INFORMATION TRANSFER AND TRANSFORMATION IN THE PRIMATE VENTRAL AUDITORY PROCESSING STREAM

by

Elliot Hansen Smith

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Neuroscience

Department of Bioengineering

The University of Utah

August 2013

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Elliot Hansen Smith**

has been approved by the following supervisory committee members:

Bradley Greger	, Chair	**4/18/2013** Date Approved
Alla Borisyuk	, Member	**5/1/2013** Date Approved
Richard Normann	, Member	**4/18/2013** Date Approved
Gary Rose	, Member	**5/1/2013** Date Approved
Yong Wang	, Member	**4/29/2013** Date Approved

and by **Kristen Keefe**, Chair of the Department of **Neuroscience**

and by Donna M. White, Interim Dean of The Graduate School.

ABSTRACT

The primate auditory system is responsible for analyzing complex patterns of pressure differences and then synthesizing this information into a behaviorally relevant representation of the external world. How the auditory cortex accomplishes this complex task is unknown. This thesis examines the neural mechanisms underlying auditory perception in the primate auditory cortex, focusing on the neural representation of communication sounds. This thesis is composed of three studies of auditory cortical processing in the macaque and human. The first examines coding in primary and tertiary auditory cortex as it relates to the possibility for developing a stimulating auditory neural prosthesis. The second study applies an information theoretic approach to understanding information transfer between primary and tertiary auditory cortex. The final study examines visual influences on human tertiary auditory cortical processing during illusory audiovisual speech perception. Together, these studies provide insight into the cortical physiology underlying sound perception and insight into the creation of a stimulating cortical neural prosthesis for the deaf.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

An NIH NIDCD 5 T32 DC008553-02 training grant and University of Utah startup funds supported this work.

CHAPTER 1

INTRODUCTION

1.1 Object identity processing in the primate auditory system

The goal of the primate auditory system is to take complex, time varying signals from the eardrums and construct a behaviorally relevant representation of the external world. In order to construct such a representation, the signal received from the eardrum must first be deconstructed in time, frequency, intensity, and space. Much of this deconstruction occurs, first in the cochea, then in the brain stem as auditory information is transferred to the cortex. This information is then synthesized through the auditory cortical circuit to form auditory perceptions. The classic view of primary auditory cortex is of a 'whiteboard' [3], containing a complete, yet partially deconstructed, representation of an incoming soundwave. More recent electrical recordings from the brains of awake, behaving animals have illuminated neural responses in early auditory cortex as being more than the sum of their spectrotemporal parts. Specifically, auditory cortical activity is susceptible to attentional modulation [4], and modulation by concurrent visual or olfactory stimuli [5, 6]. How the aforementioned processes coalesce in the electrical patterns of populations of neurons in order to form the selective sound perception exhibited in primates remains a fascinating and important question.

Auditory cortical neuroanatomy, connectivity and cytoarchitecture are well understood in the macaque (reviewed in [7]). Macaque auditory cortex is separated into three main regions: the core, belt and parabelt (Figure 1.1). The core field contains three subfields, the primary auditory cortex (AI), the rostral field (R) and the rostrotemporal field (RT). These fields are arranged tonotopically, that is, by their selectivity to stimulus frequency. Reversals in tonotopic gradients have mainly been used to define borders between AI, R, and RT. The belt field contains eight subfields and the Parabelt field consists of broadly defined rostral (PBr) and caudal (PBc) aspects [7].

Upon reaching the cortex, auditory processing is thought to occur in two cortical processing streams, analogous to the dorsal and ventral processing streams in the visual system [8]: spatial elements of a stimulus are processed in a dorsal stream extending along the parietal lobe, and stimulus identity or quality is processed in a ventral stream extending rostrally along the temporal lobe (Figure 1.2). Support for this hierarchical, multipath processing in visual and auditory cortex originated from neuroanatomical studies outlining axonal projection patterns through the cortical circuit [7, 9-11], and from inferences about functional deficits from lesion studies [12, 13].

Neuroanatomical studies of auditory cortex support this organization, as axonal projections in auditory cortex, while highly parallelized, tend to move along axes that are hierarchically organized (i.e., core to belt to parabelt in rostral and caudal directions) [14]. Neuroanatomical studies of the rostral belt and PBr also share reciprocal connections with ventrolateral prefrontal cortical areas (vlPFC) that are highly responsive to species-specific vocalizations [15]. PBr in particular has reciprocal connections with vlPFC as

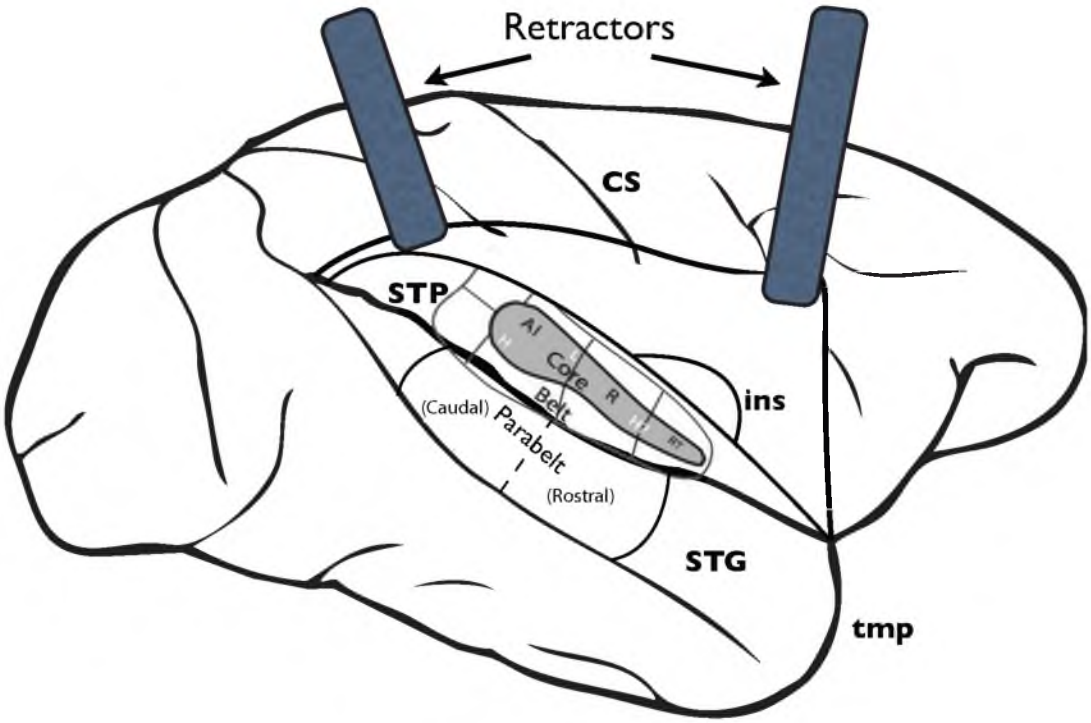


Figure 1.1 The organization of the macaque auditory cortex. Cartoon of a macaque brain with the parietal and frontal lobes retracted along the lateral fissure, exposing the superior temporal plane. Retracting the parietal and frontal lobes at the lateral fissure, exposes the temporal plane and core auditory cortex (shown in gray), which is divided into three tonotopically organized fields. The belt auditory cortex surrounds the core and is organized into eight fields. Parabelt auditory cortex is located on the surface of the superior temporal gyrus and is organized into rostral and caudal aspects. Abbreviations: STP, superior temporal plane; STG, superior temporal gyrus; CS, central sulcus; ins, insula; tmp, temporal pole; AI, primary auditory cortex; R, Rostral Core auditory cortex; RT, Rostrotemporal Core auditory cortex. Modified from Smith [2].

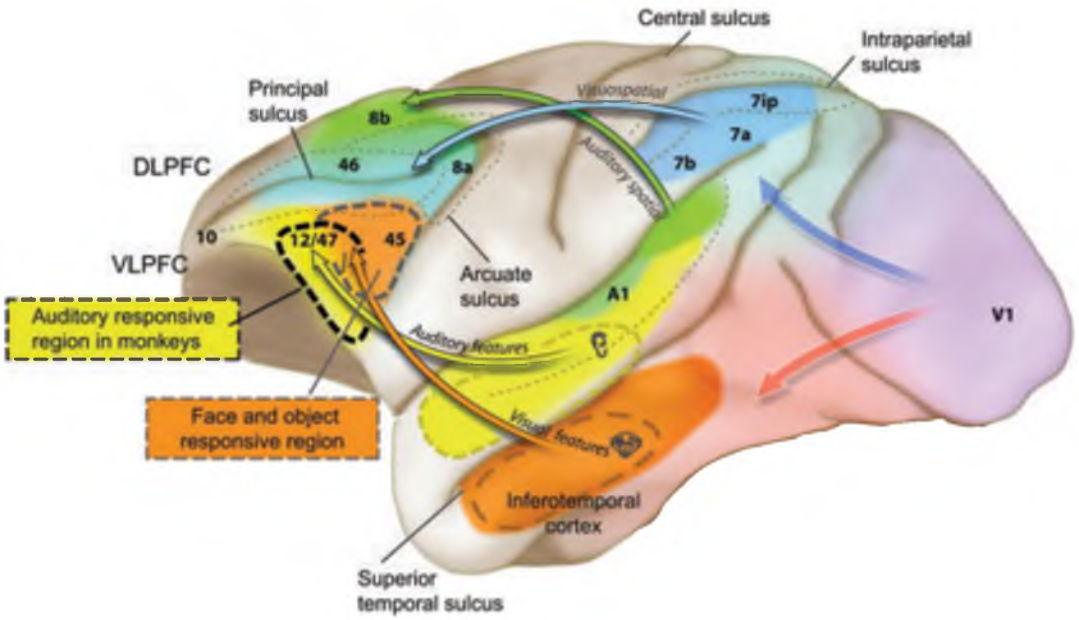


Figure 1.2 Hierarchical processing streams in the macaque cerebral cortex. Cartoon of hierarchical processing of auditory and visual information in the macaque brain. The macaque ventral auditory stream extends from AI to belt to parabelt to ventral prefrontal cortex. Modified from Romanski et al. [1]

well as adjacent rostral belt areas, superior temporal sulcus and areas further rostral on the superior temporal plane (STP) [16, 17].

While there is compelling neurophysiological evidence for a dorsal stream in audition, responsible for sound location perception [18], physiological studies of a ventral stream are limited and have shown mixed results. The strongest physiological evidence for two processing streams in auditory cortex shows increased selectivity for monkey calls at more rostral locations in belt auditory cortex, and increased selectivity for the location of a sound at more caudal locations [19].

Electrophysiological study of the ventral processing stream in primates has been carried out mainly in two areas, the vlPFC and on the STP, in core and belt auditory cortex. Additional physiological evidence for hierarchical processing comes from studies of core and belt areas in the marmoset and macaque that show less temporal synchronization and more rate-based responses at successively rostral core areas and in the belt [20-22]. Minimum latencies also increase from core to belt [23, 24]. Furthermore, a study of three sectors along the macaque STP found that increasingly rostral areas on the STP exhibit greater latencies for auditory stimuli, and show increasingly sparse representations of natural stimuli and monkey vocalizations, a characteristic of hierarchical processing streams [25]. Single neurons have also been found in the macaque vlPFC that respond with a spike rate code, rather than a temporal pattern code, to specific auditory stimuli. At the level of the vlPFC, classes of neurons respond robustly to specific macaque vocalizations and faces [1, 26, 27]. The superior temporal lobe, much of which is considered PBr (Brodmann area 42), is likely an intermediate or parallel processing area between belt auditory cortex and vlPFC.

The architectonic properties of the parabelt region of auditory cortex have been well identified in macaques [17]. The parabelt region receives input from the dorsal and medial divisions of the medial geniculate body of the thalamus [28], and has strong connections to the belt area and minimal connections to the core area of auditory cortex [29]. The rostral parabelt receives mostly input from the rostral belt and the caudal parabelt receives mostly input from the caudal belt [14]. Furthermore, corticocortical connections from the parabelt area and superior temporal gyrus to the orbitofrontal cortex, vlPFC and other prefrontal areas have been well-established [15, 30].

Despite informative anatomical studies, parabelt cortex physiology has been neither extensively nor specifically researched. Caudal parabelt cells have been shown to be responsive to a broad, multimodal range of stimuli, and are broadly selective for stimulus location in the contralateral hemifield [31]. Lesions of superior temporal gyrus in humans that left AI mostly intact yield agnosias for prosody, voice, and melodies [32]. Cebus monkeys that were preoperatively trained on a delayed match to sample auditory memory task acquired memory deficits after outer-bank STG (including parabelt) lesions [33]. In all of these legion studies, little effect on sound detection was reported, suggesting that parabelt and rostral STG are responsible for higher cortical processing.

Much of the auditory neurophysiology in macaques has reinforced the concept of a hierarchically organized ventral cortical processing stream, which is responsible for processing the identity or quality of an auditory stimulus. While single neuron responses in core and belt auditory cortex have been characterized in response to a wide array of auditory stimuli [20, 34, 35], and coding of monkey vocalizations in vlPFC are fairly well understood [36], the physiology of the ventral stream between belt and vlPFC has not

been examined electrophysiologically. Nor has the processing stream been examined in the context of a processing stream by examining causality among cortical areas, by examining object identity or by examining information flow in the auditory cortical circuit.

1.2 Language processing in the human STG

Perhaps the most important auditory object for humans is language. Language has allowed humans to cooperate in many ways, imparting a marked evolutionary advantage. The current understanding of the neural basis for human language comprehension has, until recently, been limited to descriptions of aphasias due to neurological damage. We have surmised that the parts of the brain responsible for understanding language are lateralized, concentrated in the left hemisphere [37, 38]. Using macaques as a model for humans, there is increasing evidence that macaques' neural basis for understanding vocalizations is also lateralized to the left hemisphere. Left hemisphere legions produce vocalization-specific aphasias in macaques [12]. The macaque rostral STG and parabelt cortices have also shown the greatest BOLD response to macaque vocalizations, when compared with control broadband stimuli [39]. The STG therefore is likely an important area for language processing in the primate, and while macaque vocalizations may serve as a gross model for understanding the neural representation of language, it is possible to study the electrophysiological representation of language in humans.

The potential to record electrical potentials from the surface of the human brain exists in epilepsy patients who are undergoing surgery to remove seizure foci. For these patients with pharamacoresistant epilepsies, removing the focus of the seizure can

dramatically reduce seizure frequency. In order for the neurosurgeon to know which part of the brain to remove, the patient must undergo extensive preoperative monitoring, first with EEG and then with intracranial, subdural electrodes placed near the seizure foci. These electrodes are generally large, low impedance electrocorticography (ECoG) electrodes. Due to the frequency of medial temporal seizures, these electrodes are often placed over the superior temporal lobe. Researchers are taking advantage of this rare possibility to record potentials from the surface of the human brain by recording from the clinical monitoring electrodes while the patient is performing a task or listening to stimuli. This methodology affords researchers the opportunity to acquire local field potential (LFP) signals from a variety of brain areas. In Chapter 4 we describe one such study in which we record LFP from human parabelt auditory cortex in response to vocal communication sounds.

Early ECoG recordings of human STG characterized increases in high gamma band activity and suppression of alpha band activity in response to pure tone and language stimuli [40]. Language stimuli produced increased gamma band activity on more electrodes than pure-tone stimuli in each of four patients examined. These STG responses have since been shown to be suppressed during speech [41], and causally related to recording sites on the superior temporal plane, near primary auditory cortex [42]. Furthermore, electrical stimulation of human STG evoked responses in vlPFC in the human, indicating that functional connectivity in the human ventral auditory stream is likely similar to that in the macaque [43].

More recent studies have incorporated the idea of a ventral processing stream into experimental design. These studies have focused more on the neural construction of

auditory stimulus identity through phoneme categorization [44], and coding of speech sounds in the STG [45] using recordings from ECoG and higher density mini-ECoG grids. In these patients, LFP responses showed high correlation with patients' subjective categorizations of three similar phonemes, indicating that precise information about neural representation and subjective perception of phoneme category is contained in the neural responses in STG [44]. Reconstructing the identity of full words from high gamma band activity in STG has revealed that full words can be decoded from ECoG and mini-ECoG electrodes along several centimeters of STG. These results provide evidence for neural representations of auditory elements of language identity existing in STG field potentials. Nonlinear reconstruction methods improved classification of full words, especially for temporal modulations in the stimulus. These modulations are important stimulus elements for speech perception [45]. This result is in accord with study of the macaque ventral stream using statistical classifiers, as linear decode performance decreases along the ventral stream [46] [47]. However, vocalizations can be classified accurately from the vlPFC using a nonlinear, probabilistic classifier [36], indicating that higher areas in the ventral auditory stream likely rely on sparse and nonlinear encoding.

1.3 Single-trial and population analyses of auditory system physiology

Much electrophysiological study of the nervous system has relied on examining responses from single neurons averaged over many stimulus presentations. However, in interpreting a single sensory experience, the brain likely uses noisy signals from many neurons that have been tuned by previous sensory experience. In order to better

understand coding in the brain, it is therefore important to study population codes in response to a single stimulus presentation [48].

Two complementary methods to examine population codes of single-trial neural responses are used in this thesis. The first of these methods is statistical classification, which is also known as decoding. This method mimics the description of the brain's sensory interpretation as described above: Statistical classification uses many neural responses to create a model with which to classify the identity of a single subsequent trial. Statistical classification is therefore a rigorous form of statistics in which the distribution is determined by the data at hand.

Both color and object category and identity have been decoded successfully along the visual ventral stream using linear classifiers [49, 50]. Auditory research using linear classifiers, however, has provided counterintuitive results. By far the most common decoding technique in auditory research is the Linear Pattern Discriminator (LPD) [51]. This decode operates on time-binned action potentials for a single trial over many neurons, clustering each trial in a high-dimensional space based on the similarity of the firing rate responses. The LPD algorithm's classification is then determined by calculating the Euclidian distance between the testing data point and the training data cluster.

Another way in which the current decode differs from prior studies is the method of evaluation. Most previous auditory decodes have implemented LPD as a leave-one-out cross-validation, which means a single trial was removed at random from the full data set and used as testing data, leaving remaining trials to be used for training. We found that performance was much higher for the leave-one-out cross-validation method compared to

the alternative approach where all training trials come before testing trials in time. The cross validation performance using LFP is similar to the highest performance seen using action potentials to decode among ten vocalizations [47]. Leave-one-out cross validation may therefore represent an upper bound estimate of decode performance.

The decode in Chapters 2 and 3 is not implemented as a leave one out cross-validation due to *a priori* issues with using data that came before the training data in time. The decode in Chapters 2 and 3 uses only testing data that followed training data in time in order to minimize learning or stimulus familiarity effects on decode performance. Minor changes in cortical network properties or response properties could inflate decode performance, as is seen with leave-one-out cross-validation. In Chapter 4 we implement the decode as a two-fold cross-validation in which half the trials are used as training and half of the trials are used as testing and then these data sets are exchanged.

Several studies have examined population codes in response to macaque vocalizations in auditory cortex. Using LPD, one study [46] was able to decode a set of four vocalizations with approximately 80% accuracy in each of two core areas, and three belt areas. Another study [47], however, examined neurons in belt area AL, and vlPFC, two areas further downstream, and found that neurons in vlPFC had lower decode performance than belt neurons. These results are counterintuitive, as one would expect selectivity for individual vocalizations to increase along a processing stream for stimulus identity or quality. Part of this discrepancy may be due to the decoding algorithm used. Decode performance of vocalizations in vlPFC was dramatically higher when using linear functions of the probabilistic output of a hidden Markov model, as opposed to linear functions of the spectral features of the monkey vocalizations [52]. This result

indicates that vlPFC neurons may represent the output of dynamic feature processing along an auditory ventral stream. What those features are, and how and where they are processed is unknown. This type of performance likely represents sparse and abstract encoding and is indicative of hierarchically higher areas in cortical processing streams.

While the current results agree with previous auditory decodes regarding the importance of temporal precision, overall performance was more difficult to compare. One group [35] reported lower decode performance than the two other studies in macaque [46] [47], using similar methods. It is, however, difficult to compare these results due to the number of classes decoded. The number of classes in the decode clearly changes the level of performance; therefore, evaluating the decode in a combinatorial manner is essential for future comparison. For example, we cannot compare our results with those studies that used ten classes, yet we can compare our four-class results with those from [53], who observed performance that was approximately 20% higher than our four-class results.

The second method to examine population codes of single-trial neural responses used in this thesis is information theoretic analysis. Information theory as used here quantifies the reduction in uncertainty about the identity of a stimulus based on the neural response. Again, this methodology does not rely on gaussian statistics, instead creating statistical distributions from pairwise classification frequencies from the decode.

Information theory was first described in a paper by an applied mathematician named Claude Shannon [54]. In this paper he described the parts of a communication system and the mathematical techniques to quantify the amount of information transferred through such a communication system. In Chapter 2 we apply the main

finding from this paper to the auditory cortex, treating each auditory cortical area as the receiver in a communication system. Mutual information has been used to quantify neural information about an auditory stimulus in primary and belt auditory cortex and vlPFC with similar results [51] [47] [36].

An innovative extension of information theory in this thesis includes examining conditional mutual information among a stimulus and spatially distributed neural responses. Conditional mutual information is similar to mutual information, with the inclusion of a second receiver in the communication system. This analysis has been used in several studies of the brain to analyze causality among multiple brain regions [55-57]. In Chapters 3 and 4 we apply conditional mutual information to quantify the reduction in uncertainty about stimulus identity given that a neural response in another area is known. This technique gives us insight into the interdependence and temporal progression of responses in cortical areas with respect to processing auditory object identity. While conditional mutual information's ability to quantify information transfer in the cortex would be useful for correlating interconnected brain areas, combining it with statistical classification as we have in Chapters 3 and 4 allows us to examine transfer of information about stimulus identity in the processing stream which is supposed to process auditory object identity.

Statistical classification and information theory are powerful quantitative analyses that provide the ability to describe how many neurons encode a single stimulus. Combining statistical classification and conditional mutual information provides a unique methodology to study processing of object identity in the ventral auditory processing stream.

1.4 Summary and motivation

The current understanding of the neural mechanisms of auditory perception is relatively limited. The dominant theory is that the neural mechanisms underlying auditory perception are analogous to those in vision, such that processing is separated into a ventral, audition for perception processing stream and a dorsal, audition for action processing stream. Yet unlike vision, there remain areas of the auditory ventral processing stream that have not been studied physiologically. While core and belt auditory cortices have been extensively studied and auditory responses in the vlPFC have been characterized in the macaque, parabelt auditory cortex has not been studied in the macaque. Electrophysiological responses to language stimuli in the human parabelt auditory cortex have, however, been characterized. Responses in human parabelt auditory cortex encode modulations more than linear spectrotemporal elements of a stimulus and are closely correlated with human language perception.

A major goal of the sensory brain is to take a single sensory experience and compare it with previous experience in order to determine the identity of whatever is being perceived. This is done with many neurons at a time in distributed cortical networks, with multisensory and contextual information included. It is such a picture of auditory cortical processing that has motivated the current experiments and methodologies. Using decodes and information theoretic analyses over space and time allows for understanding how a large group of neurons in a cortical network respond to a single auditory event based on a contextual model. The thesis focuses on the neural identity of communication stimuli in the primate ventral auditory stream. Using measures of information coding and flow in multiple cortical areas, the studies herein address how

information about stimulus identity transfers through the ventral auditory stream and how information about stimulus identity is changed in the process. These studies speak to several aforementioned issues: physiological considerations involved with neural prosthesis use in AI and parabelt auditory cortex, the transfer and transformation of information between AI and parabelt auditory cortex, and visual influences on parabelt auditory cortex during audiovisual language perception.

1.5 References

[1] L. M. Romanski, "Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex," *Cerebral Cortex (New York, NY : 1991),* vol. 17 Suppl 1 pp. i61-9, 2007.

[2] E. H. Smith, "Temporal processing in the auditory core: Transformation or segregation?," *Journal of Neurophysiology,* vol. 106 (6), pp. 2791-2793, 2011.

[3] I. Nelken, "Processing of complex sounds in the auditory system," *Current Opinion in Neurobiology,* vol. 18 (4), pp. 413-417, 2008.

[4] B. H. Scott, B. J. Malone, and M. N. Semple, "Effect of behavioral context on representation of a spatial cue in core auditory cortex of awake macaques," *Journal of Neuroscience,* vol. 27 (24), pp. 6489-6499, 2007.

[5] A. A. Ghazanfar, J. X. Maier, K. L. Hoffman, and N. K. Logothetis, "Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex," *Journal of Neuroscience,* vol. 25 (20), pp. 5004-5012, 2005.

[6] G. H. Otazu, L.-H. Tai, Y. Yang, and A. M. Zador, "Engaging in an auditory task suppresses responses in auditory cortex," *Nature Neuroscience,* vol. 12 (5), pp. 646-654, 2009.

[7] T. A. Hackett, "Information flow in the auditory cortical network," *Hearing Research,* vol. 271 (1-2), pp. 133-146, 2011.

[8] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nature Neuroscience,* vol. 12 (6), pp. 718-724, 2009.

[9] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[10] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex (New York, NY : 1991),* vol. 1 (1), pp. 1-47, 1991.

[11] J. A. Winer and C. C. Lee, "The distributed auditory cortex," *Hearing Research,* vol. 229 (1-2), pp. 3-13, 2007.

[12] H. E. Hefner and R. S. Heffner, "Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by japanese macaques," *Journal of Neurophysiology,* vol. 56 (3), pp. 683-701, 1986.

[13] A. U. Turken and N. F. Dronkers, "The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses," *Frontiers in System Neuroscience,* vol. 5 pp. 1-20, 2011.

[14] J. H. Kaas, T. A. Hackett, and M. J. Tramo, "Auditory processing in primate cerebral cortex," *Current Opinion in Neurobiology,* vol. 9 (2), pp. 164-170, 1999.

[15] L. M. Romanski, J. F. Bates, and P. S. Goldman-Rakic, "Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey," *The Journal of Comparative Neurology,* vol. 403 (2), pp. 141-157, 1999.

[16] L. M. Romanski, B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker, "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex," *Nature Neuroscience,* vol. 2 (12), pp. 1131-1136, 1999.

[17] T. A. Hackett, I. Stepniewska, and J. H. Kaas, "Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys," *The Journal of Comparative Neurology,* vol. 394 (4), pp. 475-495, 1998.

[18] T. M. Woods, S. E. Lopez, J. H. Long, J. E. Rahman, and G. H. Recanzone, "Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey," *Journal of Neurophysiology,* vol. 96 (6), pp. 3323-3337, 2006.

[19] B. Tian, D. Reser, A. Durham, A. Kustov, and J. P. Rauschecker, "Functional specialization in rhesus monkey auditory cortex," *Science (New York, NY),* vol. 292 (5515), pp. 290-293, 2001.

[20] D. Bendor and X. Wang, "Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys," *Journal of Neurophysiology,* vol. 100 (2), pp. 888-906, 2008.

[21] X. Wang, "Neural coding strategies in auditory cortex," *Hearing Research,* vol. 229 (1-2), pp. 81-93, 2007.

[22] B. H. Scott, B. J. Malone, and M. N. Semple, "Transformation of temporal processing across auditory cortex of awake macaques," *Journal of Neurophysiology,* vol. 105 (2), pp. 712-730, 2011.

[23] G. H. Recanzone, D. C. Guard, and M. L. Phan, "Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey," *Journal of Neurophysiology,* vol. 83 (4), pp. 2315-2331, 2000.

[24] X. Wang, T. Lu, D. Bendor, and E. Bartlett, "Neural coding of temporal information in auditory thalamus and cortex," *Neuroscience,* vol. 157 (2), pp. 484-494, 2008.

[25] Y. Kikuchi, B. Horwitz, and M. Mishkin, "Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane," *Journal of Neuroscience,* vol. 30 (39), pp. 13021-13030, 2010.

[26] L. M. Romanski and J. Hwang, "Timing of audiovisual inputs to the prefrontal cortex and multisensory integration," *Neuroscience,* vol. 214 pp. 36-48, 2012.

[27] L. M. Romanski, B. B. Averbeck, and M. Diltz, "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex," *Journal of Neurophysiology,* vol. 93 (2), pp. 734-747, 2005.

[28] J. H. Kaas and T. A. Hackett, "Subdivisions of auditory cortex and levels of processing in primates," *Audiology and Neuro-Otology,* vol. 3 (2-3), pp. 73-85, 1998.

[29] T. A. Hackett, I. Stepniewska, and J. H. Kaas, "Callosal connections of the parabelt auditory cortex in macaque monkeys," *The European Journal of Neuroscience,* vol. 11 (3), pp. 856-866, 1999.

[30] T. A. Hackett, I. Stepniewska, and J. H. Kaas, "Prefrontal connections of the parabelt auditory cortex in macaque monkeys," *Brain Research,* vol. 817 (1-2), pp. 45-58, 1999.

[31] K. Hikosaka, E. Iwai, H. Saito, and K. Tanaka, "Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey," *Journal of Neurophysiology,* vol. 60 (5), pp. 1615-1637, 1988.

[32] N. Furl, F. Hadj-Bouziane, N. Liu, B. B. Averbeck, and L. G. Ungerleider, "Functional dissociations following bilateral lesions of auditory cortex," *Journal of Neuroscience,* vol. 117 (Pt 6) pp. 1283-1301, 1994.

[33] M. Colombo, H. R. Rodman, and C. G. Gross, "The effects of superior temporal cortex lesions on the processing and retention of auditory information in monkeys (cebus apella)," *The Journal of Neuroscience,* vol. 16 (14), pp. 4501-4517, 1996.

[34] G. H. Recanzone, D. C. Guard, M. L. Phan, and T. K. Su, "Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey," *Journal of Neurophysiology,* vol. 83 (5), pp. 2723-2739, 2000.

[35] P. Kusmierek and J. P. Rauschecker, "Functional specialization of medial auditory belt cortex in the alert rhesus monkey," *Journal of Neurophysiology,* vol. 102 (3), pp. 1606-1622, 2009.

[36] B. B. Averbeck and L. M. Romanski, "Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex," *The Journal of Neuroscience,* vol. 26 (43), pp. 11023-11033, 2006.

[37] N. Geschwind, "Disconnexion syndromes in animals and man. Ii.," *Brain,* vol. 88 (3), pp. 585-644, 1965.

[38] N. Geschwind, "Disconnexion syndromes in animals and man. I.," *Brain,* vol. 88 (2), pp. 237-294, 1965.

[39] C. I. Petkov, C. Kayser, T. Steudel, K. Whittingstall, M. Augath, and N. K. Logothetis, "A voice region in the monkey brain," *Nature Neuroscience,* vol. 11 (3), pp. 367-374, 2008.

[40] N. E. Crone, D. Boatman, B. Gordon, and L. Hao, "Induced electrocorticographic gamma activity during auditory perception. Brazier award-winning article, 2001," *Clinical Neurophysiology,* vol. 112 (4), pp. 565-582, 2001.

[41] A. Flinker, E. F. Chang, H. E. Kirsch, N. M. Barbaro, N. E. Crone, and R. T. Knight, "Single-trial speech suppression of auditory cortex activity in humans," *Journal of Neuroscience,* vol. 30 (49), pp. 16643-16650, 2010.

[42] D. Boatman-Reich, P. J. Franaszczuk, A. Korzeniewska, B. Caffo, E. K. Ritzl, S. Colwell, and N. E. Crone, "Quantifying auditory event-related responses in multichannel human intracranial recordings," *Frontiers in Computational Neuroscience,* vol. 4 p. 4, 2010.

[43] P. C. Garell, H. Bakken, J. D. W. Greenlee, I. Volkov, R. A. Reale, H. Oya, H. Kawasaki, M. A. Howard, and J. F. Brugge, "Functional connection between posterior superior temporal gyrus and ventrolateral prefrontal cortex in human," *Cerebral Cortex,* 2012.

[44] E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight, "Categorical speech representation in human superior temporal gyrus," *Nature Neuroscience,* vol. 13 (11), pp. 1428-1432, 2010.

[45] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biology,* vol. 10 (1), p. e1001251, 2012.

[46] G. H. Recanzone, "Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey," *Journal of Neuroscience,* vol. 28 (49), pp. 13184-13193, 2008.

[47] B. E. Russ, A. L. Ackelson, A. E. Baker, and Y. E. Cohen, "Coding of auditory-stimulus identity in the auditory non-spatial processing stream," *Journal of Neurophysiology,* vol. 99 (1), pp. 87-95, 2008.

[48] R. Quian Quiroga and S. Panzeri, "Extracting information from neuronal populations: Information theory and decoding approaches," *Nature Reviews Neuroscience,* vol. 10 (3), pp. 173-185, 2009.

[49] C. P. Hung, "Fast readout of object identity from macaque inferior temporal cortex," *Science,* vol. 310 (5749), pp. 863-866, 2005.

[50] G. J. Brouwer and D. J. Heeger, "Decoding and reconstructing color from responses in human visual cortex," *Journal of Neuroscience,* vol. 29 (44), pp. 13992-14003, 2009.

[51] J. W. H. Schnupp, T. M. Hall, R. F. Kokelaar, and B. Ahmed, "Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex," *Journal of Neuroscience,* vol. 26 (18), pp. 4785-95, 2006.

[52] B. B. Averbeck and L. M. Romanski, "Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex," *Journal of Neuroscience,* vol. 26 (43), pp. 11023-33, 2006.

[53] G. H. Recanzone, D. C. Guard, and M. L. Phan, "Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey," *Journal of Neurophysiology,* vol. 83 (4), pp. 2315-31, 2000.

[54] C. Shannon, *Shannon: A Mathematical Theory of Communications,* Bell Syst. Tech. J, 1948.

[55] X. Li and G. Ouyang, "Estimating coupling direction between neuronal populations with permutation conditional mutual information," *NeuroImage,* vol. 52 (2), pp. 497-507, 2010.

[56] R. Salvador, M. Anguera, J. J. Gomar, E. T. Bullmore, and E. Pomarol-Clotet, "Conditional mutual information maps as descriptors of net connectivity levels in the brain," *Frontiers in Neuroinformatics,* vol. 4 p. 115, 2010.

[57] M. Vejmelka and M. Palus, "Inferring the directionality of coupling with conditional mutual information," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics,* vol. 77 (2 Pt 2), p. 026214, 2008.

CHAPTER 2

DECODING STIMULUS IDENTITY FROM MULTIUNIT ACTIVITY AND LOCAL FIELD POTENTIALS ALONG THE VENTRAL AUDITORY STREAM IN THE AWAKE PRIMATE: IMPLICATIONS FOR NEURAL PROSTHESES[1]

2.1 Introduction

The cochlear implant is the most widely used neural prosthesis. This device artificially reproduces hearing in a deaf person by stimulating the cochlea with pulses of electric current. Some deaf patients, however, do not have an auditory nerve that extends to the cochlea because of recession of the auditory nerve over time, cochlear ossification, or cranial nerve tumors. These patients require an alternative to the cochlear implant. Two such devices, which interface with subcortical nuclei, are currently being tested [1, 2]. Another possible avenue for eliciting auditory perception from stimulation of the nervous system could be through the auditory cortex.

Microstimulation of primary auditory cortext (AI) has been shown to elicit the correct behavioral responses in rats and cats trained to detect and discriminate pure tones [3, 4]. Stimulation of human auditory cortex has most commonly suppressed sound perception [5]. However, in the absence of sound stimuli, stimulation of Heschl's gyrus

[1] Reprinted with permission from the Journal of Neural Engineering and the Institue of Physics.

has elicited sound percepts [6]. Similar stimulation of human superior temporal gyrus (STG) with large, low-impedance electrodes has produced a variety of sensory percepts, most of them complex and holistic [7-9]. For example, one patient, when stimulated on the medial temporal lobe, reported hearing music; another, when stimulated on the STG, heard "a mother calling her little boy" [8]. Other patients have reported hearing "buzz," "hum," "knocking," "crickets," and "wavering" when stimulated on the STG [7]. Whether stimulation on smaller, penetrating electrodes could elicit more consistent perceptions remains to be tested. Yet, understanding how auditory information is encoded in these early auditory cortical areas will provide guidance on neural prosthetic design and use.

Visual cortex is organized into hierarchical processing streams. Spatial elements of visual stimuli are processed in a dorsal stream, while the identity of visual stimuli is processed in a ventral stream [10, 11]. Neuroanatomical and lesion studies suggest a similar separation of spatial location and stimulus identity pathways in the auditory system [12-17]. Neuroanatomical studies of the rostral belt and parabelt (PBr) support hierarchical organization in a ventral stream that extends along the STG, as these regions share reciprocal connections with ventrolateral prefrontal cortical areas (vlPFC) that are highly responsive to species-specific vocalizations [18]. PBr, in particular, has reciprocal connections with vlPFC as well as adjacent belt areas, superior temporal sulcus, and areas further rostral on the superior temporal plane (STP) [19, 20].

Whereas there is compelling neurophysiological evidence for a spatial processing stream in audition [21], physiological studies of a stimulus identity auditory stream have failed to develop a clear picture of how acoustic information is transformed through the

cortical circuit. Much like inferotemporal cortex for vision, increasingly rostral areas on the STP showed increasingly sparse representations of natural stimuli and monkey vocalizations [22]. Much of the evidence for a ventral stream stems from a study showing increased selectivity for monkey calls at more rostral locations in belt auditory cortex and increased selectivity for the location of a sound at more caudal locations [23]. Other studies have shown less selectivity via robust responses to vocalizations across auditory cortical areas, and imaging has shown areas further anterior on the STP to be selective for vocalizations [24]. Most auditory cortical data has been collected from core and belt regions, with sparse sampling along the STP and no recordings yet in PBr. There is need for further study of auditory cortex on the STG, especially in the context of a neural prosthesis, as the ventral auditory stream is likely important for constructing auditory percepts.

Decoding stimulus identity from neural responses is one way to examine stimulus selectivity along the processing stream. This methodology, however, has yielded conflicting results. Auditory research using linear classifiers on action potential (AP) data has shown less selectivity in successive areas along the processing stream. The linear pattern discriminator (LPD) is a commonly used decoding technique in auditory research [25]. Using LPD on Multiunit activity (MUA) data in awake macaques, a set of four vocalizations was decoded with approximately 90% accuracy in each of two core areas and three belt areas [26]; however, further along the ventral stream, neurons in vlPFC had lower decode performance than belt neurons [27]. Decode performance of vocalizations in vlPFC was higher when using linear functions of the probabilistic output of a hidden Markov model, as opposed to linear functions of the spectrotemporal elements of the

monkey vocalizations, as in LPD [28]. Why decode performance decreases along the ventral auditory processing stream remains an open question that has implications for neural prosthesis development. Greater understanding of how neural coding of stimulus identity at successive stages in an auditory identity processing stream will provide guidance on how to implement a cortically based auditory prosthesis.

Incorporating knowledge of neural coding into stimulation can improve prosthetic performance [29]. Therefore, as a first step towards designing a neural prosthesis based upon cortical microstimulation, we decoded neural signals recorded on chronically implanted microelectrode arrays during species-specific vocalizations. The impact of spatial, temporal, and frequency parameters on decoding performance were examined. Here we examine selectivity of MUA and local field potentials (LFP) for stimulus identity in an awake, behaving primate to better understand the possibility for an auditory cortical neural prosthesis. This study used macaque vocalization stimuli to examine discriminability of auditory stimulus identity in two cortical areas in the ventral processing stream. While there was no direct auditory cortex stimulation in this study, we show that decoding stimulus identity from auditory cortex has implications for basic auditory cortical physiology and provides guidance on the implementation of a cortically based auditory neural prosthesis.

2.2 Materials and methods

2.2.1 Research subject

The neural data examined in this study were recorded from 192 electrodes (96 per electrode array) in two cortical areas in one male rhesus macaque over a period of five

months. All experiments were performed according to National Institutes of Health guidelines for animal care and use and with approval from the University of Utah Institutional Animal Care and Use Committee.

2.2.2 Micro-electrode array implantation

Penetrating microelectrode arrays (Blackrock Microsystems, LLC, Salt Lake City, UT) were implanted in AI and rostral PBr (Figure 2.1). The monkey's temporalis muscle was detached at the top of the skull and retracted, and a craniotomy was made to expose the superior temporal and inferior frontal lobes. The parietal lobe was carefully dissected away from the temporal lobe, exposing a small area on the STP. The parietal lobe was retracted from the surface of the STP, and the AI microelectrode array was inserted by hand at about 5 mm rostral to the interaural axis in stereotaxic coordinates, and from 3 to 5 mm within the lateral fissure. The PBr array was then implanted with a pneumatic insertion device on the surface of the STG [30]. We were unable to use the pneumatic insertion device to insert the lateral fissure array, as the parietal lobe prevented perpendicular access to the medial superior temporal plane. The craniotomy was sealed and temporalis muscle reattached.

2.2.3 Experimental paradigm

Seven macaque vocalization exemplars [31] were delivered free field through a piezoelectric speaker (ES1, Tucker Davis Technologies, Alachua, FL) randomly at 64 dB SPL. There was at least 1 second between each stimulus presentation. Stimuli were

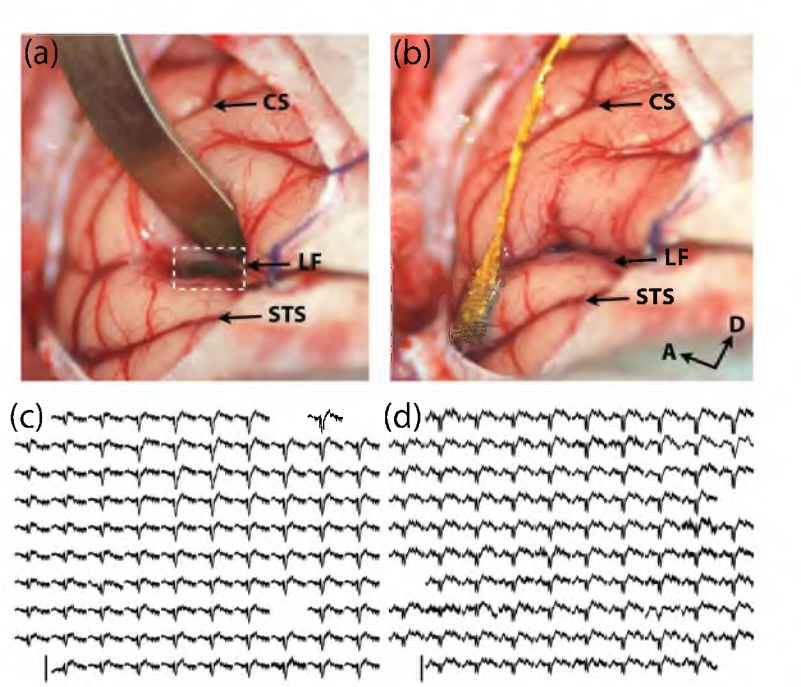


Figure 2.1. Microelectrode array implantation in AI and PBr. (a) Retraction of the parietal lobe to expose the microelectrode array implanted several millimeters within the lateral fissure (white dashed box). (b) The implanted microelectrode array in PBr after AI array implantation. Acronyms: CS: Central Sulcus, LF: Lateral Fissure, STS: Superior Temporal Sulcus, A: Anterior, D: Dorsal. Averaged evoked potentials across the array frootprint in AI (c) and PBr (d) in response to 15 presentations of the *harmonic arch* vocalization. Each evoked potential represents 1.8 seconds of averaged LFP data starting 0.2 seconds before stimulus onset. Amplitude scale bar extends from -500 to 300 microvolts in (c) and -300 to 200 microvolts in (d).

presented while the monkey sat in an acoustically, optically, and electromagnetically shielded chamber (Acoustic Systems, Austin, TX).

Task control software was custom-built using LabVIEW (National Instruments, Austin, TX) and run in real time on a National Instruments PXI-embedded computer system. Digital markers and analog waveforms of auditory stimuli were recorded synchronously with the neural data.

2.2.4 Data collection and preprocessing

Neural data were recorded with a Cerebus System (Blackrock Microsystems). Electrophysiological signals were highpass (0.3 Hz, one pole) and lowpass (7500 Hz, three pole) filtered and pseudo-differentially amplified with a gain of 5000×. High-frequency MUA data were digitally filtered (eight pole, highpass 250 Hz, lowpass 7.5 kHz) and sampled at 30 kHz. Individual spikes were detected offline using t-distribution E-M sorting [32]. Any MUA waveform that exceeded -3.5 RMS times the highpass filtered voltage signal was included as part of the MUA signal for the electrode on which the waveform was recorded. Large motion artifact waveforms, and waveforms with a shorter interspike interval than 1 millisecond were excluded from any analyses. The broadband data were recorded at 30 kHz and later filtered and downsampled to 2 kHz for analysis of the LFP.

LFPs were recorded on 96 electrodes for 144 trials per class in AI (1008 total trials) and on 95 electrodes for 112 trials per class in PBr (784 total trials) over multiple days. To common average rereference our LFP data, the average voltage from all electrodes for each trial was subtracted from the trial-by-trial response of each individual

electrode. Spectrograms were calculated using multitapered analysis with a time-bandwidth product of 5, 9 leading tapers, 200-msec windows, and 10-msec step sizes. Spectrograms with either 77 or 304 frequency bins ranging from 0 to 300 Hz were calculated, to examine the effect of frequency resolution on decode performance.

MUA analysis utilized the same trials used for LFP data (1008 in AI and 784 in PBr). For MUA data, only the electrodes with a significant change in the average firing rate (Wilcoxon rank-sum test, $P<0.05$) between 500 milliseconds before the beginning of the vocalization and 500 milliseconds after the beginning of the vocalization were included in the analysis (65 electrodes in PBr and 62 electrodes in AI). Multitapered spectrograms and peri-stimulus time histograms (PSTHs) were generated using an open source neural analysis package [33]. Single-trial PSTHs and spectrograms were generated separately for four durations after the beginning of each vocalization (200, 400, 800, and 1600 msec). PSTHs were calculated using Gaussian kernels of five widths (5, 10, 25, 50, and 100 msec). These kernel widths constitute the five temporal resolutions we used in examining the effect of temporal resolution on decode performance.

2.2.5 Feature selection and processing

We extended the method of Kellis et al. [34] that simultaneously incorporates features representing dynamics in time, space, and frequency to apply to both MUA data and LFP data. We use the term *class* to refer to one type of conspecific vocalization (e.g., "grunt") and the term *trial* to refer to one instance of a vocalization being played for the monkey. We also use the term channel to refer to the neural signals from one electrode.

To select training features for LFP data, two-dimensional spectrograms were calculated for a subset of seven trials from each class and each recording channel (Figure 2.2). These multidimensional data were unwrapped to produce a two-dimensional matrix in which each row contained all the time and frequency features from all channels for a single trial. The feature matrix was z-scored, orthogonalized using PCA, and projected into the principal component space using a sufficient number of leading principal components to retain 95% of the variance in the data. A subset of seven trials from each class, directly following the training trials in time, was selected for testing the classifier, and spectrograms were calculated for each class and each recording channel (142 training and testing sets in AI and 110 training and testing sets in PBr). These four-dimensional data tensors were matricized, or unwrapped, into a two-dimensional matrix, z-scored, and projected into the principal component space calculated during training.

As with the LFP-based decode, seven trials from each class were selected for training and seven trials following the training trials in time for each class were used as testing data for MUA data (142 training and testing sets in AI and 110 training and testing sets in PBr). For both training and testing, MUA data were collected into a large two-dimensional matrix where each row represented all firing rate data from all channels for a given trial. The data were then z-scored, orthogonalized, and projected into the principal component space using the aforementioned process for LFP data.

2.2.6 Evaluation

Using features derived from either MUA or LFP, data were classified on a trial-by-trial basis using linear discriminant analysis (LDA) [35] (Figure 2.3). All possible

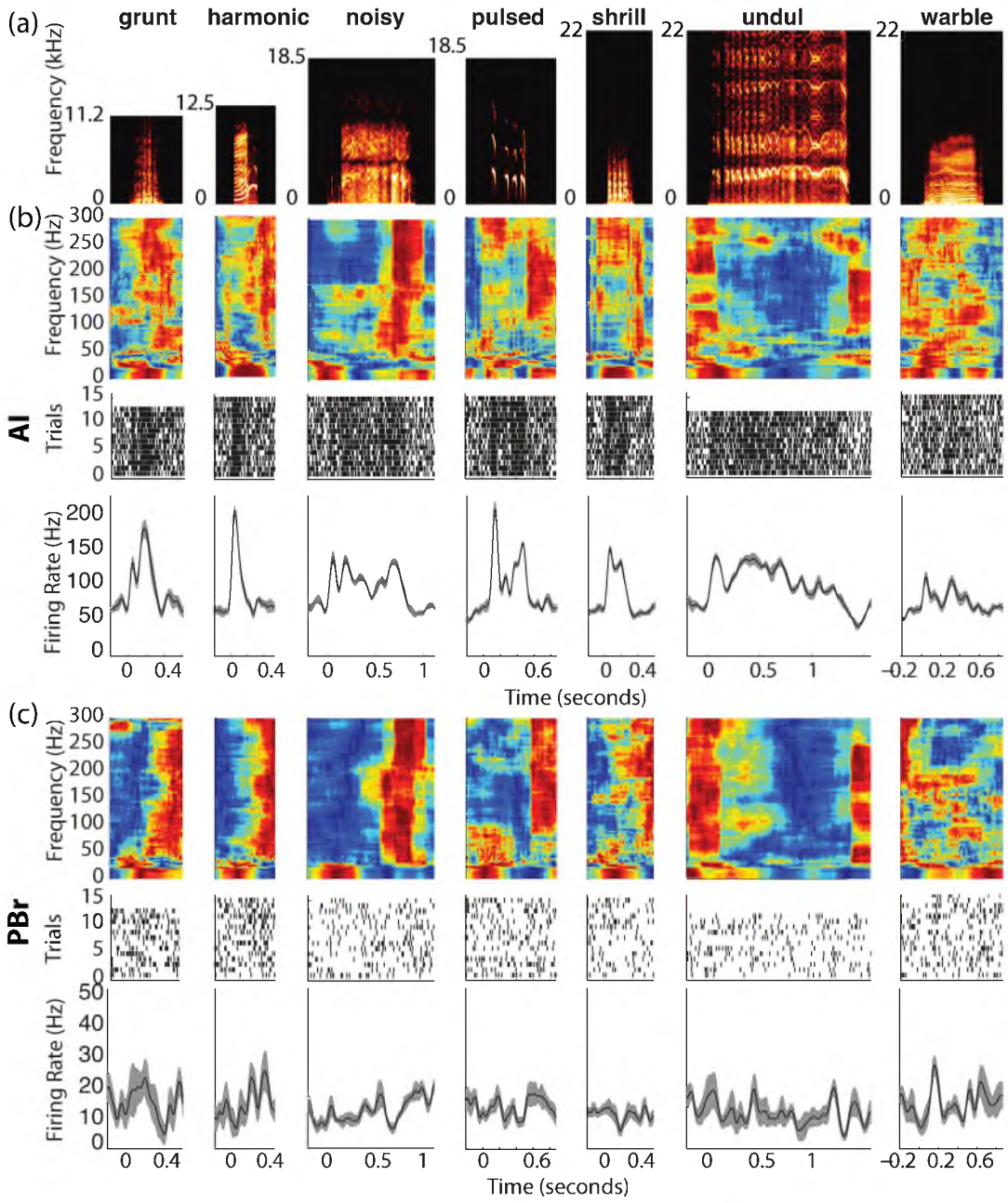


Figure 2.2 Neural data used for classification for each vocalization. (a) Spectrograms of sound stimuli for each vocalization. (b) Trial-averaged spectrograms of LFP responses to vocalizations for a single channel and trial-averaged PSTHs for multiunit responses for AI. (c) Trial averaged spectrograms of LFP responses to vocalizations for a single channel and trial averaged PSTHs for MUA responses for PBr. All PSTH kernel widths are 25 msec.

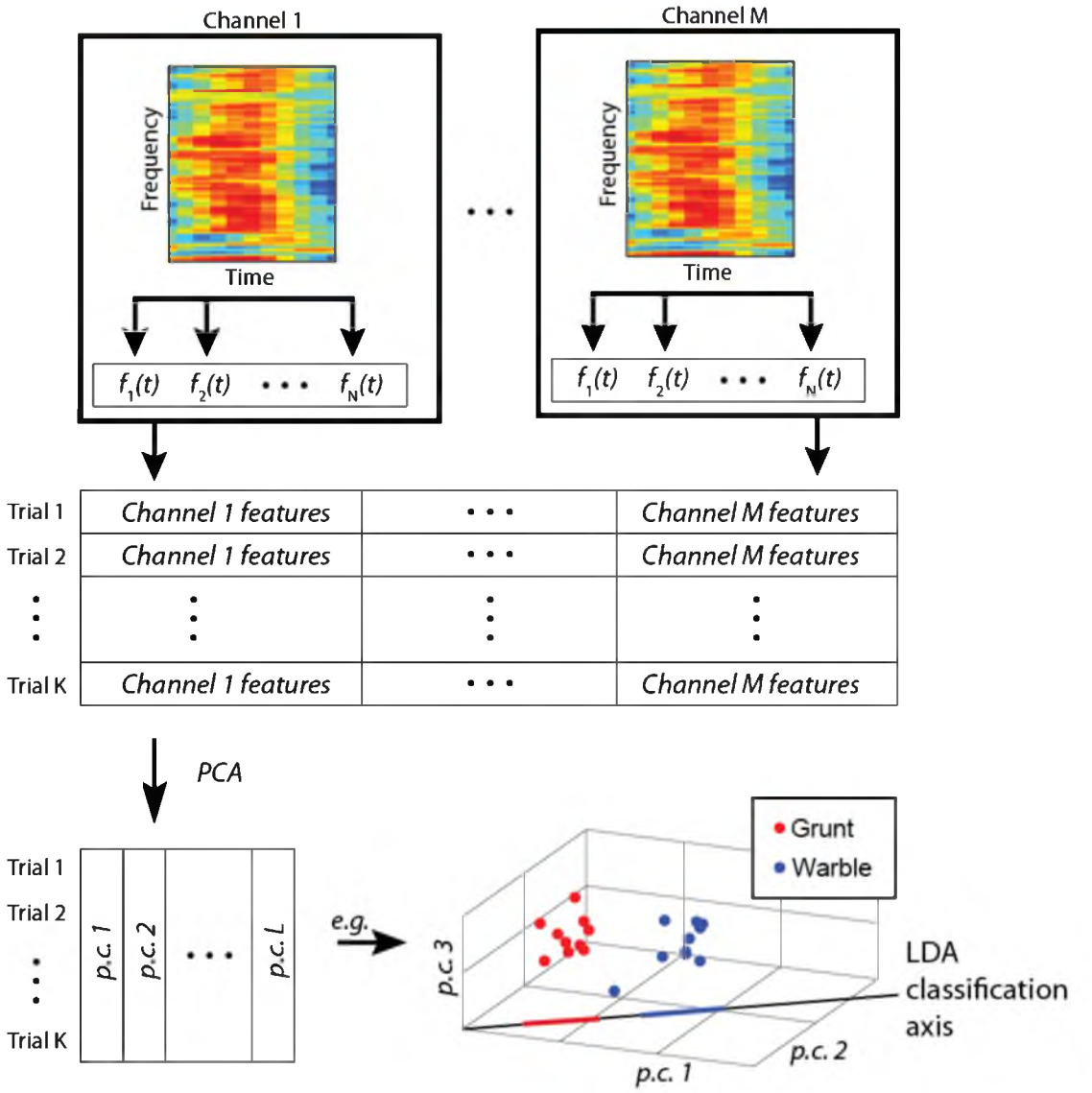


Figure 2.3. Summary of decoding method. We began with spectrograms for each trial over all channels. All neural features were unwrapped along the trials dimension and PCA was applied to this matrix. The PCA reconstruction determined from the training set was then applied to the testing set and LDA was used to classify the identity of the vocalization for each trial. A comparable method was used for MUA-based decodes.

combinations of two through seven classes were evaluated. Only one outcome exists for a combination of seven classes; otherwise, mean, median, standard deviation, interquartile range (IQR), and the 95% confidence interval were computed. We also evaluated performance for each single electrode on the AI microelectrode array and for varying numbers of electrodes (4, 7, 14, 24, 48, and 95 electrodes for LFP; and 4, 8, 16, 31, and 62 electrodes for MUA). These electrodes were chosen by successively removing twice as many electrodes between those used for the decode, while maintaining an even spatial sampling. These combinations, as well as class-by-class comparisons, were evaluated for the best performing durations after data onset (for MUA and LFP data) and best temporal resolutions (for MUA data).

Classification accuracy was measured against the level of chance, which was defined as equal likelihood for all classes. Chance performance therefore ranged from 50.0% for two classes to 14.3% for seven classes. Consistent classification above the level of chance indicated that the decode was finding and operating on relevant features from what could otherwise be stochastic physiological data. That is, the decode is able to predict the stimulus identity based on the similarity of the testing data to the training data.

2.2.7 Information theoretic analysis

We examined the mutual information between the vocalizations and the responses in AI and PBr. Probability distributions $p(s)$, $p(r)$, and $p(s,r)$ were taken from confusion matrices of pair-wise classification frequencies for all MUA- and LFP-based decodes, where $p(s)$ is the sum across one dimension divided by the total number of trials, $p(r)$ is the sum across the other dimension divided by the total number of trials, and $p(s,r)$ is the

diagonal divided by the total number of trials. *p(s,r)* is also the product of the probability that a certain stimulus was shown, *p(s)*, and the probability that a stimulus was shown given that a certain response occurred, *p(s|r)*. The equation

$$I(S;R) = \sum_{s \in S} \sum_{r \in R} p(s,r) \log_2 \left(\frac{p(s,r)}{p(s)p(r)} \right)$$

was then evaluated for these probability distributions. This analysis was performed on the confusion matrix for all pairwise classification frequencies for both MUA- and LFP-based decodes. Mutual information was also calculated between stimulus and response for the five different kernel widths used for MUA-based decodes.

2.3 Results

2.3.1 Decode performance in AI relative to PBr for both MUA and LFP.

Decode performance in AI was above chance, whereas decode performance in PBr for both MUA and LFP was near chance. For MUA-based decodes, the best decode performance in AI was 93.7±10.5% (mean ± std) (IQR: 7.9%; 95% confidence interval: 65.9%≤μ≤100%) for pair-wise comparisons (Figure 2.4) and 73.2% for seven classes. The best MUA-based decode performance in PBr was 51.6±16.6% (IQR: 17.7%; 95% confidence interval: 25.8%≤μ≤90.8%) for pair-wise comparisons and 28.6% for seven classes. For LFP-based decodes, best performance in AI ranged from 83.5±14.1% (IQR: 19.9%; 95% confidence interval 41.3%≤μ≤97.8%) for two classes to 57.3% for seven classes. The best LFP-based decode performance in PBr was 53.8±9.3% (IQR:

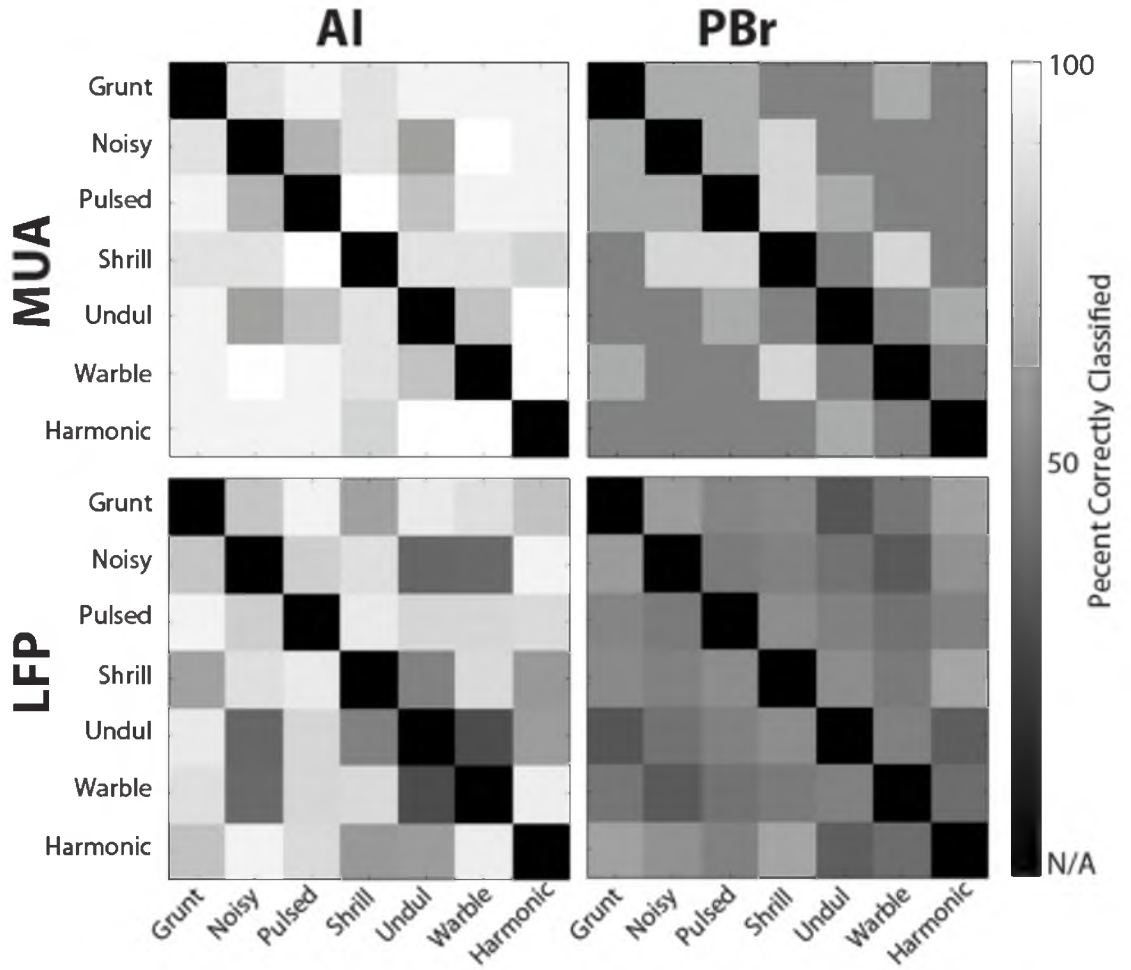


Figure 2.4. Pairwise decode performance for MUA- and LFP-based decodes in AI and PBr. The top two panels show stimulus-response matrices for MUA-based performance for AI and PBr, while the bottom two show LFP-based performance. Mean performance is indicated in grayscale from 0 to 100%. Chance performance is 50%. The black boxes on the diagonal indicate the irrelevant condition of classifying the same stimulus that was presented.

12.1%; 95% confidence interval: 36.8%≤μ≤65.2%) for two classes to 21.2% for seven classes (Figure 2.5).

2.3.2 Temporal and spectral classification dynamics

Since vocalizations were different lengths (mean length = 0.53±0.37 seconds), we felt it necessary to examine the spectral and temporal parameters that may affect decoding vocalizations from AI and PBr. To examine temporal aspects of stimulus decoding, we applied the decode to data of different durations and temporal resolutions. Each temporal resolution for single-trial PSTHs is represented by a different hue line in Figure 2.6(a). In AI, MUA-based decode performance improved with increased temporal resolution (Figure 2.6(a)). The 5-msec temporal resolution outperformed other temporal resolutions (mean performance over data durations = 59.8±6.6%). MUA-based decode results were consistently near chance in PBr, regardless of data duration or kernel width (Figure 2.6(a)). The best performing MUA-based decode in PBr was for the 1600-millisecond data duration and both the 50 and 25 -millisecond temporal resolutions (28.5±0.1 for 50-millisecond temporal resolution and 28.5±0.04 for 25-millisecond temporal resolution).

Temporal dynamics of LFP-based decodes were similar to temporal dynamics of MUA-based decodes. In A1, LFP-based decodes performed best with the 800-millisecond data duration and in PBr decodes performed best with the 400-millisecond data duration (47.3±0.02%) and 22.4%±0.1%, respectively). To examine influence of spectral resolution on classification results, we decoded vocalizations using spectrograms with two different spectral resolutions. Increasing frequency resolution in the

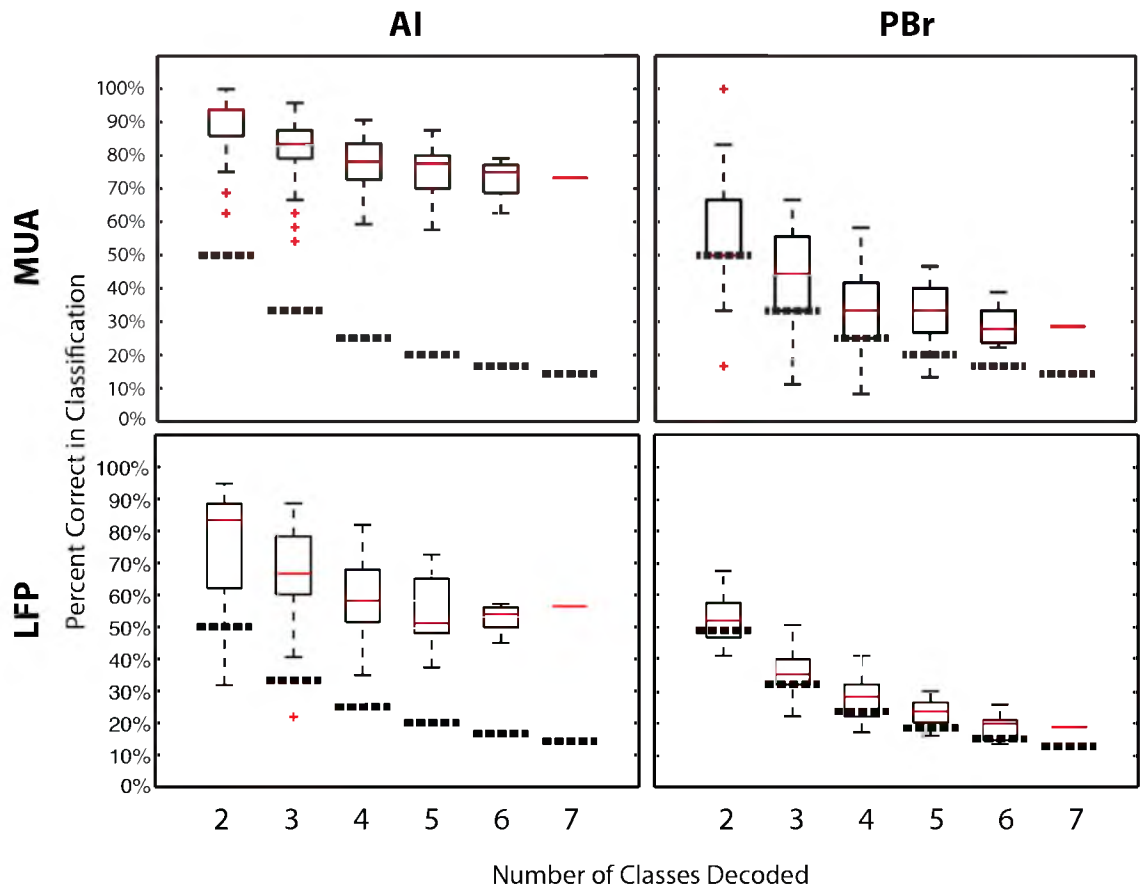
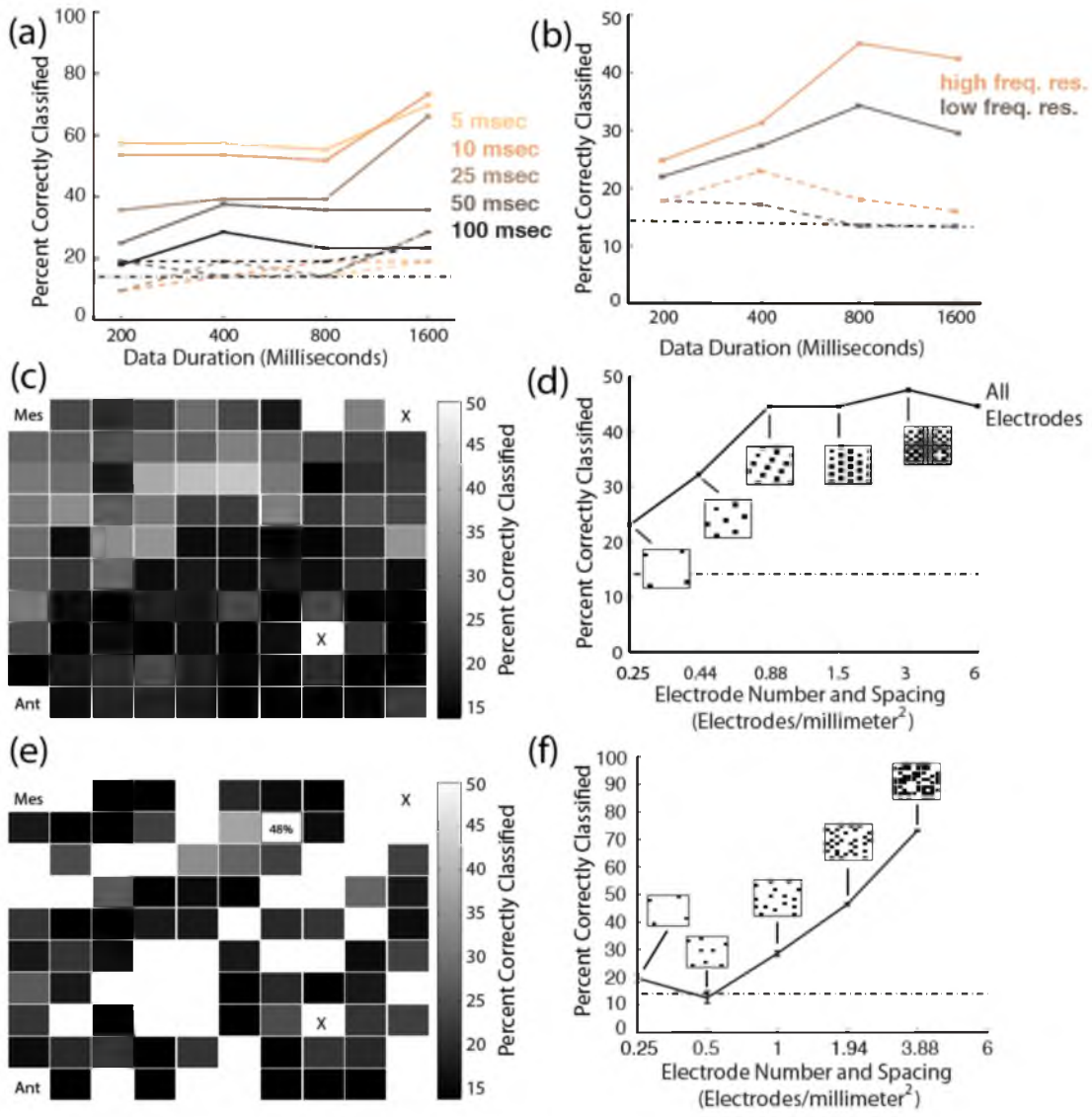


Figure 2.5. Best classifier performance for MUA- and LFP-based decodes. The top two panels show box plots for MUA-based decode performance for AI and PBr, while the bottom two show LFP-based decode performance. Mean performance is indicated by red lines across combinations of vocalizations for two through seven classes. Red crosses indicate outliers. Thick dotted lines indicate chance performance across vocalization combinations.

Figure 2.6. Decode performance over time and space. (a) Seven-class performance for MUA-based decodes. Solid lines show performance for AI and dotted lines show performance for PBr. (b) Seven-class performance for LFP-based decodes. Solid lines show performance for AI and dotted lines show performance for PBr. Brown lines show performance for spectrograms with the lowest frequency resolution. Copper lines show performance for spectrograms with higher frequency resolution. Standard error bars are shown in (a), (b), (d), and (f); which are only slightly larger than the line width. The dash-dotted line in (a), (b), (d), and (f) indicates the level of chance. (c) LFP-based decode performance level for each electrode for seven classes is superimposed onto a map of the AI microelectrode array. "Ant" and "Mes" indicate the corners of the array that are pointing in the anterior direction (Ant) and towards the midline (Mes) on the STP. "X" indicates electrodes that were not connected. (d) LFP-based decode performance for different electrode densities across the array. Seven-class performance is shown for different electrode configurations, which are shown in black on the footprint of the AI electrode array for each electrode density tested. (e) Performance level for each MUA channel superimposed onto a map of electrodes with significant responses. "Ant" and "Mes" indicate the corners of the array that are pointing in the anterior direction and towards the midline on the STP. "X" indicates electrodes that were not connected. (f) MUA-based decode performance for different electrode densities. Seven-class performance is shown for different electrode configurations, which are shown in black on the footprint of the AI electrode array for each electrode density tested.


(a)
Percent Correctly Classified
Data Duration (Milliseconds)
5 msec
10 msec
25 msec
50 msec
100 msec
(b)
high freq. res.
low freq. res.
Percent Correctly Classified
Data Duration (Milliseconds)
(c)
Mes
Ant
Percent Correctly Classified
(d)
All Electrodes
Percent Correctly Classified
Electrode Number and Spacing
(Electrodes/millimeter²)
(e)
Mes
Ant
48%
Percent Correctly Classified
(f)
Percent Correctly Classified
Electrode Number and Spacing
(Electrodes/millimeter²)

spectrograms improved LFP-based decode performance in both AI and PBr (Figure 2.6(b)). While PBr results remained near chance overall, the 400-msec data duration for increased frequency resolution reached 22.6±0.01% performance. In AI, higher frequency resolution also improved performance.

2.3.3 Spatial classification dynamics

Since decode performance was above chance in AI, we further explored decodability on the AI microelectrode array to better understand the topography for interfacing with AI. The decodability of each electrode on the array was examined by executing the LFP- and MUA-based decodes for each individual electrode. Best electrode performance for seven classes ranged from 47.8% (one electrode) to below 15% (each of 46 electrodes) for LFP-based decodes and from 48.6% (one electrode) to below 15% (each of 31 electrodes). The majority of the electrodes performing better than chance were located on the medial half of the electrode array for LFP-based decodes (Figure 2.6(c)). MUA performance was more variable across the array (Figure 2.6(e)).

To examine the spatial scale of information processing in AI we explored LFP- and MUA-based decode performance for fewer, and more sparsely spaced, electrodes on the electrode array. For LFP-based decodes, classifier performance was similar to performance using all electrodes for all electrode densities of 0.88 electrodes/mm^2 and above (14 electrodes). Performance dropped for 7 electrodes (0.44 electrodes/mm^2) to 34% and again for 4 electrodes to 23% (0.25 electrodes/mm^2) (Figure 2.6(d)). Therefore, the minimum spacing required to maintain decode performance in AI is the inverse of the electrode density, (1/0.88 electrodes/mm^2) or 1.14 mm^2/electrode.

2.3.4 Mutual information

To examine information content over all neural responses recorded in AI and PBr, we calculated the mutual information between stimulus and response in the two cortical areas. Mutual information between stimulus and response was calculated for all MUA-based decodes for each temporal resolution. Decode performance in PBr was consistent among temporal resolutions and data durations for MUA-based decodes (Figure 2.7(a)). Mutual information increased in AI for longer data durations and higher temporal resolutions (Figure 2.7(a)).

To compare information content in MUA- and LFP-based decodes, mutual information between stimulus and response was then calculated for the best-performing temporal resolution for MUA-based decodes (5-msec temporal resolution) and best frequency resolution for LFP-based decodes (1024-point) in both AI and PBr. We found that responses in PBr contained little information about the stimulus for all data durations (0.65±0.11 bits for LFP-based decodes and 0.76±0.02 bits for MUA-based decodes). AI decodes contained more information for longer data durations and contained more information than PBr decodes overall (Figure 2.7(a)). In AI, MUA responses contained more information than LFP responses (0.56±0.13 difference in bits across data durations).

2.4 Discussion

We have investigated the nature of LFPs and MUAs from AI and PBr in the context of conspecific vocalization stimulus to explore the potential for interfacing an auditory neural prosthesis directly with the neocortex. We observed that AI provided greater discriminability between vocalizations than PBr using linear statistical

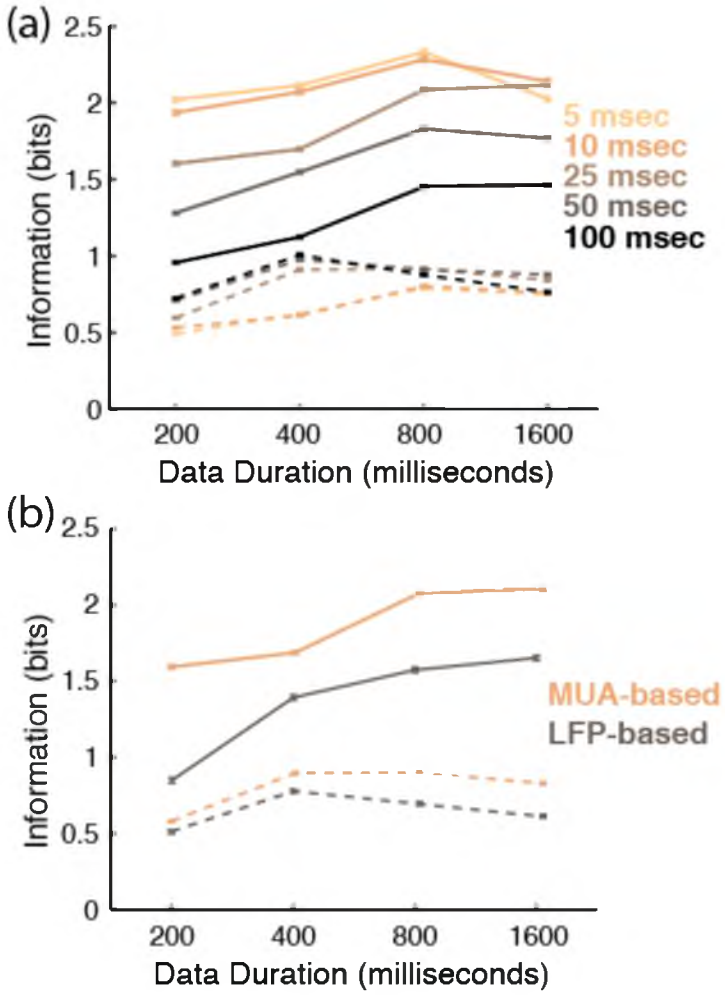


Figure 2.7. Information theoretic analysis. Solid lines show stimulus-response information content across data durations for AI. Dotted lines show stimulus-response information content across data durations for PBr. (a) Information content for five different kernel widths for the MUA-based decode. Different kernel widths are color coded. (b) Information content over data durations using the best performing kernel width and frequency resolution determined from Figure 2.6. Copper lines show information content across data durations for all MUA-based decodes. Brown lines show information content across data durations for all LFP-based decodes.

classification. We found that the decode performance was dynamic in space, time, and frequency. Increased temporal resolution improved MUA-based decode performance, and greater spectral resolution improved LFP-based decode performance. From this primary result, we

suggest that information coding in AI relies on precise dynamics in both time and frequency domains. Spatial analyses estimated a lower limit of electrode spacing at 1.14 mm^2 for electrophysiological interface with AI. While increased spectral resolution improved LFP-based decode performance in PBr, most of the PBr decodes remained near chance. The process of implantation and data acquisition undertaken for this work illustrates the potential for chronic electrophysiological interface with awake, behaving primate AI over a period of months, similar to the type of interface that would be used for a cortical neural prosthesis.

2.4.1 Temporal precision of MUA responses

We found that increased temporal precision yields superior decode results in AI. We also found the highest information content for temporal resolutions of less than 10 milliseconds. Two of the three auditory linear pattern discriminator AP decode studies in macaques have shown increased performance for higher temporal precision (using binned spikes rather than Gaussian kernels) [26, 36]. These results also are in accord with decodes of marmoset vocalizations, where the greatest mutual information from primary auditory cortical neurons utilized bins smaller than 10 milliseconds [25], and with previous studies of temporal integration in AI in the marmoset and macaque showing

high spike timing reliability and short latencies in AI [37, 38]. We can therefore conclude that high temporal precision is an important feature of stimulus coding in AI.

2.4.2 Effect of Duration on decode performance

We have shown that increased data durations in AI improved decode results and provided increased information content up to the 800-millisecond data duration, however information about the stimulus increases in AI up to 1600 milliseconds for both MUA and LFP. This suggests that linear spectrotemporal encoding in AI relies on the temporal precision of firing rates, as opposed to vocalization duration. The best MUA-based decode performance in PBr, however, utilized the 400-millisecond data duration. Whether this means that PBr could be utilizing a firing rate-based coding schema remains to be determined. Effects of data duration may also be a product of working memory capacity, as opposed to auditory processing.

One possible limitation of this study holds true for all auditory decodes of natural stimuli that vary in length: the varying length of vocalizations could be a potential source of artificial discriminability among vocalizations. Trial-averaged spectrograms from both AI and PBr showed a broadband burst of activity coincident with the vocalizations' end, which could contribute to principal component reconstruction and therefore skew the results. However, the raw data suggest that this is not the case. PBr spectrograms demonstrated bursts of low-frequency power at the beginning of the vocalizations, and high-frequency power at the end of the vocalizations, and therefore represented the duration of the vocalization in the raw responses. If these indicators of vocalization length accounted for artificial discriminability between classes, decodes in PBr would

perform much better than the current results suggest. PBr showed very little decodability despite the presence of the spectral information about vocalization length. The AI spectrograms show far more structure during the call, which may account for more overall between-class variance and temporal precision than the features associated with the end of vocalizations. This overall variability is likely what PCA is operating on in the decode.

2.4.3 Frequency resolution of LFP responses

Previous LFP decodes have focused on motor and visual modalities using similar classification methods as the current study [39, 40]. These and other LFP decodes selected features on the basis of defined neural bands. For example, spectral data near the gamma range were the most useful for decoding rat limb movements (70-120 Hz) [39], bistable visual perceptions (50-70 Hz) [40], and high-frequency LFP (100–400 Hz) provided the most information about monkey limb position [41]. Our study used frequencies between 0 and 300 Hertz as features in the decode. We show that increasing the frequency resolution of the spectrograms used in the decode from 400 to 1024 points improves decode performance. Performance was increased in both AI and PBr when frequency resolution was increased.

A recent study reconstructed auditory stimuli that human patients heard using LFP recorded from the surface of the posterior STG with high-density micro-electrocorticographic (μECoG) electrodes as well as standard clinical ECoG electrodes [42]. This study found using high gamma band (75-150 Hz) LFP produces the highest accuracy in reconstructing auditory stimuli. In addition, stimulus reconstruction accuracy

was improved using a nonlinear model, which was based on spectrotemporal modulations, when compared with a linear spectrogram. This result indicates that encoding in PBr may utilize nonlinear encoding schemes, or PBr may represent spectrotemporal modulations in a stimulus more than linear changes in spectrotemporal features of a stimulus. This type of encoding scheme, as well as sparse representations in PBr, or the possibility that vocalizations are not the optimal stimuli for PBr, could account for why PBr decodes were close to chance.

2.4.4 Spatial scale and electrode density

The LFP-based decode results over electrode densities provide evidence for a lower limit on functional electrode density for electrophysiological recording of about one electrode per square millimeter (interelectrode spacing of 1.14 mm). This inter-electrode spacing is larger than those utilized in many penetrating electrode arrays, yet smaller than ECoG interelectrode spacing (~1 cm) and the reported lower limit on spacing is smaller than measurements of LFP activation in acoustic space [43]. The lower limit for electrode spacing found in the current study is almost double the optimal spacing predicted by spatial spectral models, which report a best electrode spacing of 0.6 mm [44]. The reported spacing is, however, on par with μECoG electrodes, which show independent processing at 1- and 1.4-mm electrode spacing [34].

We show that MUA-based decode performance increases as more channels are added to the decode, at a higher density. Whereas classification accuracy for MUA-based decodes over all electrodes is almost 30% higher than the best individual electrode, LFP-based decodes over all electrodes is near the best electrode performance. This result

indicates that variability over electrodes is adding to decode performance to a greater extent in the MUA-based decode than in the LFP-based decode. The smaller spatial scale of information from neurons provides more channels of information.

Electrical microstimulation of the cerebral cortex is likely to act at the scale of LFP and cortical columns, rather than the scale of individual neurons. Recent work has shown that stimulation thresholds in visual cortex are lower when more channels are used for stimulation, and that it was possible to evoke spatially distinct visual percepts with microstimulation of ~1 mm^2 of primary visual cortex [45, 46]. While the spatial resolution results presented here are inadequate for determining perceptual resolution of electrical microstimulation of auditory cortex, these results provide the first fixed geometry electrode array study of primate AI and provide some insight into characterizing spatial elements of an electrical interface with AI. How spatial and temporal parameters of microstimulation of auditory cortex affect perceptual discrimination remains an open question that needs to be examined through micro-stimulation of auditory cortex.

2.4.5 Implications for an auditory cortical neural prosthesis

Although the cochlear implant is a successful neural prosthesis for the treatment of deafness, there is some need for an auditory neural prosthesis that can bypass the auditory nerve. There are at least 100 cases per year in the United States of cochlear nerves being destroyed by neuromas brought on by neurofibromatosis type II. These patients would be excellent candidates for a cortical auditory neural prosthesis. In addition to showing functionality of the type of neural interface that could be used in a

stimulating cortical prosthetic, the current study provides insight into the physiology of the auditory ventral stream that is useful for design and use of a neural prosthesis. Through exploration of cortical coding we address several aspects of implementing an auditory cortical prosthesis.

While microstimulation of AI in trained rats has been shown to elicit behavioral responses similar to those elicited by tones [3], whether this stimulation evokes auditory perception that is useful in constructing an auditory object remains to be determined. Intracortical microstimulation that allowed for discrimination behavior in rats used a 200-Hz stimulation rate [3]. The current study suggests that high temporal precision is an important feature of coding in AI. Rapidly dynamic temporal stimulation patterns may therefore prove more successful for encoding information in an auditory cortical neural prosthesis.

In addition to potentially meeting the needs of a patient group without an auditory nerve, an additional benefit is the possibility for achieving increased frequency resolution through interface with core auditory cortex. This concept is derived from the large accessible area of core auditory cortex (approximately 24 mm^2 in humans and 20 mm^2 in macaques) [47]. It is possible to fit several 96-electrode microelectrode arrays on the human core auditory cortex, thereby oversampling the tonotopic map in two core areas (AI and rostral core) [47]. Between the lower limit on electrode density determined by electrophysiology in this study and the actual density of the electrode arrays we used, 112 to 768 electrodes could be fit on macaque core auditory cortex.

Because of vascularization and cortical folding, application of penetrating electrode arrays in the human auditory cortex may face challenges, yet methods exist to

implant depth electrodes with microelectrode contacts similar to those in a μECoG, which can take advantage of the full tonotopic map in humans [48]. Early human visual prostheses utilized surface electrode grids to evoke phosphenes from stimulation of primary visual cortex [49] and macroelectrodes on STG to evoke varied auditory perceptions [8]. Stimulating with high-density μECoG in primary cortical areas may be a functional alternative to using penetrating electrode arrays, which may have complex interactions with the cerebral cortex [45]. A recent study, which could have particular application to primate AI, provided proof of principle for stimulating the brain with a high-density μECoG inside a sulcus [50].

There has been a great deal of work done on implantable neural prostheses for the cochlear nucleus, as well as the inferior colliculus, to address the need to bypass the auditory nerve. The potential benefits of bypassing these areas and stimulating the cortex directly are surgical ease, patient safety, and increased frequency resolution. This study suggests that high, dynamic stimulation rates in AI could be a feasible solution for dramatically increasing the channel counts and frequency resolution with a cortical auditory neural prosthesis.

2.5 Conclusions

We have examined auditory stimulus coding at early stages along the hierarchical processing stream in neocortex in order to assess the possibility for a cortically based auditory neural prosthesis. We report a lower limit on electrode spacing for electrophysiological recording interface with AI, encoding in AI may improve with higher-resolution temporal and spectral information, and linear spectrotemporal coding

for stimulus identity is higher in AI than PBr. These results together serve as a design input for a human auditory cortical neural prosthesis, and provide guidance on interface design, microstimulation parameters, and anatomical placement.

2.6 References

[1] M. S. Schwartz, S. R. Otto, R. V. Shannon, W. E. Hitselberger, and D. E. Brackmann, "Auditory brainstem implants," *Neurotherapeutics: The Journal of the American Society for Experimental Neurotherapeutics,* vol. 5 (1), pp. 128-36, 2008.

[2] H. H. Lim, M. Lenarz, and T. Lenarz, "Auditory midbrain implant: A review," *Trends in Amplification,* vol. 13 (3), pp. 149-80, 2009.

[3] K. J. Otto, P. J. Rousche, and D. R. Kipke, "Cortical microstimulation in auditory cortex of rat elicits best-frequency dependent behaviors," *Journal of Neural Engineering,* vol. 2 (2), pp. 42-51, 2005.

[4] P. J. Rousche and R. A. Normann, "Chronic intracortical microstimulation (icms) of cat sensory cortex using the utah intracortical electrode array," *IEEE Transactions on Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society,* vol. 7 (1), pp. 56-68, 1999.

[5] S. R. Sinha, N. E. Crone, R. Fotta, F. Lenz, and D. F. Boatman, "Transient unilateral hearing loss induced by electrocortical stimulation," *Neurology,* vol. 64 (2), pp. 383-385, 2005.

[6] A. J. Fenoy, M. A. Severson, I. O. Volkov, J. F. Brugge, and M. A. Howard, "Hearing suppression induced by electrical stimulation of human auditory cortex," *Brain Research,* vol. 1118 (1), pp. 75-83, 2006.

[7] W. H. Dobelle, S. S. Stensaas, M. G. Mladejovsky, and J. B. Smith, "A prosthesis for the deaf based on cortical stimulation," *The Annals of Otology, Rhinology, and Laryngology,* vol. 82 (4), pp. 445-463, 1973.

[8] W. Penfield, "Some mechanisms of consciousness discovered during electrical stimulation of the brain," *Proceedings of the National Academy of Sciences of the United States of America,* pp. 51-66, 1958.

[9] A. Selimbeyoglu and J. Parvizi, "Electrical stimulation of the human brain: Perceptual and behavioral phenomena reported in the old and new literature," *Frontiers in Human Neuroscience,* vol. 4 (1), pp. 1-11, 2010.

[10] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex (New York, NY : 1991),* vol. 1 (1), pp. 1-47, 1991.

[11] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[12] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[13] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex,* vol. 1 (1), pp. 1-47, 1991.

[14] T. A. Hackett, "Information flow in the auditory cortical network," *Hearing Research,* vol. 271 (1-2), pp. 133-46, 2011.

[15] J. A. Winer and C. C. Lee, "The distributed auditory cortex," *Hearing Research,* vol. 229 (1-2), pp. 3-13, 2007.

[16] H. E. Hefner and R. S. Heffner, "Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by japanese macaques," *Journal of Neurophysiology,* vol. 56 (3), pp. 683-701, 1986.

[17] A. U. Turken and N. F. Dronkers, "The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses," *Frontiers in Systems Neuroscience,* vol. 5 pp. 1-20, 2011.

[18] L. M. Romanski, J. F. Bates, and P. S. Goldman-Rakic, "Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey," *Journal of Comparative Neurology,* vol. 403 (2), pp. 141-57, 1999.

[19] L. M. Romanski, B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker, "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex," *Nature Neurosci,* vol. 2 (12), pp. 1131-6, 1999.

[20] T. A. Hackett, I. Stepniewska, and J. H. Kaas, "Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys," *Journal of Comparative Neurology,* vol. 394 (4), pp. 475-95, 1998.

[21] T. M. Woods, S. E. Lopez, J. H. Long, J. E. Rahman, and G. H. Recanzone, "Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey," *Journal of Neurophysiology,* vol. 96 (6), pp. 3323-3337, 2006.

[22] Y. Kikuchi, B. Horwitz, and M. Mishkin, "Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane," *Journal of Neuroscience,* vol. 30 (39), pp. 13021-13030, 2010.

[23] B. Tian, D. Reser, A. Durham, A. Kustov, and J. P. Rauschecker, "Functional specialization in rhesus monkey auditory cortex," *Science,* vol. 292 (5515), pp. 290-3, 2001.

[24] C. I. Petkov, C. Kayser, T. Steudel, K. Whittingstall, M. Augath, and N. K. Logothetis, "A voice region in the monkey brain," *Nature Neuroscience,* vol. 11 (3), pp. 367-374, 2008.

[25] J. W. H. Schnupp, T. M. Hall, R. F. Kokelaar, and B. Ahmed, "Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex," *Journal of Neuroscience,* vol. 26 (18), pp. 4785-95, 2006.

[26] G. H. Recanzone, "Representation of con-specific vocalizations in the core and belt areas of the auditory cortex in the alert macaque monkey," *Journal of Neuroscience,* vol. 28 (49), pp. 13184-13193, 2008.

[27] B. E. Russ, A. L. Ackelson, A. E. Baker, and Y. E. Cohen, "Coding of auditory-stimulus identity in the auditory non-spatial processing stream," *Journal of Neurophysiology,* vol. 99 (1), pp. 87-95, 2008.

[28] B. B. Averbeck and L. M. Romanski, "Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex," *Journal of Neuroscience,* vol. 26 (43), pp. 11023-33, 2006.

[29] S. Nirenberg and C. Pandarinath, "Retinal prosthetic strategy with the capacity to restore normal vision," *Proceedings of the National Academy of Sciences of the United States of America,* 2012.

[30] P. J. Rousche and R. A. Normann, "A method for pneumatically inserting an array of penetrating electrodes into cortical tissue," *Annals of Biomedical Engineering,* vol. 20 (4), pp. 413-422, 1992.

[31] J. P. Rauschecker, B. Tian, and M. Hauser, "Processing of complex sounds in the macaque nonprimary auditory cortex," *Science,* vol. 268 (5207), pp. 111-4, 1995.

[32] S. Shoham, M. R. Fellows, and R. A. Normann, "Robust, automatic spike sorting using mixtures of multivariate t-distributions," *Journal of Neuroscience Methods,* vol. 127 (2), pp. 111-22, 2003.

[33] H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, and P. P. Mitra, "Chronux: A platform for analyzing neural signals," *Journal of Neuroscience Methods,* vol. 192 (1), pp. 146-51, 2010.

[34] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering,* vol. 7 (5), p. 056007, 2010.

[35] W. J. Krzanowski, *Principles of multivariate analysis*, Oxford University Press, USA; Rev Sub edition. December 28, 2000.

[36] B. E. Russ, A. L. Ackelson, A. E. Baker, and Y. E. Cohen, "Coding of auditory-stimulus identity in the auditory non-spatial processing stream," *Journal of Neurophysiology,* vol. 99 (1), pp. 87-95, 2008.

[37] D. Bendor and X. Wang, "Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys," *Journal of Neurophysiology,* vol. 100 (2), pp. 888-906, 2008.

[38] B. H. Scott, B. J. Malone, and M. N. Semple, "Transformation of temporal processing across auditory cortex of awake macaques," *Journal of Neurophysiology,* pp. 1-75, 2010.

[39] M. W. Slutzky, L. R. Jordan, E. W. Lindberg, K. E. Lindsay, and L. E. Miller, "Decoding the rat forelimb movement direction from epidural and intracortical field potentials," *Journal of Neural Engineering,* vol. 8 (3), p. 036013, 2011.

[40] Z. Wang, N. K. Logothetis, and H. Liang, "Decoding a bistable percept with integrated time-frequency representation of single-trial local field potential," *Journal of Neural Engineering,* vol. 5 (4), pp. 433-42, 2008.

[41] J. Zhuang, W. Truccolo, C. Vargas-Irwin, and J. P. Donoghue, "Decoding 3-d reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex," *IEEE Transactions on Bio-medical Engineering,* vol. 57 (7), pp. 1774-1784, 2010.

[42] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biology,* vol. 10 (1), p. e1001251, 2012.

[43] Y. Kajikawa and C. E. Schroeder, "How local is the local field potential?" *Neuron,* vol. 72 (5), pp. 847-858, 2011.

[44] M. W. Slutzky, L. R. Jordan, T. Krieg, M. Chen, D. J. Mogul, and L. E. Miller, "Optimal spacing of surface electrode arrays for brain-machine interface applications," *Journal of Neural Engineering,* vol. 7 (2), p. 26004, 2010.

[45] K. Torab, T. S. Davis, D. J. Warren, P. A. House, R. A. Normann, and B. Greger, "Multiple factors may influence the performance of a visual prosthesis based on intracortical microstimulation: Nonhuman primate behavioural experimentation," *Journal of Neural Engineering,* vol. 8 (3), p. 035001, 2011.

[46] R. P. Tyler Davis, Paul House, Elias Bagley, Suzanne Wendelken, Richard A Normann and Bradley Greger, "Spatial and temporal characteristics of v1 microstimulation during chronic implantation of a microelectrode array in a behaving macaque," *Journal of Neural Engineering,* 2012.

[47] T. A. Hackett, T. M. Preuss, and J. H. Kaas, "Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans," *Journal of Comparative Neurology,* vol. 441 (3), pp. 197-222, 2001.

[48] C. G. Reddy, N. S. Dahdaleh, G. Albert, F. Chen, D. Hansen, K. Nourski, H. Kawasaki, H. Oya, and M. A. Howard, "A method for placing heschl gyrus depth electrodes," *Journal of Neurosurgery,* vol. 112 (6), pp. 1301-7, 2010.

[49] W. H. Dobelle, J. Turkel, D. C. Henderson, and J. R. Evans, "Mapping the representation of the visual field by electrical stimulation of human visual cortex," *American Journal of Ophthalmology,* vol. 88 (4), pp. 727-35, 1979.

[50] T. Matsuo, K. Kawasaki, T. Osada, H. Sawahata, T. Suzuki, M. Shibata, N. Miyakawa, K. Nakahara, A. Iijima, N. Sato, K. Kawai, N. Saito, and I. Hasegawa, "Intrasulcal electrocorticography in macaque monkeys with minimally invasive neurosurgical protocols," *Frontiers in Systems Neuroscience,* vol. 5 pp. 1-9, 2011.

CHAPTER 3

INFORMATION TRANSFER ALONG THE VENTRAL AUDITORY PROCESSING STREAM IN THE AWAKE MACAQUE

3.1 Introduction

Neocortical sensory processing is thought to occur simultaneously in multiple cortical streams as the brain transforms sensory information into more cognitive representations. Support for this hierarchical processing in visual and auditory cortex stems from neuroanatomical studies outlining projection patterns through the cortical circuit [1-4], and from lesion studies [5, 6]. These studies have identified two processing streams: spatial elements of a stimulus are processed in a "dorsal" stream extending along the parietal lobe, and stimulus identity or quality is processed in a "ventral" stream extending rostrally along the temporal lobe. While there is compelling neurophysiological evidence for a "dorsal" stream in audition [7], physiological studies of a "ventral" stream are limited.

Researchers have examined selectivity of neurons along the ventral auditory processing stream using differences in firing rate [8], statistical classifiers [9], and mutual information [10]. None of these measures, however, examines cortical physiology in the context of information transfer along the processing stream. Modeling work and studies in

other sensorimotor modalities have used cross-correlation [11], Granger causality [12], and directed mutual information [13] to inform the understanding of cortical connectivity and information flow in cortical circuits. These measures all infer causality from a time delay.

Conditional mutual information (CMI) is the mutual information between two random variables, given a third random variable is known. This measure has been used to infer connectivity from resting fMRI data [14]. For sensory cortex, CMI allows one to calculate the information between a sensory stimulus and a neural response, given that the neural response in another cortical area is known. CMI is therefore a relatively simple way of analyzing information transfer among multiple, interconnected cortical areas. The advantages of CMI in neural responses are that CMI does not rely on Gaussian statistics, or infer causality from a time delay.

The goal of the present study was to characterize the physiology of the two cortical areas in the auditory ventral stream in the context of a processing stream. We therefore examined information transfer of neural responses in two cortical areas in an awake primate. Specifically we looked at action potential (AP) and local field potential (LFP) responses to species-specific vocalizations from primary auditory cortex (AI) and rostral parabelt auditory cortex (PBr).

3.2 Materials and methods

3.2.1 Research subject

The neural data examined in this study were recorded from two microelectrode arrays chronically implanted in an adult male rhesus macaque. Each microelectrode array

contained 96 electrodes and was implanted in AI and PBr (Figure 3.1). All experiments were performed according to NIH guidelines for animal care and use and with approval from the University of Utah Institutional Animal Care and Use Committee.

3.2.2 Experimental setup

Each of seven macaque vocalization exemplars was randomly presented free field through a piezoelectric speaker (ES1, Tucker Davis Technologies, Alachua, FL). There was at least 1 second in time between each stimulus presentation. Stimuli were presented while the monkey sat in an acoustically, optically, and electromagnetically shielded chamber (Acoustic Systems, Austin, TX).

3.2.3 Data acquisition and preprocessing

Neural data were recorded with a Cerebus System (Blackrock Microsystems, Salt Lake City, UT). The broadband data were recorded at 30 kHz and later filtered and down-sampled to 2 kHz for analysis of the LFP. High-frequency AP data were digitally filtered (eight pole, high-pass 250Hz, low-pass 7.5kHz) and sampled at 30 kHz. Individual spikes were detected offline using t-distribution E-M sorting [15]. Only units that were well isolated from the noise cluster in 3-D PCA space were retained for further analysis (12 in PBr, and 24 in AI).

LFPs were recorded on 96 channels for 144 trials per class in AI (1008 total trials), and on 95 channels for 112 trials per class in PBr (784 total trials) over multiple weeks. Multitapered spectrograms and peri-stimulus time histograms (PSTHs) were generated using an open source neural analysis package (www.chronux.org)

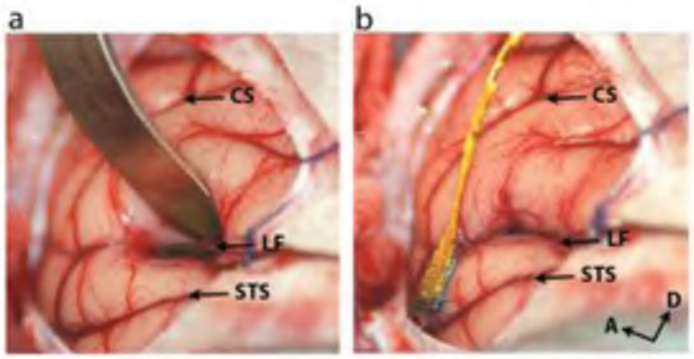


Figure 3.1. Electrode arrays implanted in AI and PBr. (a) AI and (b) PBr. CS, Central Sulcus; LF, Lateral Fissure; STS, Superior Temporal Sulcus; A, Anterior; D, Dorsal.

[16]. Single-trial PSTHs and LFP spectrograms were generated separately for four durations after the beginning of each vocalization (200, 400, 800, and 1600 msec). PSTHs were calculated using Gaussian kernels of five widths (5, 10, 25, 50, and 100 msec). Spectrograms were calculated using multitapered analysis with a time-bandwith product of 5, 9 leading tapers, 200 msec windows, 10 msec step sizes, and using 1024-point fast Fourier transforms.

3.2.4 Statistical classification

We extended the method of Kellis et al. [17] to simultaneously incorporate features representing dynamics in time, space, and frequency for both AP data and LFP data. These are similar to the neural features mentioned in Chapter 2, except we used isolated single units here. We use the term *class* to refer to one type of species-specific vocalization (e.g., "grunt"), and the term *trial* to refer to one instance of a vocalization being played for the monkey.

To select training features for LFP responses, two-dimensional spectrograms were calculated for a subset of trials from each class and each recording channel. These multidimensional data were unwrapped to produce a two-dimensional matrix in which each row contained all the time and frequency features from all channels for a single trial. The feature matrix was z-scored, orthogonalized using principal component analysis (PCA), and projected into the principal component space using a sufficient number of leading principal components to retain 95% of the variance in the data. A different subset of trials was selected for testing the classifier, and spectrograms were calculated for each class and each recording channel. These data were unwrapped into a two-dimensional

matrix, z-scored, and projected into the same principal component space calculated for training features.

To produce training and testing sets from AP data, we selected consecutive days for which at least four common channels had units with similar firing responses (six pairs of days for AI, four units for each day; two pairs of days for PBr, six units for each day). For each pair of consecutive days, the data from the first day were used for training, and data from the second day were used as testing. For both training and testing, AP data were collected into a large two-dimensional matrix where each row represented all firing rate data from each unit for a given trial. Then, the data were z-scored, orthogonalized, and projected into the principal component space using the same process described above for LFP data.

The classifier was evaluated for all pair-wise combinations of vocalizations using linear discriminant analysis. Pairwise classifications were examined in order to determine probability distributions to be used in information theoretic analyses.

3.2.5 Information theoretic analyses

To examine cortical information transfer we examined CMI between a stimulus and a response in one cortical area given that there was a response in the other cortical area. Let R_x and R_y be two random variables representing neural response classes in two different cortical areas, we calculated CMI as follows:

$$I(S;R_x|R_y) = \sum_{s\in S}\sum_{x\in R_x}\sum_{y\in R_y} p(s,x,y)\log_2\left(\frac{p(y)p(s,x,y)}{p(s,y)p(x,y)}\right).$$

where probability distributions $p(y)$, $p(s,y)$, $p(x,y)$, and $p(s,x,y)$ were taken from confusion matrices of pair-wise classification frequencies for all AP and LFP responses (Figure 3.2). Equation (1) was evaluated for R_x being responses in AI and R_y being responses in PBr to calculate CMI for stimulus and response in AI given that the response in PBr is known ($I(S;R_{AI}\,|R_{PBr})$). CMI for stimulus and response in PBr given a known response in AI ($I(S;R_{PBr}\,|R_{AI})$) was calculated in the same way. Mean information values across data durations are reported as mean ± standard deviation.

3.3 Results

3.3.1 Transfer information varies for different time scales of auditory processing

To examine cortical information transfer between AI and PBr for the seven species-specific vocalization stimuli, we calculated the mutual information between an auditory stimulus and the neural response in one cortical area given the response in the other cortical area was known. For AP responses, when CMI is averaged over all kernel-widths, information in AI given PBr was greater than the information in PBr given A1 for all four data durations ($I(S;R_{AI}\,|R_{PBr})$ - $I(S;R_{PBr}\,|R_{AI})$ = 1.68 ± 0.51 bits) (Figure 3.3a). For LFP responses, the information in A1 given PBr was also greater than the information in PBr given A1 for the 400, 800 and 1600 millisecond data durations ($I(S;R_{AI}\,|R_{PBr})$ - $I(S;R_{PBr}\,|R_{AI})$ = 0.49 ± 0.40 bits). However, for the 200 millisecond data duration for LFP responses information in PBr given A1 was greater than the information in A1 given PBr ($I(S;R_{PBr}\,|R_{AI})$ - $I(S;R_{AI}\,|R_{PBr})$ = 3.33 bits) (Figure 3.3b). The highest

$$I(S;R_x \mid R_y) = \sum_{s \in S} \sum_{x \in R_x} \sum_{y \in R_y} p(s,x,y) \log_2 \left(\frac{p(y)\,p(s,x,y)}{p(s,y)\,p(x,y)} \right)$$

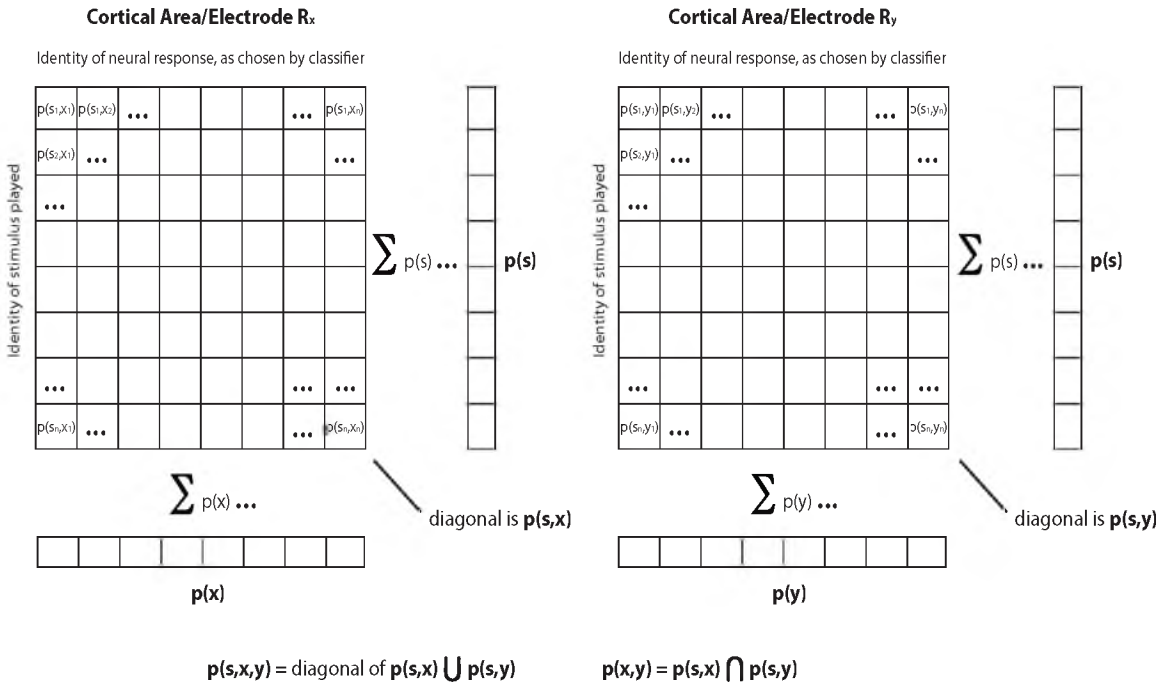


Figure 3.2. Probability distributions derived from pairwise classification frequencies for conditional mutual information. Confusion matrices of pairwise classification frequencies are shown for cortical areas R_x and R_y. Visual representations of derivation of probability distributions used in the CMI equation are outlined with the derived probability distribution shown in bold. Probability distributions which include both R_x and R_y are described below as being derived from the unions or intersections of the two confusion matrices shown above.

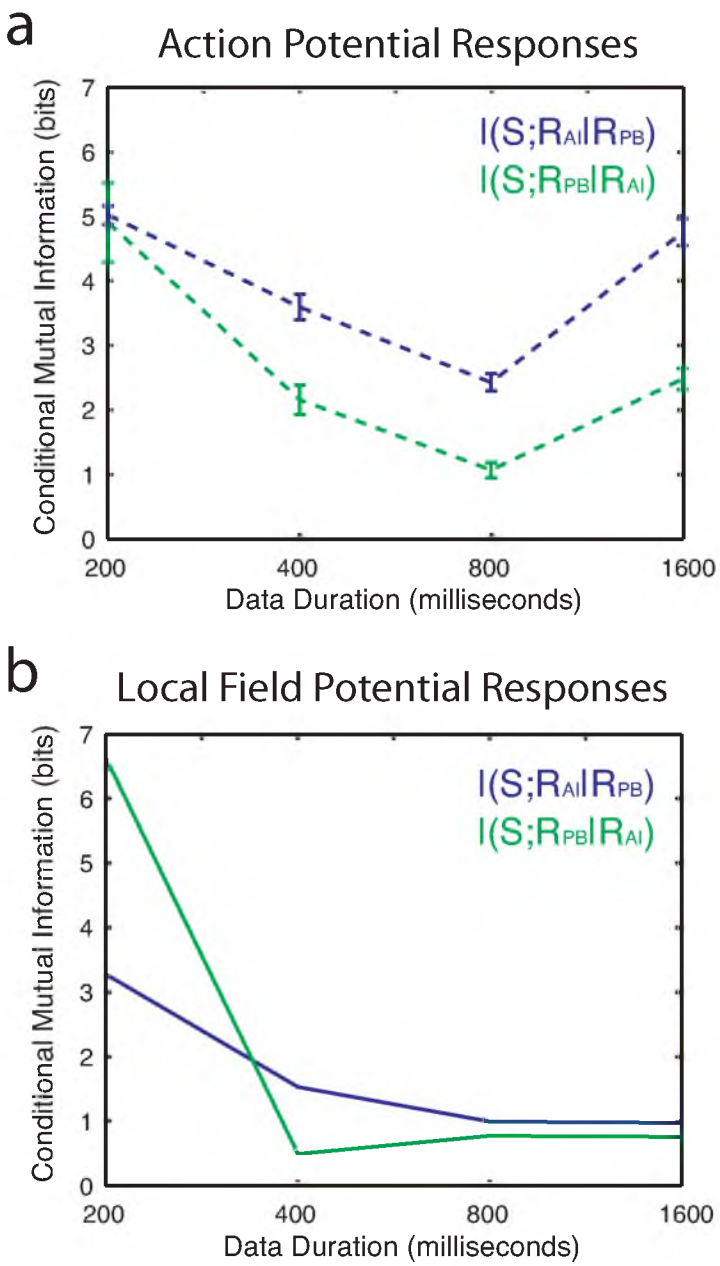


Figure 3.3. CMI in LFP responses and average AP responses. (a) Average CMI for AP responses. Average CMI over kernel-widths for AP data is shown. Bars represent standard error. (b) CMI for LFP responses. In both (a) and (b), blue lines indicate CMI between stimulus and responses in AI given responses in PBr are known and green lines indicate CMI between stimulus and AP responses in PBr given the response in AI is known.

CMI observed across all LFP responses was for the 200-millisecond data duration for responses in PBr, given responses in AI (6.64 bits) (Figure 3.3b).

3.3.2 Transfer information for short time-scale AP responses is dependent on PSTH temporal resolution

While the bias of information transfer reverses for 200 millisecond versus 400, 800 and 1600 millisecond LFP responses, this change in direction is not evident for the AP responses averaged over the different kernel-widths. However, this change in bias is apparent for the 25 and 50 millisecond kernel widths (red and blue lines are higher for the 200 millisecond data duration in Figure 3.4b, whereas all colored lines are lower in Figure 3.4a). The highest CMI observed for AP responses was for the 200 millisecond data duration for responses in PBr, given responses in AI for the 50-millisecond kernel width (8.77 bits) (Figure 3.4b).

3.4 Discussion

At the beginning of a cortical response to an auditory stimulus, information transfer is largely feed-forward, or dependent on lower auditory nuclei to filter and propagate the transduced neural information to the cerebral cortex. Then, as the identity of a stimulus is recognized, information transfer feeds back from higher cortical areas. The current results reflect this schema by showing the first 200 milliseconds of a response in PBr being informed by the first 200 milliseconds of a response in AI, and longer responses in AI being informed by longer responses in PBr. In examining CMI between AI and PBr, we found that information in AI given responses in PBr was higher for the 400, 800 and 1600

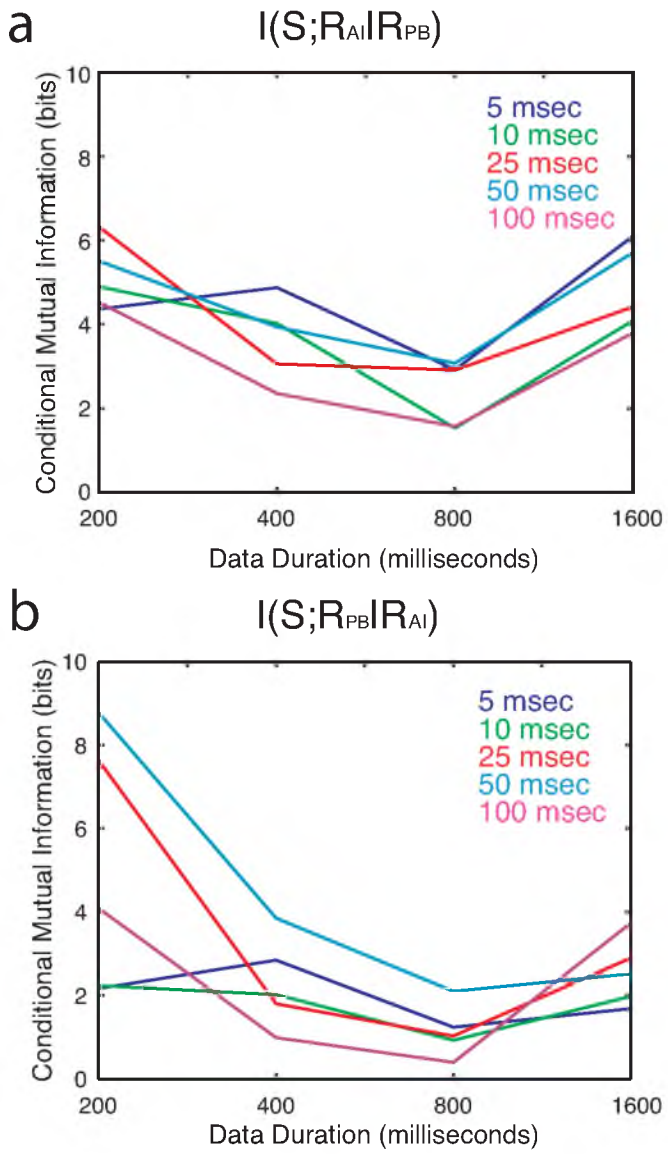


Figure 3.4. CMI in AP responses over kernel widths. (a) CMI between stimulus and AP responses in AI given the response in PBr is known. Shows CMI for action potential responses for different data durations and PSTH kernel-widths (colored lines). (b) CMI between stimulus and AP responses in PBr given the response in AI is known. Shows CMI for action potential responses for different data durations and PSTH kernel widths (colored lines), as in (a).

millisecond data durations. Interestingly, information in PBr given responses in AI was highest for the 200 millisecond data duration for LFP responses. Information from the 200 millisecond data duration in PBr given responses in AI was also higher for two of the five kernel widths (25 and 50 milliseconds).

Responses to full-length vocalizations in AI contained more information about the stimulus given that responses in PBr were known, when compared with information in PBr given responses in AI were known. The duration of the vocalizations we played ranged from 0.23 to 1.27 seconds (average vocalization length was 0.53 ± 0.37 seconds). The 400 through 1600 millisecond data durations therefore each contain the full duration of at least one vocalization, whereas the 200 millisecond data duration does not contain the full duration of any of the seven vocalizations. These results accord with the theory that cortico-cortical feedback is evident in offset, or at least later, neural responses.

The output of our classifier provided more information about the first 200 milliseconds of responses in PBr given that the first 200 milliseconds of responses in AI were known. Further examination of the CMI from AP responses shows that CMI from the 5, 10 and 100 millisecond kernel-widths is lower in PBr given AI than that in AI given PBr. The lack of information in PBr given AI for the 5 and 10 millisecond kernel-widths points to a particular window of temporal integration in PBr. A recent study examining temporal integration in macaque auditory cortex shows that neurons from the belt region of auditory cortex (between AI and parabelt in the processing stream) have difficulty synchronizing to stimulus modulations above and below 5 Hz [18]. PBr may have a similar window of temporal integration, as the high information for the 25 and 50 millisecond kernel-widths corresponds to a temporal integration window of between 20

and 40 Hz. The 100 millisecond kernel-width is wide enough to be interpreted as an average firing rate for the 200 millisecond data duration. Therefore, we can conclude that temporal fluctuations in the neural responses, in addition to average firing rate, provide information about a stimulus in the first 200 milliseconds in PBr. Together these results suggest that coding in PBr relies on information from overall firing rate and precise temporal fluctuations of responses in AI, though neurons in PBr might not be able to synchronize to rapid temporal modulations, like those in AI.

3.5 Conclusions

We have presented a study of neural responses in two cortical areas along the ventral processing stream in the context of information transfer along the cortical processing stream. We have shown that responses in AI contain more information about species-specific vocalizations when late responses in PBr are known. In contrast, early responses in PBr contain more information about stimuli when responses in AI are known. These results elucidate the time scale and directionality of processing along the ventral auditory processing stream.

3.6 References

[1] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavior and Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[2] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex,* vol. 1 (1), pp. 1-47, 1991.

[3] T. A. Hackett, "Information flow in the auditory cortical network," *Hearing Research,* vol. 271 (1-2), pp. 133-46, 2011.

[4] J. A. Winer and C. C. Lee, "The distributed auditory cortex," *Hearing Research,* vol. 229 (1-2), pp. 3-13, 2007.

[5] H. E. Hefner and R. S. Heffner, "Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by japanese macaques," *Journal of Neurophysiology,* vol. 56 (3), pp. 683-701, 1986.

[6] A. U. Turken and N. F. Dronkers, "The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses," *Frontiers in Systems Neuroscience,* vol. 5 pp. 1-20, 2011.

[7] T. M. Woods, S. E. Lopez, J. H. Long, J. E. Rahman, and G. H. Recanzone, "Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey," *Journal of Neurophysiology,* vol. 96 (6), pp. 3323-3337, 2006.

[8] X. Wang, "Neural coding strategies in auditory cortex," *Hearing Research,* vol. 229 (1-2), pp. 81-93, 2007.

[9] J. W. H. Schnupp, T. M. Hall, R. F. Kokelaar, and B. Ahmed, "Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex," *Journal of Neuroscience,* vol. 26 (18), pp. 4785-95, 2006.

[10] L. M. Romanski, B. B. Averbeck, and M. Diltz, "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex," *Journal of Neurophysiology,* vol. 93 (2), pp. 734-47, 2005.

[11] A. Adhikari, T. Sigurdsson, M. A. Topiwala, and J. A. Gordon, "Cross-correlation of instantaneous amplitudes of field potential oscillations: A straightforward method to estimate the directionality and lag between brain areas," *Journal of Neuroscience Methods,* vol. 191 (2), pp. 191-200, 2010.

[12] A. K. Seth, "A matlab toolbox for granger causal connectivity analysis," *Journal of Neuroscience Methods,* vol. 186 (2), pp. 262-273, 2010.

[13] K. So, A. C. Koralek, K. Ganguly, M. C. Gastpar, and J. M. Carmena, "Assessing functional connectivity of neural ensembles using directed information," *Journal of Neural Engineering,* vol. 9 (2), p. 026004, 2012.

[14] R. Salvador, M. Anguera, J. J. Gomar, E. T. Bullmore, and E. Pomarol-Clotet, "Conditional mutual information maps as descriptors of net connectivity levels in the brain," *Frontiers in Neuroinformatics,* vol. 4 p. 115, 2010.

[15] S. Shoham, M. R. Fellows, and R. A. Normann, "Robust, automatic spike sorting using mixtures of multivariate t-distributions," *Journal of Neuroscience Methods,* vol. 127 (2), pp. 111-22, 2003.

[16] H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, and P. P. Mitra, "Chronux: A platform for analyzing neural signals," *Journal of Neuroscience Methods,* vol. 192 (1), pp. 146-51, 2010.

[17] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering,* vol. 7 (5), p. 056007, 2010.

[18] B. H. Scott, B. J. Malone, and M. N. Semple, "Transformation of temporal processing across auditory cortex of awake macaques," *Journal of Neurophysiology,* pp. 1-75, 2010.

CHAPTER 4

AUDITORY PERCEPTIONS CORRELATE WITH THE NEURAL REPRESENTATIONS OF VISUAL STIMULI MORE THAN AUDITORY STIMULI IN AUDITORY CORTEX DURING THE MCGURK ILLUSION

4.1 Summary paragraph

Most interpersonal communication occurs in contexts where the listener can see as well as hear the speaker. Furthermore, it has been shown that visual stimuli can subtly change a listener's auditory perception [1]. A salient example of this phenomenon is the McGurk illusion [2]. In this illusion a video of someone speaking a phoneme is played concurrently with audio of a different phoneme. In some cases this results in the illusory auditory perception of the phoneme articulated by the mouth in the video stimulus, and not the actual phoneme present in the auditory stimulus. While many studies have shown visual influences in auditory cortex [3, 4], studies have yet to link visual influences on the neural representation of language with subjective language perception. Here, we show that vision directly influences the electrophysiological representation of phonemes in human auditory cortex. Using the McGurk effect to dissociate the subjective perception of phonemes from the auditory stimuli, we demonstrate that neural representations in auditory cortex are closer to the illusory subjective perception, i.e. the visual stimulus of

mouth articulation, than the auditory stimulus. Conditional mutual information analysis showed that information about visual and auditory stimuli transfered in the caudal-rostral direction along the superior temporal gyrus during phoneme perception. These results show that visual stimuli can influence the neural representation of phonemes in auditory cortex, and in some cases override the subjective perceptions of an auditory stimulus.

4.2 Introduction

The McGurk effect is an auditory illusion that occurs when the perception of a phoneme's auditory identity is changed by a concurrently played video of a mouth articulating a different phoneme [2]. Most subjects will report hearing the phoneme articulated by the mouth in the video, and not the actual phoneme present in the auditory stimulus [5]. The concurrent visual stimulus is presumably altering the neural representation, and therefore subjective perception, of the auditory stimulus. Understanding how, and where, neural representations are changed and perceptual identity is altered will provide important insight into the neural mechanisms of speech perception.

The perceptual identity of a sound is thought to be processed hierarchically in the human brain along the superior temporal lobe in a cortical processing stream analogous to the ventral visual processing stream in the inferior temporal lobe [6-9]. Studies of the neural representation of language have therefore focused on the neural construction of phonemic identity in the superior temporal lobe. Electrical recordings from the surface of the human brain have determined that local field potentials (LFP)

correlated with subjective phoneme categorization [10], and showed topographic coding of specific speech sounds in the superior temporal gyrus (STG) [11, 12].

Where in the brain, and to what extent, vision influences auditory perception is not well understood. Visual enhancement and suppression of auditory responses have been observed at the level of primary auditory cortex in macaques [13]. Electrophysiological recordings through the medial to lateral extent of the human temporal lobe have determined that vision influences audition early in time and visual influence extends to hierarchically lower cortical areas [4, 14, 15]. Magnetoencephalography and electroencephalography showed that auditory representations in the superior temporal lobe were altered by visual influences and argue that visual influences play a predictive role in determining language identity [16, 17]. These visual influences on auditory cortex, however, have not been linked to subjective perception. Understanding how visual influences alter auditory perception during the McGurk illusion, a potent example of vision's effect on auditory perception, will provide insight into the neural mechanisms of quotidian speech perception.

To explore this issue, we examined electrical activity recorded from subdural electrodes in four human patients (two right hemispheres and three left hemispheres) with pharmacologically intractable epilepsy who were undergoing monitoring for seizure activity. Using the McGurk effect we were able to dissociate the perceptual identity of an auditory stimulus from the auditory stimulus provided to the ear. Subjects performed an audio-visual speech perception task in which a video stimulus of a mouth articulating one of four syllables, ("BA", "GA", "VA", and "THA") was paired with an audio stimulus of one of the same four syllables (/BA/, /GA/, /VA/, and /THA/) (Figure 4.1(a)).

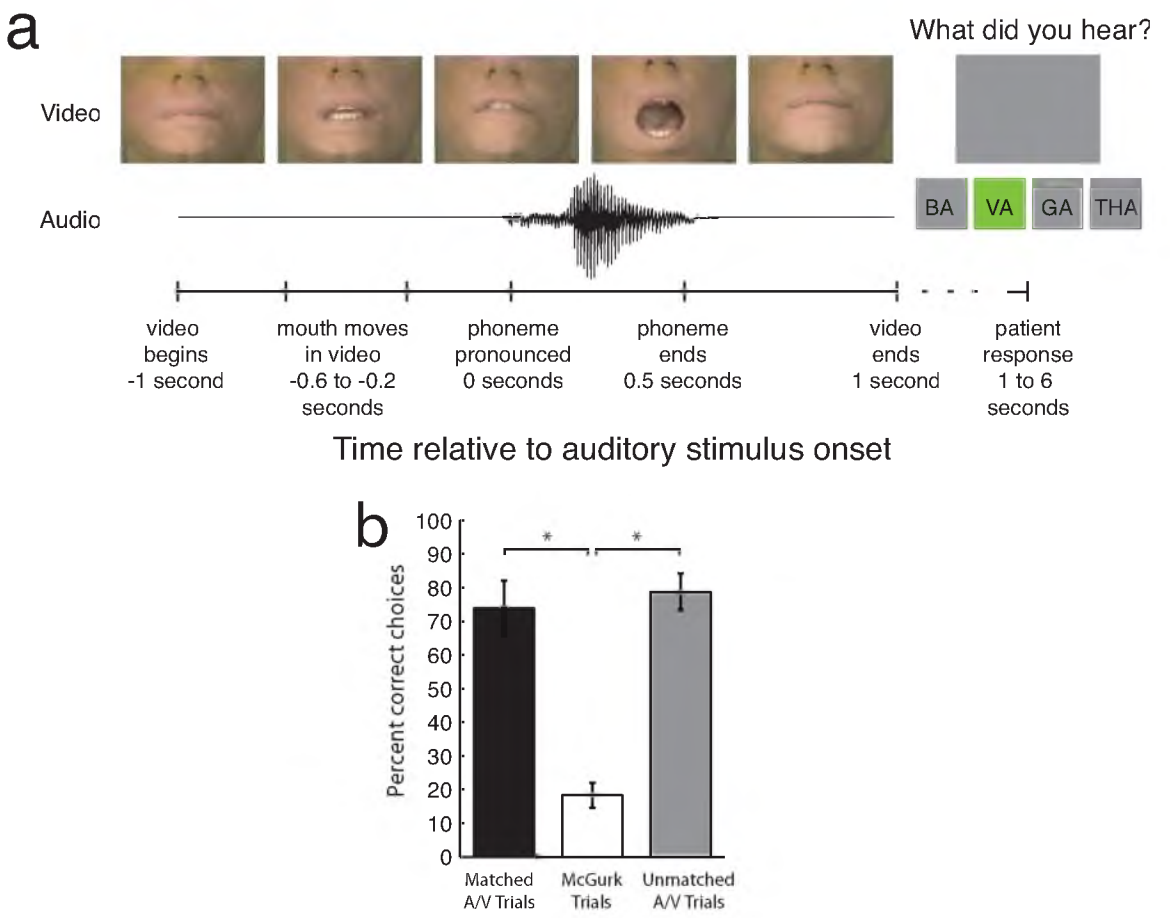


Figure 4.1 Task description and performance. (a) The task consisted of a video of a mouth pronouncing one of four phonemes, this video was randomly paired with audio of a male pronouncing one of the same four syllables. The video times here are shown in the text below the timeline. There was 1 second of video before the audio began, during which the mouth moved slightly in order to position to speak the starting phoneme. The audio syllable lasted half a second, and there was 1 second of video after the audio had finished. After a brief randomized delay the subject was cued to respond. The patient had 5 seconds to respond before a new trial was initiated. (b) Task performance for three conditions. Patients performed significantly better on "Matched A/V" (73.81%, N = 186) trials and "Unmatched" (78.77%, N = 299) trials when compared with "McGurk" (18.29%, N = 152) trials (ANOVA, Tukey-Kramer method for multiple comparisons, $p<0.01$ for both comparisons)

4.3 Methods

These experiments were approved by University of Utah Institutional Review Board.

4.3.1 Subjects

Four human patients were implanted with electrocorticography (ECoG) electrodes for clinical monitoring of epilepsy. Two patients were implanted with frontotemporal ECoG grids with 64 electrodes in the left hemisphere (L; patients 1 and 2), one patient was implanted with a frontotemporal grid with 48 electrodes in the right hemisphere (R; patient 3), and one patient was implanted with strips of ECoG electrodes in both the left and right hemispheres (L and R; patient 4; see Figure 4.2(a)).

4.3.2 Task design and behavioral testing

These patients performed a multisensory speech perception task in which syllables were delivered binaurally from flat frequency response, closed-back, headphones concurrently with videos of a mouth articulating syllables shown on a monitor. We use the term syllable to describe phoneme pairs. Four audio (/BA/, /GA/, /VA/, and /THA/) and four video ("BA", "GA", "VA", and "THA") syllables pairs were randomly paired, creating 16 stimulus combinations. Audio syllables were all from the same male speaker and were paired with commonly used McGurk stimulus videos [18].

Stimulus combinations were grouped into three categories based on the patients' behavioral responses: "Matched A/V" trials were those in which the video and audio had the same phonemic identity. "McGurk" trials were those in which the patient reported

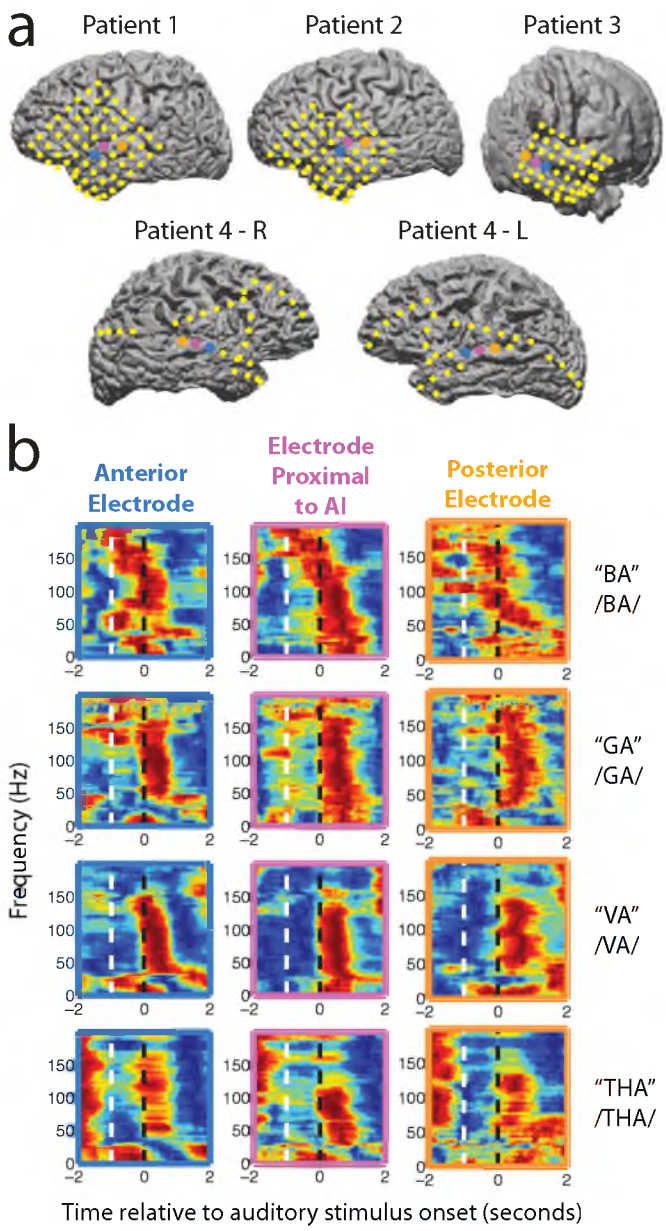


Figure 4.2 Electrode locations and responses to auditory stimuli on STG electrodes. (a) Electrode locations for five hemispheres in four patients. Yellow dots indicate electrode placements we did not use for analyses. Electrodes used for analysis have been color coded. Blue represents the anterior electrode, pink represents the electrode proximal to AI, and orange represents the posterior electrode. (b) Representative neural responses for each phoneme on three STG electrodes. White dotted lines indicate the start of the video. Black dotted lines indicate the start of the audio. Responses are outlined to match their respective electrode locations.

hearing the identity of the video more often than chance, and reported hearing the identity of the audio less often than chance. "Unmatched A/V" trials were those in which the video and audio had different phonemic identities and the patient reported hearing the audio identity more often than chance and the video identity less often than chance.

4.3.3 Data collection and preprocessing

Neural data were collected using a Cerebus system (Blackrock Microsystems). Electrophysiological signals were pseudodifferentially amplified at a gain of 5000x and sampled at 10 kilosamples per second for patients 1, 3 and 4. Data for patient 2 was sampled at 1 kilosamples per second. All data were filtered at 500 Hz, and downsampled to 1 kHz for further analysis. Data for each electrode was then referenced against all other ECoG electrodes in the same hemisphere for each patient by subtracting the mean across electrodes for each trial. This referencing procedure acts as a large, low-impedance monopolar reference.

Multitapered spectral analysis was used to generate spectrograms [19]. A 500 millisecond-long moving window and step size, with seven and eleven leading tapers were used to generate spectrograms for trial averaged spectral analysis. Averaged spectrograms were subtracted and the mean of the absolute value of the resulting difference spectrogram is quantified in Figure 4.3(b).

4.3.4 Electrode selection

For each patient, three electrodes in the superior temporal gyrus were used for analysis. These were the electrodes with the greatest spectral power between 75 and 200

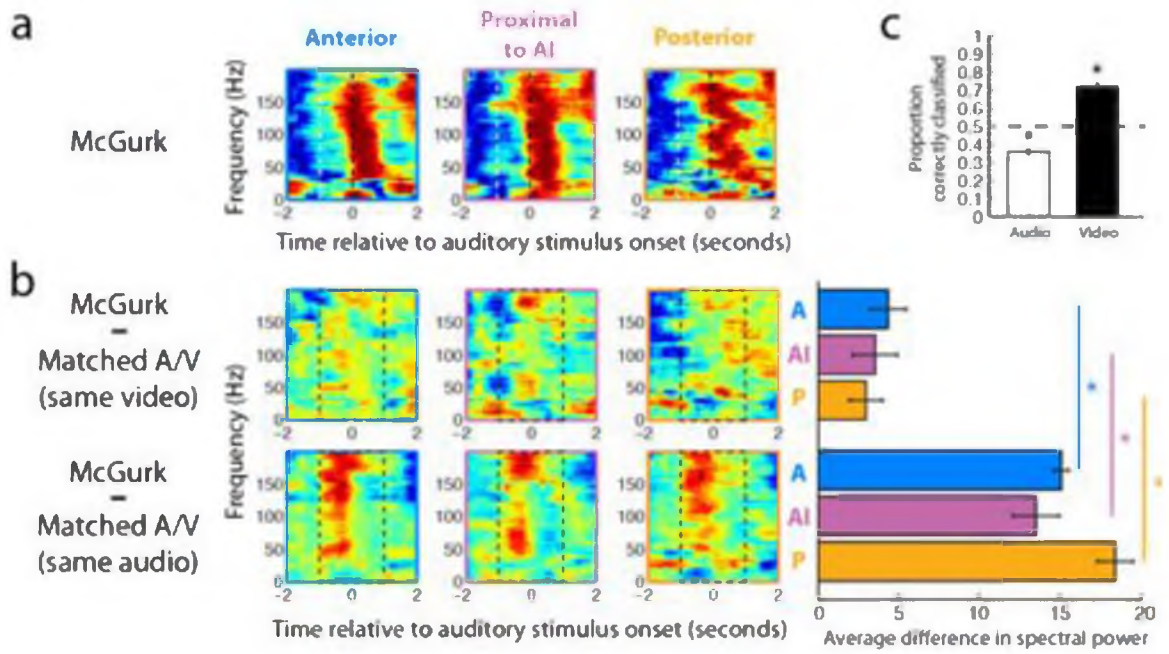


Figure 4.3 Visual representations in parabelt auditory cortex. (a) Example spectrograms for the "McGurk" condition ("VA"&/BA/). Spectrograms were normalized by frequency band. White dotted lines indicate the start of the video. Black dotted lines indicate the start of the audio**.** (b) Example difference spectrograms for all three electrodes from one patient. "Matched A/V" spectrograms were subtracted from "McGurk" spectrograms ("VA"&/BA/ - "VA"&/VA/ and "VA"&/BA/- "BA"&/BA/) between -1 and 1 seconds relative to auditory stimulus onset (black dashed box in spectrograms). "McGurk" spectrograms were significantly less different from "Matched A/V" spectrograms with the same video identity than "Matched A/V" spectrograms with the same audio identity, as shown in the bar graph to the left of the difference spectrograms. Electrode locations are color coded and labeled (A, anterior electrode; AI, electrode proximal to AI; P, posterior electrode). (c) A statistical classifier accurately classified "McGurk" trials when tested on the identity of the video (74.33%), however the classifier consistently chose the wrong auditory identity for "McGurk" trials (36.17%). The dashed line represents chance level classification (50.00 %).

Hz during auditory stimulus presentation, averaged over all stimuli. Electrodes were localized using custom software [20] based on Statistical Parametric Mapping 8 (functional neuroimaging group, University College London). After coregistering and reslicing anatomical preoperative MRIs and postoperative computed tomography (CT) images, electrodes from the CTs are projected onto cortical surfaces generated from the MRIs. Patient 3's MRIs were taken over a year before the CTs and parts of his brain were removed before the CTs were taken, so the cortical rendering is rougher than the other three patients. The superior temporal and transverse temporal gyri, were still visible in patient 3's rendered cortex (Figure 4.2(b)).

4.3.5 Classification of phoneme identity

Single-trial spectrograms and voltage traces ranging in time from the onset of the video to the end of the phoneme were used as neural features in the statistical classifier [21], [22]. These multidimensional data were unwrapped to produce a two-dimensional matrix in which each row contained all the voltage, time, and frequency features from all channels for a single trial. The feature matrix was z-scored, orthogonalized using PCA, and projected into the principal component space using a sufficient number of leading principal components to retain 95% of the variance in the data. Based on these neural features, data were then classified on a trial-by-trial basis using linear discriminant analysis (LDA) [23]. The classifier was trained on half the trials and tested on the other half. Training and testing sets were then interchanged to obtain two fold cross validation. Classification accuracy was measured against the level of chance, which was 50% for all

classifications. The one sample Wilcoxon signed rank test was used to determine the level of significance for classification results.

4.3.6 Information transfer measures

Conditional mutual information measures were calculated using probability distributions derived from the pair wise classification frequencies generated from all statistical classifications [21]. The equation

$$I(S;R_x|R_y) = \sum_{s\in S}\sum_{x\in R_x}\sum_{y\in R_y} p(s,x,y)\log_2\left(\frac{p(y)\,p(s,x,y)}{p(s,y)\,p(x,y)}\right) \quad (1)$$

was evaluated for these probability distributions, where R_x and R_y are the response identities on each of two electrodes, and $p(x)$ and $p(y)$ are the corresponding probability distributions generated from pair wise classification frequencies. We define information tendency as the net information transfer between adjacent electrodes. Information tendency is therefore the difference between the two opposing conditional mutual information measures for adjacent electrodes (e.g. $I(S;R_x|R_y) - I(S;R_y|R_x)$ indicates information transfer from electrode Y to electrode X).

4.4 Experiments and results

Video and audio stimuli were randomly paired, creating sixteen possible stimulus combinations. After each audiovisual stimulus had been delivered, four buttons in the task control software appeared, cueing the subject to indicate which phoneme she had heard. We grouped these trials into three categories. The first category, "Matched A/V" trials (N = 186), were those trials in which the audio and video stimuli had the same

phonemic identity. Trials in which the audio and video stimuli did not match were grouped into two categories: "McGurk" trials (N = 152) were those in which the video and audio stimuli did not match and the patient reported hearing the phonemic identity of the video stimulus and not the phonemic identity of the audio stimulus, and "Unmatched A/V" trials (N = 299) were those in which the video and audio stimuli did not match, and did not produce a McGurk illusion. Patients performed significantly better at identifying the audio stimuli on "Matched A/V" (73.81%) trials and "Unmatched A/V" (78.77%) trials compared with "McGurk" (18.29%) trials (ANOVA, Tukey-Kramer method for multiple comparisons, $p<0.01$) (Figure 4.1(b)).

We restricted our analyses of neural signals to three electrodes per patient. These were the electrodes with the greatest spectral power in the 75-200 Hz range during the 1000 ms when the auditory phoneme is pronounced. Further analysis of the spatial location of these electrodes, based on preoperative magnetic resonance images (MRI) and postoperative computed tomography (CT) images [20], indicated that all three electrodes were on STG in Brodmann's areas 41 and 42, or parabelt auditory cortex (Figure 4.2(a)). Figure 4.2(b) shows example responses averaged over "Matched A/V" trials for one patient. For all patients we observed bursts in spectral power coincident with the presentation of auditory stimuli on all three of the electrodes that were used for analysis.

The McGurk effect is robust enough to be perceived even if the viewer knows the illusion is occurring, suggesting that visual stimuli can influence early representations of auditory stimuli. We wanted to examine how far visual influences extend into the auditory cortex, and to what extent these visual influences alter neural representation in the auditory cortex. We began by examining the similarity of neural representations

during "McGurk" and "Matched A/V" condition stimuli. Spectrograms from "McGurk" condition trials were similar to spectrograms from "Matched A/V" condition when they had the same video stimuli, even though the auditory stimuli were different. Conversely, spectrograms from "McGurk" condition trials were dissimilar to spectrograms from "Matched A/V" condition when they had the different video stimuli, even though the auditory stimuli were same. This trend was evident in data recorded from all three parabelt electrodes. To quantify this observation, the mean differences between neural responses on each electrode to the "McGurk" condition and the "Matched A/V" condition were compared for all four patients. For each electrode, spectral differences between "McGurk" spectrograms and "Matched A/V" spectrograms which had the same audio stimuli different video stimuli were significantly greater than the spectral differences between "McGurk" spectrograms and "Matched A/V" spectrograms which had different audio stimuli and the same video stimuli ($p<0.01$, Wilcoxon signed rank test) (Figure 4.3(b)). This result indicates that in parabelt auditory cortex, neural representations of the McGurk illusion corresponded to the video stimuli more than the audio stimuli, i.e., the neural representation of the visual stimulus was closer to the patients' illusory auditory perceptions.

To further probe the electrophysiological representation of audiovisual language stimuli in parabelt on a trial-by-trial basis, a statistical classifier was trained on LFP waveforms and spectrograms from all three electrodes for half of the "McGurk" trials. This classifier was then used to decode the identity of the video or the audio stimuli of each trial from the remaining half of the "McGurk" trials [21, 22]. The training and testing sets were then switched to obtain a two-fold cross-validation. Classifier

performance was significantly better than chance for “McGurk" stimuli when the correct classification was chosen as the video identity (Wilcoxon signed rank test, $p<0.01$). However, the classifier consistently chose the wrong identity when the correct classification was chosen as the sound identity (Wilcoxon signed rank test, $p<0.01$). These results indicate that trial-by-trial neural representations for phoneme stimuli in parabelt auditory cortex encoded the identity of the video stimuli during “McGurk” condition trials.

Visual information about object identity likely flows from caudal to rostral into the auditory cortex along the ventral visual pathway [24], and similarly flows from caudal to rostral onto frontal cortex along the ventral auditory pathway [7, 8]. To examine this idea we used a conditional information theoretic analysis [21, 25] to determine the transfer of information about visual or auditory stimulus identity among the three STG electrodes. This analysis quantifies the average reduction in uncertainty about the identity of a stimulus from observing a single neural response, given that we already know the response identity in another area. Information transfer was examined for three time intervals: 1) a baseline interval in which no audio or video stimulus is present; 2) when the video of the mouth articulation is being shown, yet the audio stimulus has not begun; and 3) when both the audio stimulus of the phoneme being pronounced and the video stimulus of the mouth articulation are being concurrently presented.

Information transfer about the identity of the visual stimuli was increased between the posterior electrode and the electrode proximal to AI during the period when the mouth begins articulating the phoneme, yet no auditory stimulus is present (Figure 4.4(b)). Information about the phonemic identity from the video stimuli was transferred

in the caudal-rostral direction along the STG before the onset of the auditory stimuli, suggesting that the ventral visual "what" pathway is providing input into cortical areas in the superior temporal lobe early on during audiovisual language perception. Information transfer about the identity of both the visual and auditory stimuli was increased between the electrode proximal to AI and the anterior electrode during the period when the visual stimulus of mouth articulation and the auditory stimulus of phoneme pronunciation were concurrently presented (Figure 4.4(c)). The temporal dynamics of the caudal to rostral information transfer along the STG shows early visual information influencing the neural representation of phonemes in auditory cortex, and later visual and auditory information passing into the audition-for-perception processing stream [26].

4.5 Conclusions

Results from electrocorticography in human STG depict language representations in tertiary auditory cortex as being altered by attention or the context of a sound [27], and suggest multimodal influences on the early stages of auditory processing [28]. We demonstrate that auditory representations in the STG are altered by early visual stimuli and these visual influences are predictive in determining the subjective perception of phonemes, i.e., the neural representations of phonemic identity from visual input can extend into auditory cortex and affect the perception of language. The time course and direction of auditory and visual information transfer about the identity of phonemes in parabelt auditory cortex showed that visual information transfers caudorostrally through STG before any auditory stimuli was presented. This suggests a mechanistic

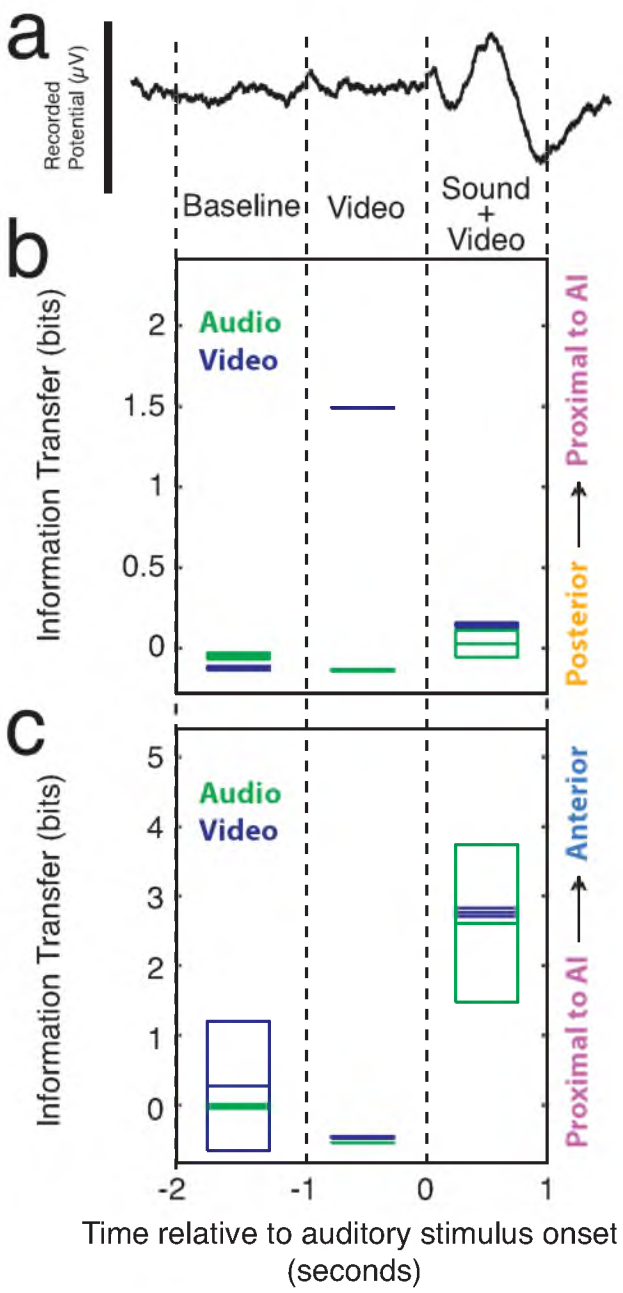


Figure 4.4 Transfer information in the superior temporal lobe. (a) An averaged evoked potential from the middle electrode is shown above plots for reference. The scale bar for the evoked potential is 400 μV. (b) Plot of information transfer between the posterior electrode and the electrode proximal to AI for three second-long time periods through the duration of the trial. (c) Plot of information transfer between the electrode proximal to AI and the anterior electrode for three second-long time periods through the duration of the trial. For both (b) and (c) positive values indicate transfer of information in the caudal-rostral direction and negative values indicate transfer of information in the rostral-caudal direction. Green box plots indicate information about the identity of the audio stimuli. Blue plots indicate information about the video stimuli.

underpinning for the McGurk effect in that the information from the visual cortex may be instructing the auditory cortex which phoneme to "hear" before an auditory stimulus is received. This understanding of multisensory neocortical language processing provides insight into the multisensory neural mechanisms underlying quotidian language perception and has implications for rehabilitation therapy and neural prostheses.

4.6 References

[1] B. E. Stein and M. A. Meredith, *The merging of the senses*: Mit Pr, 1993.

[2] H. McGurk and J. Macdonald, "Hearing lips and seeing voices," *Nature,* vol. 264 (5588), pp. 746-748, 1976.

[3] G. A. Calvert, "Crossmodal processing in the human brain: Insights from functional neuroimaging studies," *Cerebral Cortex (New York, NY : 1991),* vol. 11 (12), pp. 1110-1123, 2001.

[4] J. Besle, C. Fischer, A. Bidet-Caulet, F. Lecaignard, O. Bertrand, and M.-H. Giard, "Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans," *Journal of Neuroscience,* vol. 28 (52), pp. 14301-14310, 2008.

[5] J. Jiang and L. E. Bernstein, "Psychophysics of the mcgurk and other audiovisual speech integration effects," *Journal of Experimental Psychology. Human Perception and Performance,* vol. 37 (4), pp. 1193-1209, 2011.

[6] T. A. Hackett, "Information flow in the auditory cortical network," *Hearing Research,* vol. 271 (1-2), pp. 133-46, 2011.

[7] J. P. Rauschecker and S. K. Scott, "Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing," *Nature Neurosci,* vol. 12 (6), pp. 718-724, 2009.

[8] L. M. Romanski, "Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex.," *Cerebral Cortex (New York, NY : 1991),* vol. 17 Suppl 1 pp. i61-9, 2007.

[9] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[10] E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight, "Categorical speech representation in human superior temporal gyrus," *Nature Neuroscience,* vol. 13 (11), pp. 1428-1432, 2010.

[11] T. Blakely, K. Miller, R. Rao, M. Holmes, J. Ojemann, "Localization and classification of phonemes using high spatial resolution electrocorticography (ecog) grids," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference,* pp. 1-4, 2008.

[12] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biology,* vol. 10 (1), p. e1001251, 2012.

[13] A. A. Ghazanfar, C. Chandrasekaran, and N. K. Logothetis, "Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys," *Journal of Neuroscience,* vol. 28 (17), pp. 4457-4469, 2008.

[14] J. Besle, O. Bertrand, and M.-H. Giard, "Electrophysiological (eeg, seeg, meg) evidence for multiple audiovisual interactions in the human auditory cortex," *Hearing Research,* vol. 258 (1-2), pp. 143-151, 2009.

[15] D. Senkowski, T. R. Schneider, J. J. Foxe, and A. K. Engel, "Crossmodal binding through neural coherence: Implications for multisensory processing," *Trends in Neurosciences,* vol. 31 (8), pp. 401-409, 2008.

[16] M. Sams, R. Aulanko, M. Hämäläinen, R. Hari, O. V. Lounasmaa, S. T. Lu, and J. Simola, "Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex," *Neuroscience Letters,* vol. 127 (1), pp. 141-145, 1991.

[17] V. van Wassenhove, K. W. Grant, and D. Poeppel, "Visual speech speeds up the neural processing of auditory speech," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102 (4), pp. 1181-1186, 2005.

[18] L. D. Rosenblum, M. A. Schmuckler, and J. A. Johnson, "The mcgurk effect in infants," *Perception & Psychophysics,* vol. 59 (3), pp. 347-357, 1997.

[19] H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, and P. P. Mitra, "Chronux: A platform for analyzing neural signals," *Journal of Neuroscience Methods,* vol. 192 (1), pp. 146-51, 2010.

[20] D. Hermes, K. J. Miller, H. J. Noordmans, M. J. Vansteensel, and N. F. Ramsey, "Automated electrocorticographic electrode localization on individually rendered brain surfaces," *Journal of Neuroscience Methods,* vol. 185 (2), pp. 293-298, 2010.

[21] E. H. Smith, S. S. Kellis, P. House, and B. Greger, "Information transfer along the ventral auditory stream in the awake macaque," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference,* vol. 1 2012.

[22] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering,* vol. 7 (5), p. 056007, 2010.

[23] W. J. Krzanowski, "Principles of multivariate analysis," vol. 3 p. xxii+563, 1990.

[24] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys," *Behavioural Brain Research,* vol. 6 (1), pp. 57-77, 1982.

[25] X. Li and G. Ouyang, "Estimating coupling direction between neuronal populations with permutation conditional mutual information," *NeuroImage,* vol. 52 (2), pp. 497-507, 2010.

[26] L. M. Romanski, B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic, and J. P. Rauschecker, "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex," *Nature Neurosci,* vol. 2 (12), pp. 1131-6, 1999.

[27] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature,* vol. 485 (7397), pp. 233-236, 2012.

[28] A. A. Ghazanfar and C. E. Schroeder, "Is neocortex essentially multisensory?" *Trends in Cognitive Sciences,* vol. 10 (6), pp. 278-285, 2006.

CHAPTER 5

SUMMARY AND PERSPECTIVES

5.1 Summary

This thesis describes my contribution to the study of the ventral auditory stream in macaque and human primates. This work includes the first implant of an electrode array approved for human use by the FDA into the lateral fissure of a macaque, including the first published electrode array recordings from macaque AI. These recordings were carried out simultaneously with the first published electrophysiological recordings from PBr in the macaque. Both LFP and action potential responses were examined in each area. These studies extend beyond the study of trial-averaged responses, developing techniques to examine how information about single stimulus presentations is transferred through the ventral stream, and how that information is changed in the process. Transfer and transformation of information were examined in both macaque and human ventral streams and have been correlated with illusory and subjective perception in humans. This work is important for understanding the neural basis for auditory object perception and has application to brain machine interfaces to create hearing in the deaf or allow locked-in patients to communicate with the external world.

5.2 On the possibility for an auditory cortical neural prosthesis

The cochlear implant (CI) is the most widely used neural prosthesis, providing hundreds of thousands of deaf people with the ability to hear and understand speech [1]. Yet for many CI users there remains some dissatisfaction due to poor frequency resolution and poor hearing in noisy environments [2]. Many of the cues used for speech comprehension are relatively long temporal cues that enable users to discriminate phonemes [3, 4]. Hearing other sounds without the same broad temporal organization as speech, or sounds that require information about pitch or timbre to imbue their meaning, like music, can cause discomfort in many CI users. An important issue in auditory neuroprosthetics, therefore, is improving the capability for encoding specific, and varied, frequencies with a prosthesis.

It has become apparent that there may be a biophysical limitation to stimulating the cochlea, which limits frequency resolution. Despite the 23 stimulating electrodes used in the latest cochlear prostheses, these electrodes do not reach the upper 1.5 turns of the cochlea, and there is a functional upper limit of between 10 and 12 independent electrodes [5]. This upper limit on electrode independence likely arises from the biophysical limitations involved with stimulating the nervous system via a well of conductive fluid: the cochlear perilymph. Attempts at current steering using bipolar or tripolar electrode configurations have been unable to overcome this limitation [6]. These electrode configurations often result in increased thresholds with little change in electrode independence [7, 8]. The lateral spread of current through the cochlear perilymph during cochlear prosthesis use therefore limits electrode independence. Lack

of independence limits the use of the cochlear tonotopy and thus the ability to impart adequate frequency resolution to CI users.

Stimulating the cortex directly overcomes many of these issues, allowing for many more electrodes to be used to stimulate, with lower currents and greater electrode independence. Chapter 2 determines a lower limit on electrode density for interfacing with primary auditory cortex, which allows for the use of hundreds of electrodes in stimulating the cortex. The estimate put forth in Chapter 2 is between 112 and 768 electrodes, depending on electrode density, which could be placed along the human core auditory cortex [9]. At the lower limit, this allows for four times as many channels of input than is possible in a CI as being possible in a cortical prosthesis.

While the results in Chapter 2 depict AI as being a more suitable location for a cortical prosthesis, Chapter 4 shows an intimate relationship between neural representations in the STG and audiovisual language perception. The STG's ease of access and the high correlation between parabelt responses and subjective perception could make this area a good candidate for an auditory cortical neural prosthesis. An STG prosthesis would dramatically reduce the surgical risk involved with accessing the core auditory cortex. Previous studies of macroelectrode stimulation in this area have shown complex and holistic auditory perceptions [10, 11]. Whether the STG can represent the lower level sounds that could be encoded with the types of speech processors used in cochlear implants remains to be seen. Recording with smaller electrodes, microwires or even penetrating electrodes and using simpler stimuli in order to characterize encoding in this area, while still maintaining the behavioral relevance of simple stimuli through the

use of a task, would be a key step forward in determining the possibility for a cortical neural prosthesis.

5.3 Parabelt auditory cortical physiology

Reflecting on the results of the thesis as a whole, it is apparent that many of the results speak to the little-known physiology of parabelt auditory cortex. Chapter 2 presents the first published penetrating recordings in primate PBr, and explores coding in response to natural vocal communication sounds in PBr. While responses in AI had high firing rates and temporal structure that matched the vocalizations' structure, responses in parabelt had lower overall firing rates and seemed to respond only to certain elements of monkey vocalizations (see Figures 2.2 and 2.4 for examples). The low firing rates could be due to sparse or nonlinear coding in parabelt. Human parabelt has been shown to encode language in a nonlinear manner [12]. Yet representations of human language sounds in parabelt are not sparse, but represent a few language sounds in a broad, distributed fashion [13, 14]. How then do we reconcile responses in PBr that show signs of sparse coding with results from human auditory cortex that show broad, distributed coding of a few, specific language sounds?

One possibility involves the behavioral relevance of the stimuli. It has been shown that task performance significantly modulates neural activity in primary auditory cortex [15, 16]. This modulation is likely greater in tertiary auditory cortex, which is closer to association cortex in organization, structure, and proximity. Greater task-related modulation of neural responses has been observed in hierarchically higher auditory areas in the ferret [17] and in fMRI responses in human STG [18]. The natural sounds we used

in our experiments were macaque vocalizations that relate to food or social cues [19]. These stimuli therefore have poignant evolutionary significance, which theoretically predisposes the vocalizations to being behaviorally relevant for the monkey. However, these sounds were delivered while the monkey was passively listening, and not engaging in a task that would make the sounds important for the monkey to obtain a liquid reward. The stimuli's apparent lack of relevance for the task at hand may have contributed to low firing rates and poor decode performance.

Another possibility for observing sparsity in the parabelt data is the relatively anterior recording location. The recordings from Chapter 2 were the first published penetrating recordings from PBr auditory cortex in the primate. Therefore there are few results available for comparison with those presented here. The few other studies that have examined parabelt physiology in the macaque do so from the standpoint of fMRI or examining behavioral deficits after lesions [20-22]. Parabelt auditory cortex in the primate is coarsely divided into a rostral and caudal aspect. The microelectrode array implanted in PBr (Chapters 2 and 3) was implanted towards the rostral aspect of PBr. One other study has examined auditory responses to monkey calls and environmental stimuli in areas rostral to core and belt auditory cortex on the superior temporal plane [23]. This study recorded in AI and in two areas in the rostral superior temporal plane. As the recording location moved further rostral in this study, neurons became more selective for specific monkey calls and environmental sounds, and became increasingly suppressed by all stimuli, and decreasingly excited by all stimuli. These results corroborate the hypothesis that sparsity of auditory representations increases rostrally in the superior temporal lobe.

The sparsity observed in our parabelt recordings could also be a product of the recording technology we used. It is possible that the electrode arrays we used to record are damaging the cortex we implanted them into. While we show in Chapter 2 that there were LFP evoked responses to vocalizations on 95 of 96 electrodes in PBr, the electrodes on the microelectrode array may not have been suitable for recording single neurons. The length of electrodes may not have been suitable to access the appropriate lamina in PBr. It is also a possibility that the fixed geometry of the electrode arrays did not give us preferential access to the neurons in PBr that responded to monkey vocalizations. Single-electrode recordings would allow recordings from specific classes of neurons that represented the stimuli at hand.

There have been many more intracranial recordings in human parabelt auditory cortex than in macaque auditory cortex. These are recordings from epilepsy patients, which means that the clinical monitoring requirements predetermine electrode placement. Strategically placed, high-density, grids of electrodes have therefore rarely been placed over the rostral STG. Auditory recordings from STG have mostly been located near AI or in caudal parabelt. Studies which use only auditory stimuli show increased activation in the caudal STG or PBc [12, 14, 24], whereas another study which looked at simple audiovisual stimuli [25] and Chapter 4 of this thesis show robust responses in PBr. Audiovisual integration, or some sort of multisensory auditory object reinforcement may be required to activate areas in the rostral STG, as is indicated in Figure 4.4.

More strategic experiments utilizing recordings from auditory STG are required to further understand how this area of the human brain is processing auditory communication stimuli. These experiments are rare and difficult to carry out, mainly due

to the unique patient population. Furthermore, the results from Mesgarani and Chang [24] and those presented in Chapter 4 of this thesis are counterintuitive to straightforward auditory processing in the STG. Instead these results indicate that auditory processing in the STG is dramatically altered by the stimulus and task context and is closely linked to perception.

Another aspect of parabelt neurophysiology addressed in this thesis is parabelt's window of temporal integration, that is the minimum time period over which two distinct acoustic events can be resolved. By examining different temporal resolutions in the parabelt decode, we were able to estimate the best temporal resolution for auditory processing in PBr. The 50 millisecond and 25 millisecond temporal resolutions deviated the furthest from chance in the 7 class decode (Figure 2.6), indicating that that these are the best temporal resolutions for our natural stimuli. The information theoretic analysis presented in Chapter 4 adds to this estimate, showing that information transfer between AI and PBr is highest for the 50 millisecond window. This estimate constitutes the first published estimate of the window of temporal integration in parabelt auditory cortex.

Scott et al. [26] estimated the windows of temporal integration for macaque core and belt auditory cortex using sinusoidally amplitude modulated tones. This study estimated a window of 20-30 milliseconds in AI. In Chapters 2 and 3 we estimate a window for AI that is closer to that in the marmoset, at below 10 Hz [27]. Belt neurons in Scott et al. [26] showed more selective windows of temporal integration that centered around 200 milliseconds. Our estimate of temporal integration for neurons in PBr is between 25 and 50 milliseconds, which is wider, yet shorter than that found in belt, and is narrower, yet longer than that found in AI. These estimates are shorter than estimates of

temporal integration in the rostral field of core auditory cortex, and therefore support the hypothesis that processing of temporal elements of auditory stimuli occurs caudorostrally on the superior temporal plane. Future work in this area should utilize monkey calls, band-pass noise, and amplitude modulated stimuli to determine how differences in stimuli affect measurements of temporal integration.

5.4 Conclusions

This thesis explores the neural mechanisms of auditory perception from several different, yet complementary angles. Coding of primate vocal communication in primary and tertiary auditory cortex is examined in the context of the possibility for creating a stimulating neural prosthesis for the deaf. An innovative information theoretic approach to understanding transfer and transformation of information in the cerebral cortex is used to quantify directionality of neural information of vocal communication sounds between primary and tertiary auditory cortex. Finally, visual influences on auditory cortical processing are described during illusory auditory perception. Electrophysiological recordings directly over human auditory cortex are utilized to explain the extent to which the neural mechanisms underlying auditory subjective perception of language can be affected by vision. Together these studies provide insight into the neural mechanisms of sound perception and the possibility for interfacing with the auditory cortex in order to restore hearing in deaf individuals.

5.5 References

[1] N. C. f. H. Statistics, "Vital and health statistics series 10, no. 235 (12/2007)," pp. 1-163, 2007.

[2] K. Oshima, S. Suchert, N. H. Blevins, and S. Heller, "Curing hearing loss: Patient expectations, health care practitioners, and basic science," *Journal of Communication Disorders,* vol. 43 (4), pp. 311-318, 2010.

[3] B. J. Kwon and C. van den Honert, "Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users," *The Journal of the Acoustical Society of America,* vol. 120 (1), pp. EL1-6, 2006.

[4] B. J. Kwon and C. van den Honert, "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," *The Journal of the Acoustical Society of America,* vol. 119 (5 Pt 1), pp. 2994-3002, 2006.

[5] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *The Journal of the Acoustical Society of America,* vol. 110 (2), pp. 1150-1163, 2001.

[6] J. A. Bierer, "Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration," *The Journal of the Acoustical Society of America,* vol. 121 (3), pp. 1642-1653, 2007.

[7] D. J. Morris and B. E. Pfingst, "Effects of electrode configuration and stimulus level on rate and level discrimination with cochlear implants," *JARO - Journal of the Association for Research in Otolaryngology,* vol. 1 (3), pp. 211-223, 2000.

[8] J. A. Bierer and J. C. Middlebrooks, "Cortical responses to cochlear implant stimulation: Channel interactions," *JARO - Journal of the Association for Research in Otolaryngology,* vol. 5 (1), pp. 32-48, 2004.

[9] E. Smith, S. Kellis, P. House, and B. Greger, "Decoding stimulus identity from multi-unit activity and local field potentials along the ventral auditory stream in the awake primate: Implications for cortical neural prostheses," *Journal of Neural Engineering,* vol. 10 (1), p. 016010, 2013.

[10] W. Penfield, "Some mechanisms of consciousness discovered during electrical stimulation of the brain," *PNAS,* pp. 51-66, 1958.

[11] A. Selimbeyoglu and J. Parvizi, "Electrical stimulation of the human brain: Perceptual and behavioral phenomena reported in the old and new literature," *Frontiers in Human Neuroscience,* vol. 4 (1), pp. 1-11, 2010.

[12] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biology,* vol. 10 (1), p. e1001251, 2012.

[13] T. Blakely, K. Miller, R. Rao, M. Holmes, J. Ojemann, "Localization and classification of phonemes using high spatial resolution electrocorticography (ecog) grids," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference,* pp. 1-4, 2008.

[14] E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight, "Categorical speech representation in human superior temporal gyrus," *Nature Neuroscience,* vol. 13 (11), pp. 1428-1432, 2010.

[15] G. H. Otazu, L.-H. Tai, Y. Yang, and A. M. Zador, "Engaging in an auditory task suppresses responses in auditory cortex," *Nature Neuroscience,* vol. 12 (5), pp. 646-654, 2009.

[16] B. H. Scott, B. J. Malone, and M. N. Semple, "Effect of behavioral context on representation of a spatial cue in core auditory cortex of awake macaques," *Journal of Neuroscience,* vol. 27 (24), pp. 6489-6499, 2007.

[17] J. B. Fritz, S. V. David, S. Radtke-Schuller, P. Yin, and S. A. Shamma, "Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex," *Nature Neuroscience,* vol. 13 (8), pp. 1011-1019, 2010.

[18] D. L. Woods, G. C. Stecker, T. Rinne, T. J. Herron, A. D. Cate, E. W. Yund, I. Liao, and X. Kang, "Functional maps of human auditory cortex: Effects of acoustic features and attention," *PloS One,* vol. 4 (4), p. e5183, 2009.

[19] J. P. Rauschecker, B. Tian, and M. Hauser, "Processing of complex sounds in the macaque nonprimary auditory cortex," *Science (New York, NY),* vol. 268 (5207), pp. 111-114, 1995.

[20] C. I. Petkov, C. Kayser, M. Augath, and N. K. Logothetis, "Functional imaging reveals numerous fields in the monkey auditory cortex," *PLoS Biology,* vol. 4 (7), p. e215, 2006.

[21] H. E. Hefner and R. S. Heffner, "Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by japanese macaques," *Journal of Neurophysiology,* vol. 56 (3), pp. 683-701, 1986.

[22] H. E. Heffner and R. S. Heffner, "Temporal lobe lesions and perception of species-specific vocalizations by macaques," *Science (New York, NY),* vol. 226 (4670), pp. 75-76, 1984.

[23] Y. Kikuchi, B. Horwitz, and M. Mishkin, "Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane," *Journal of Neuroscience,* vol. 30 (39), pp. 13021-13030, 2010.

[24] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature,* vol. 485 (7397), pp. 233-236, 2012.

[25] J. Besle, C. A. Schevon, A. D. Mehta, P. Lakatos, R. R. Goodman, G. M. McKhann, R. G. Emerson, and C. E. Schroeder, "Tuning of the human neocortex to the temporal dynamics of attended events," *Journal of Neuroscience,* vol. 31 (9), pp. 3176-3185, 2011.

[26] B. H. Scott, B. J. Malone, and M. N. Semple, "Transformation of temporal processing across auditory cortex of awake macaques," *Journal of Neurophysiology,* vol. 105 (2), pp. 712-730, 2011.

[27] D. Bendor and X. Wang, "Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys," *Journal of Neurophysiology,* vol. 100 (2), pp. 888-906, 2008.