.

Telkemeyer S, Rossi S, Koch SP, Nierhaus T, Steinbrink J, Poeppel D, Obrig H, Wartenburger I (2009) Sensitivity of newborn auditory cortex to the temporal structure of sounds. J Neurosci 29:14726-14733.

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

ten Donkelaar HJ (2011) Clinical Neuroanatomy: Brain Circuitry and Its Disorders. Berlin Heidelberg: Springer-Verlag.

Terhardt E (1974) Pitch, consonance, and harmony. J Acoust Soc Am 55:1061-1069.

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

Tremblay K, Kraus N, McGee T, Ponton C, Otis B (2001) Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. Ear Hear 22:79-90.

Tuckwell HC, Jost J (2012) Analysis of inverse stochastic resonance and the long-term firing of Hodgkin-Huxley neurons with Gaussian white noise. Physica A 391:5311-5325.

Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD (2006) Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage 31:1284-1296.

Van Essen DC (2003) Organization of visual areas in macaque and human cerebral cortex. In: The Visual Neurosciences(Chalupa, L. and Werner, J. S., eds), pp 507-521: MIT Press.

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

Vaughan HG, Ritter W (1970) Sources of auditory evoked responses recorded from human scalp. Electroencephalogr Clin Neurophysiol 28:360-&.

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

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48-55.

Vuilleumier P (2005) How brains beware: neural mechanisms of emotional attention. Trends Cogn Sci 9:585-594.

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

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

Warren JE, Wise RJ, Warren JD (2005) Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci 28:636-643.

Wartenburger I, Steinbrink J, Telkemeyer S, Friedrich M, Friederici AD, Obrig H (2007) The processing of prosody: Evidence of interhemispheric specialization at the age of four. Neuroimage 34:416-425.

Werker JF, Tees RC (1999) Influences on infant speech processing: toward a new synthesis. Annu Rev Psychol 50:509-535.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. Journal of Cognitive Neuroscience 13:1-7.

Wiesenfeld K, Moss F (1995) Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. Nature 373:33-36.

Wilden I, Herzel H, Peters G, Tembrock G (1998) Subharmonics, biphonation, and deterministic chaos in mammal vocalization. The international Journal of Animal Sound and its Recording 9:171-196.

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

Winer JA, Diehl JJ, Larue DT (2001) Projections of auditory cortex to the medial geniculate body of the cat. J Comp Neurol 430:27-55.

Winer JA, Schreiner CE (eds.) (2011) The Auditory Cortex. New York: Springer.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 10:420-422.

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

Yetkin FZ, Roland PS, Christensen WF, Purdy PD (2004) Silent functional magnetic resonance imaging (FMRI) of tonotopicity and stimulus intensity coding in human primary auditory cortex. Laryngoscope 114:512-518.

Yost WA (1996a) Pitch of iterated rippled noise. J Acoust Soc Am 100:511-518.

Yost WA (1996b) Pitch strength of iterated rippled noise. J Acoust Soc Am 100:3329-3335.

Yost WA (1997) Pitch strength of iterated rippled noise when the pitch is ambiguous. J Acoust Soc Am 101:1644-1648.

Zaske R, Schweinberger SR, Kaufmann JM, Kawahara H (2009) In the ear of the beholder: neural correlates of adaptation to voice gender. Eur J Neurosci 30:527-534.

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

Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. J Neurosci 24:3637-3642.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization Of Phonetic And Pitch Discrimination In Speech Processing. Science 256:846-849.

Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. Philos Trans R Soc Lond B Biol Sci 363:1087-1104.

CURRICULUM VITAE

William J. Talkington

EMAIL: btalkington@hsc.wvu.edu PHONE: 931-216-5472

Education

2007-2013 **West Virginia University** **Morgantown, WV**

Ph.D. Candidate (Projected defense date: March 2013)

Dissertation: Early and Late Stage Mechanisms for Vocalization Processing in the Human Auditory System

2003-2007 **Austin Peay State University** **Clarksville, TN**

Bachelor of Science in Physics,

Minor in Mathematics

Professional Skills

Project Management: hypothesis construction, experimental design, regular oversight of project progress to meet goals and deadlines, manage critical lab functions (computational and scientific equipment resource acquisition, maintenance, and upgrades, IRB and HIPPA protocol design and conformance, human subject personal health information recording and management), rapid engagement of new analytical and productivity software tools

Leadership and Instruction: primary investigator on several simultaneous HIPPA-approved projects involving human subjects, training of junior lab members and interns on standard protocols, classroom instruction of theoretical and practical scientific topics

Writing Proficiency: Authorship of numerous peer-reviewed scientific journal publications and conference abstracts, regularly serve as reviewer in peer-review process

Scientific Presentation: Regular oral presentations of research findings to peers, faculty collaborators, senior academic faculty and leadership

---

Technical Skills

Functional Magnetic Resonance Imaging (fMRI): experience with both GE and Siemens MRI systems, functional and structural pulse sequence design and optimization, third-party equipment interface, statistical analysis of volumetric and surface-based functional data, computer cluster processing of cortical surface reconstructions, psychophysical task design

Electrophysiology: experience with NeuroScan SynAmps electroencephalography (EEG) recording equipment, third-party equipment interface and fabrication of electronic componentry, psychophysical task design

Software and Data Analysis: experienced user of fMRI analysis software including Analysis of Functional Neuroimages (AFNI), Surface Mapping with AFNI (SUMA), FreeSurfer cortical surface reconstruction package, remote secure-shell (SSH) interfacing with High Performance Computing (HPC) clusters, EEG analysis software EEGLAB, MatLab (Matrix Laboratory) matrix-based programming, scripting, and data-visualization environment, Praat acoustic analysis and scripting software, PsychoPy (Python-based psychophysics programming and scripting environment), data analysis in JMP (SAS Institute Inc.), Presentation stimuli environment, proficient user of Microsoft Office Suite products including Word, Excel, PowerPoint, EndNote bibliographic software

---

Professional Experience

Ph.D. Candidate 2007-2013 (expected graduation March 2013)
West Virginia University, Morgantown, WV
Advisor: James W. Lewis, Ph.D.

- Developed and standardized experimental paradigms and data analysis techniques for fMRI, EEG, and psychophysical experiments
- Refined data analysis techniques and adopted latest procedures
- Supervised and instructed undergraduate lab members and interns
- Co-author on four peer-reviewed manuscripts

- Primary author on two peer-reviewed manuscripts

Awards

- **National Defense Science and Engineering Graduate (NDSEG) Fellowship**
  (2009-2012)

Scientific Publications

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(7):2283-2296.

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* 2011, 32(12):2241-2255.

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* 2011, 23(8):2079-2101.

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.

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(23):8084-8093.

Talkington WJ, Smith BD, Khoo SK, Frum CA, Lewis JW: Late auditory evoked potentials exhibiting sensitivity to harmonic signal content. Submitted to *Euro J Neurosci*. 2013.

Talkington WJ, Taglialatela JP, Lewis JW: Using naturalistic utterances to investigate vocal communication processing and development in human and non-human primates. Invited Review for *Hearing Research*. 2013.