2016

Neural And Behavioral Correlates Of Auditory Representation, Perception, And Categorization In Humans

Adam Martin Gifford
University of Pennsylvania, mada@mail.med.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Neuroscience and Neurobiology Commons

Recommended Citation
https://repository.upenn.edu/edissertations/2304

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2304
For more information, please contact repository@pobox.upenn.edu.
Neural And Behavioral Correlates Of Auditory Representation, Perception, And Categorization In Humans

Abstract
Auditory perceptual representations (i.e., “sounds”) reflect the brain’s ability to group or segregate acoustic features based on detected regularities in the acoustic environment. These perceptual representations provide information on the putative sound sources in the environment, which are combined with knowledge of auditory categories to both imbue sounds with meaning and inform appropriate behavioral actions. Critical aspects regarding the cortical mechanisms responsible for regularity representation and perception as well as how prior knowledge of auditory categories influences perceptual judgments remain unanswered. This dissertation had two main goals: (1) to test how neural activity encodes regularity representation and perception; and (2) to test how a listener uses prior category knowledge to inform categorical judgments when a stimulus’ category membership is ambiguous. To achieve these goals, I employed a combination of neurophysiological, behavioral, and computational analyses in humans. I found that the phase of population-level neural activity is a more reliable indicator of regularity than power and that a variety of brain regions exhibited reliable modulations that distinguished stimulus and behavioral differences related to regularity violation. Additionally, I found that human listeners learn approximations of auditory categories and are varied in their ability to use prior category information to inform categorical judgments. Finally, I found that categorization behavior was consistent with an ideal decision strategy that includes trial-by-trial variability in a listener’s estimates of the prior probability of each category. These findings build upon previous work on the mechanisms underlying regularity processing in auditory perception and that future research should focus on a variety of brain regions beyond the classical auditory pathways in cortex. Additionally, the categorization findings are the first to extend previous work in visual categorization into the auditory domain and reformulates the issue of categorization in a manner that can help to interpret the results of previous research within a generative framework.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Neuroscience

First Advisor
Yale E. Cohen

Keywords
auditory, categorization, electrocorticography, perception, regularity

Subject Categories
Neuroscience and Neurobiology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2304
NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY REPRESENTATION, PERCEPTION, AND CATEGORIZATION IN HUMANS

Adam M. Gifford

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

________________________

Yale E. Cohen

Professor of Otorhinolaryngology: Head and Neck Surgery

Graduate Group Chairperson

________________________

Joshua I. Gold
Professor of Neuroscience

Dissertation Committee:

David H. Brainard, RRL Professor of Psychology
Michael J. Kahana, Professor of Psychology
Daniel D. Lee, Professor of Electrical and Systems Engineering
Danielle S. Bassett, Assistant Professor of Innovation in Bioengineering
Charles E. Schroeder, Professor of Psychology, Columbia University
NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY REPRESENTATION, PERCEPTION, AND CATEGORIZATION IN HUMANS

COPYRIGHT

2016

Adam M. Gifford

This work is licensed under the
Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License

To view a copy of this license, visit

https://creativecommons.org/licenses/by-nc-sa/3.0/us/
I dedicate this dissertation to my family for all of the love, support, and motivation they provided me throughout my life, and in particular during my thesis years. And to my fiancée, Jasmine, because if it were not for my graduate studies, we may have never met.
ACKNOWLEDGMENTS

First, I would like to thank the ~30 patients with intractable epilepsy who agreed to participate in the electrophysiological studies on the neural correlates of auditory perception in humans. Without their participation, none of this work would be possible. Similarly, I would like to thank the clinical teams at both Thomas Jefferson University Hospital and the Hospital of the University of Pennsylvania for the help and collaboration I received that allowed my thesis work to be conducted. And, of course, I have to think the 6 subjects who volunteered to participate in my behavioral study on auditory categorization.

Second, I would like to thank Dr. Michael J. Kahana, who was a co-advisor of sorts, a committee member, and the facilitator of our collaboration with the clinical teams at both hospitals. He was an invaluable resource for the study and analysis of electrocorticography, and a critical contributor to the success of my work. I would also like to thank the Kahana lab in general, including the many lab members who have guided my training in electrocorticographic research and analysis and the numerous lab techs who conducted all of the behind-the-scenes work running the experiments and performing necessary data processing that prepared the raw signals for further analysis.

Third, I would like to thank my other committee members, Drs. David Brainard, Daniel Lee and Danielle Bassett for their support, guidance, and ideas throughout my graduate training. Fourth, I would like to thank all of the faculty and research staff who participated in the systems-neuroscience laboratory meetings that provided valuable feedback and recommendations on my research. Their insight and support greatly strengthened the quality of my thesis work.

Finally, I would like to thank the members of my thesis lab. From participating in pilot studies, to offering critical advice and feedback, to simply making my experience as a graduate student a (relatively) pleasant experience, this dedicated and thoughtful group of scientists played
an invaluable role in my training and development. In particular, I have to thank my advisor, Dr. Yale E. Cohen. Over the past 6+ years, he has been nothing short of the perfect mentor for me, always available to offer valuable guidance and advice, and always supportive of my research and extracurricular endeavors. Without his unconditional support and belief in my abilities, I most certainly would not have accomplished this milestone.
ABSTRACT

NEURAL AND BEHAVIORAL CORRELATES OF AUDITORY REPRESENTATION, PERCEPTION, AND CATEGORIZATION IN HUMANS

Adam M. Gifford

Yale E. Cohen, Ph.D.

Auditory perceptual representations (i.e., “sounds”) reflect the brain’s ability to group or segregate acoustic features based on detected regularities in the acoustic environment. These perceptual representations provide information on the putative sound sources in the environment, which are combined with knowledge of auditory categories to both imbue sounds with meaning and inform appropriate behavioral actions. Critical aspects regarding the cortical mechanisms responsible for regularity representation and perception as well as how prior knowledge of auditory categories influences perceptual judgments remain unanswered. This dissertation had two main goals: (1) to test how neural activity encodes regularity representation and perception; and (2) to test how a listener uses prior category knowledge to inform categorical judgments when a stimulus’ category membership is ambiguous. To achieve these goals, I employed a combination of neurophysiological, behavioral, and computational analyses in humans. I found that the phase of population-level neural activity is a more reliable indicator of regularity than power and that a variety of brain regions exhibited reliable modulations that distinguished stimulus and behavioral differences related to regularity violation. Additionally, I found that human listeners learn approximations of auditory categories and are varied in their ability to use prior category information to inform categorical judgments. Finally, I found that categorization behavior was consistent with an ideal decision strategy that includes trial-by-trial variability in a listener’s estimates of the prior probability of each category. These findings build upon previous work on the mechanisms
underlying regularity processing in auditory perception and that future research should focus on a
variety of brain regions beyond the classical auditory pathways in cortex. Additionally, the
categorization findings are the first to extend previous work in visual categorization into the auditory
domain and reformulates the issue of categorization in a manner that can help to interpret the
results of previous research within a generative framework.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ IV

ABSTRACT .......................................................................................................................... VI

LIST OF TABLES ................................................................................................................... X

LIST OF FIGURES ................................................................................................................ XI

1. GENERAL INTRODUCTION ............................................................................................ 1
   The problem of auditory perception .................................................................................. 2
   Spectrotemporal regularities and prediction in auditory perception .................................. 3
   The ventral pathway for auditory perception ................................................................. 7
   Neural rate-place codes for sound segregation .............................................................. 7
   Complementary role of temporal coherence in sound segregation .................................. 9
   Mechanisms of temporal coherence and neural oscillations ......................................... 10
   Neural oscillations in relation to audition ..................................................................... 11
   Missing gaps in the neural-oscillatory correlates of audition ....................................... 13
   Auditory categorization ................................................................................................. 15
   Missing gaps in understanding auditory categorization ................................................. 17

2. NEURAL-PHASE ALIGNMENT IS A MECHANISM FOR TRACKING DYNAMIC CHANGES IN ACOUSTIC SPECTRAL REGULARITY .................................................... 18
   ABSTRACT ...................................................................................................................... 18
   INTRODUCTION .............................................................................................................. 19
   METHODS .................................................................................................................... 22
   RESULTS ....................................................................................................................... 35
   DISCUSSION .................................................................................................................. 54

3. NEURAL OSCILLATORY CORRELATES OF CONCURRENT SPECTROTEMPORAL-REGULARITY REPRESENTATION AND DEVIANCE DETECTION ........................................... 62
   ABSTRACT ...................................................................................................................... 62
   INTRODUCTION .............................................................................................................. 63
   METHODS .................................................................................................................... 66
   RESULTS ....................................................................................................................... 78
   DISCUSSION .................................................................................................................. 108
4. CHARACTERIZING THE IMPACT OF CATEGORY UNCERTAINTY ON HUMAN AUDITORY CATEGORIZATION BEHAVIOR .......................................................... 114
   ABSTRACT ........................................................................................................ 114
   AUTHOR SUMMARY .......................................................................................... 115
   INTRODUCTION ................................................................................................. 116
   METHODS .......................................................................................................... 118
   RESULTS ............................................................................................................ 126
   DISCUSSION ....................................................................................................... 142

5. GENERAL DISCUSSION ..................................................................................... 145
   Summary of the oscillatory correlates of regularity representation .................... 145
   Strengths of the current investigation on regularity representation and perception .... 146
   Pitfalls of the current investigation on regularity representation and perception ...... 149
   Future directions on the study of regularity representation and perception ............ 150
   Summary of the computational strategies in auditory categorization .................... 152
   Strengths of the current investigation of auditory categorization ......................... 153
   Pitfalls of the current investigation of auditory categorization ............................... 154
   Future directions in the study of auditory categorization ..................................... 155
   Conclusions ........................................................................................................ 156

REFERENCES ........................................................................................................... 157
LIST OF TABLES

2. NEURAL-PHASE ALIGNMENT IS A MECHANISM FOR TRACKING DYNAMIC CHANGES IN ACOUSTIC SPECTRAL REGULARITY

Table 2.1: Regions of interest .................................................................................................................. 34
Table 2.2: Summary statistics for event-related potentials and high-gamma activity to sequence onsets .......................................................................................................................... 38
Table 2.3: Summary statistics for high-gamma sensitivity to tone-burst frequency .............................. 38
Table 2.4: Summary statistics for power and phase-alignment modulations to temporal regularity ................................................................................................................................. 41
Table 2.5: Summary correlation statistics for pairwise phase consistency in band-specific analyses ............................................................................................................................... 50
Table 2.6: Summary correlation statistics for pairwise phase consistency in band-specific analyses ............................................................................................................................... 51
Table 2.7: Summary correlation statistics for amplitude in band-specific analyses .......................... 52

3. NEURAL OSCILLATORY CORRELATES OF CONCURRENT SPECTROTEMPORAL-REGULARITY REPRESENTATION AND DEVIANCE DETECTION

Table 3.1: Band names and pass bands for band-specific analyses ...................................................... 72
Table 3.2: Regions of interest .................................................................................................................. 75
Table 3.3: Summary statistics for power modulations to the standard sequence ............................... 81
Table 3.4: Summary statistics for phase-alignment modulations to the standard sequence ........... 82
LIST OF FIGURES

2. NEURAL-PHASE ALIGNMENT IS A MEchanism FOR TRACKING DYNAMIC CHANGES IN ACOUSTIC SPECTRAL REGULARITY

Figure 2.1: Stimulus information and calculation of regularity metric for local configurations ................................................. 25
Figure 2.2: Selected subsequences used to test spectral regularity ................................................................. 31
Figure 2.3: Identification of electrodes with significant event-related potentials ............................................. 36
Figure 2.4: Identification of electrodes with significant high-gamma (HG) activity to sequence onsets ................................................................. 37
Figure 2.5: Wide-band neural frequency response to tone-burst sequences ....................................................... 40
Figure 2.6: Identification of electrodes with modulations to temporal regularity ............................................. 40
Figure 2.7: Wide-band frequency relationships between pairwise phase consistency or power and spectral regularity in deviant comparison ....................................................... 43
Figure 2.8: Wide-band frequency relationships between pairwise phase consistency or power and spectral regularity in alternating comparison ....................................................... 44
Figure 2.9: Example electrode traces aligned to subsequences with different local configurations of $F_1$ and $F_2$ ................................................................. 45
Figure 2.10: Relationship between spectral regularity and pairwise phase consistency ............................................. 47
Figure 2.11: Relationship between spectral regularity and amplitude ............................................................. 49
Figure 2.12: Electrode sensitivity to acoustic information by brain region and analysis ............................................. 53

3. NEURAL OSCILLATORY CORRELATES OF CONCURRENT SPECTROTEMPORAL-REGULARITY REPRESENtATION AND DEVIANCE DETECTION

Figure 3.1: Stimulus design and trial progression .................................................................................................. 68
Figure 3.2: Deviant detection task ................................................................................................................. 70
Figure 3.3: Average performance in pilot study ............................................................................................... 79
Figure 3.4: Grand-mean power modulations to standard sequence across subjects ........................................ 83
Figure 3.5: Grand-mean phase-alignment modulations to standard sequence across subjects ........ 84
Figure 3.6: Across-subject mean power and PPC modulations by brain region .............................................. 85
Figure 3.7: Analysis of behavioral performance in neurophysiological experiment ........................................ 86
Figure 3.8: Hit rates by level of frequency deviation .......................................................................................... 88
Figure 3.9: Performance as a function of session number .................................................................................. 88
Figure 3.10: Results of the H vs. CR classifier analysis ......................................................................................... 89
Figure 3.11: Results of the stimulus classifier analysis ......................................................................................... 91
Figure 3.12: Results of the stimulus classifier after first separating deviants by type ........................................ 92
Figure 3.13: Results of the behavioral classifier analysis ..................................................................................... 93
Figure 3.14: Results of the behavioral classifier after first separating by deviant type ........................................ 94
Figure 3.15: Results of behavioral x stimulus classifier .......................................................... 95
Figure 3.16: Results of behavioral x stimulus classifier after first separating by deviant type ...... 96
Figure 3.17: Time-averaged results from stimulus classifier by brain region ................................ 97
Figure 3.18: Time-averaged results from behavioral classifier by brain region ....................... 99
Figure 3.19: Time-averaged results from behavioral x stimulus classifier by brain region on standard trials.......................................................... 100
Figure 3.20: Time-averaged results from behavioral x stimulus classifier by brain region on deviant trials........................................................................................................ 101
Figure 3.21: Correlation between behavioral performance and classification accuracy ............ 102
Figure 3.22: Comparison of behavioral x stimulus classifier results between good and poor performers............................................................................................................. 103
Figure 3.23: Wideband differences in power and PPC as a function of stimulus and behavior... 105
Figure 3.24: Wideband differences as a function of behavior on standard and deviant trials...... 106
Figure 3.25: Results of wideband-difference analyses after first separating by the type of deviant stimulus.......................................................... 107

4. CHARACTERIZING THE IMPACT OF CATEGORY UNCERTAINTY ON HUMAN AUDITORY CATEGORIZATION BEHAVIOR

Figure 4.1: Schematic diagram of the categorical priors employed in the categorization task ..... 122
Figure 4.2: Graph of the Bayesian model .................................................................................. 123
Figure 4.3: The discrimination thresholds for each subject.......................................................... 127
Figure 4.4: Effects of category priors on psychometric data for individual subjects ............... 129
Figure 4.5: Effects of learning ................................................................................................. 130
Figure 4.6: Predictions of the Bayesian model with different categorization behaviors .......... 131
Figure 4.7: Model comparisons using objective priors and individually measured sensory noise σv ............................................................................................................... 132
Figure 4.8: Normalized likelihoods for the Bayesian model predictions.................................... 133
Figure 4.9: Model comparisons using subjective prior distributions with observed individual responses ............................................................................................................. 134
Figure 4.10: Normalized likelihoods for the Bayesian-model fits assuming subjective priors...... 135
Figure 4.11: Individual subjects’ reconstructed category distributions from the model fits for MATCH behavior ........................................................................................................... 136
Figure 4.12: Fitted model parameters for MATCH behavior of the model with subjective priors.. 137
Figure 4.13: Likelihood comparisons for model fits.................................................................... 138
Figure 4.14: Simulations of behavior under the assumption of additional sources of categorical-prior noise ............................................................................................................... 140
Figure 4.15: Variability in category prior probabilities computed as running estimates of P(C=... 141
CHAPTER 1

1. General Introduction

The overarching goal of this dissertation was to understand—both behaviorally and physiologically—how human listeners perceive and categorize sounds. To this end, the experiments conducted for this dissertation aimed to address several key factors underlying auditory perception and decision-making. First, we tested how the statistical properties of acoustic stimuli—in particular, their spectral regularity—were encoded in population-level neural activity. Second, we tested how neural activity correlated with human patients’ behavioral reports, during a task in which they reported on the spectral regularity of an acoustic stimulus. For both of these studies, the patients had medically intractable epilepsy and underwent surgery to implant subdural recording electrodes to localize epileptogenic brain regions. From these electrodes, we recorded large-scale neural-population (e.g., oscillatory) activity in response to acoustic stimuli while patients were engaged in either a passive-listening task (Chapter 2) or an active detection task (Chapter 3). Finally, we tested how learned experiential (i.e., prior) information influences categorization. We tested auditory categorization with the help of volunteer healthy subjects that performed a set of
psychophysical tasks to test how different computation models of perceptual behavior predicted auditory categorization in these subjects (Chapter 4).

This introductory chapter lays the foundation for my thesis work. In it, I describe the concepts of auditory perception and categorization. Next, I describe what is currently known about the nature of the behavioral and/or cortical representations that reflect key aspects of each concept, including the motivation for studying the relationship between neural oscillatory activity and auditory processing. Finally, I describe key missing gaps in our current understanding of auditory perception that is addressed in my thesis work.

The problem of auditory perception

A principle goal of the auditory system is to receive acoustic information from the environment and transform it into perceptual representations that can be used for understanding and interacting with the external world (Cohen et al., 2005; Bizley and Cohen, 2013). In the auditory system, acoustic information from an environmental sound source is first decomposed and represented as a set of acoustic features (e.g., the spectral–or frequency–components in the stimulus). This process begins in the auditory periphery and continues in the early central auditory system (Schnupp et al., 2011). Different acoustic features are either grouped together or segregated, which ultimately results in the formation of perceptual representations (i.e., sounds). These perceptual representations reflect the distinct sound sources in the environment (Bregman, 1994; Cusack, 2005).

In audition, transforming sensory representations into perceptual representations is complicated for two main reasons. First, because acoustic stimuli from distinct sound sources mix, the information that reaches the ear is a mixture of all of the acoustic events in the environment. Consequently, there often is no explicit information in the raw stimulus that conveys which or how many sound sources are present in the environment. Second, because acoustic information from any sound source inherently evolves over time (e.g., speech), the auditory system’s perceptual
representations must reflect the time-varying nature of acoustic stimuli. In the case of speech, this means not only (1) segregating each of an individual’s speech utterances from other sounds but also (2) maintaining a representation of the individual’s speech as the collection of all of his/her utterances over time.

**Spectrotemporal regularities and prediction in auditory perception**

To understand how the auditory system creates perceptual representations, let us first consider how a listener might be able to segregate a speech mixture comprised of the voices of a male and female speaker. One way to distinguish between the two voices might be by differences in their pitch, a perceptual property of sound relating to acoustic frequency. Because male and female speakers are likely to produce speech sounds with different pitch content (due to differences in their vocal apparatuses), the auditory system can exploit this fact to segregate each utterance of the male and female speakers. In a similar manner, the auditory system can use timing information between individual acoustic events to help segregate the male and female voices. Each utterance of the male (or female) speaker will likely be produced at relatively regular points in time, one after the other. In contrast, the utterances between speakers will likely fall at more random times with respect to one another due to the differences in the words that each speaker is producing and their production rate. The auditory system can use this contrast in timing information within and between speakers to help segregate the voices. Finally, if the auditory system detects that the pitch or timing of the current utterance suddenly changes, it must determine whether that stimulus should be grouped with one of the existing perceptual representations or if the stimulus reflects a new speaker entirely. If the utterance suggests the presence of a new speaker, then the auditory system should create a new perceptual representation to reflect the current stimulus environment.

As described in the above example, the auditory system uses implicit *spectral* (i.e., frequency) and *temporal* regularities that exist in acoustic stimuli to group or segregate feature representations into distinct perceptual representations. Experimentally, it has been shown that
many spectrotemporal regularities influence the perceptual segregation of acoustic stimuli, including acoustic frequency (van Noorden, 1975; Bregman, 1994; Cusack, 2005; Micheyl et al., 2005; Winkler et al., 2009), timbre (Iversen, 1995), spatial location (Hill et al., 2011), and amplitude modulation rate (Grimault et al., 2002). Additionally, changes in the rate at which successive stimuli are presented affects perceptual segregation (van Noorden, 1975; Bregman, 1994).

These regularities are useful for perceptual segregation of sounds because acoustic stimuli from distinct sound sources—as discussed above with the female and male speakers—are characterized by distinct spectrotemporal regularities. As a result, the perceptual representations, which are formed on the basis of these regularities, can be used to predict the perceptual qualities of future acoustic events that could be elicited from each sound source (Winkler et al., 2009; Sedley et al., 2016). If an incoming stimulus is predicted by the current organization of perceptual representations, it provides evidence that the current organization should be maintained. In contrast, if an incoming stimulus violates a prediction, it provides evidence against the current organization in favor of an alternative organization (Winkler et al., 2009).

The above examples describe how auditory perception can be conceptually redefined as a predictive process with a prominent role for the detection and tracking of spectrotemporal regularities (Bregman, 1994; Denham and Winkler, 2006; Winkler et al., 2009). First, the auditory system detects spectrotemporal regularities in an acoustic stimulus and uses them to form competing perceptual organizations that differentially group acoustic features into one or more perceptual representations. Second, each competing organization makes distinct predictions regarding the perceptual qualities of future acoustic events based on the regularities assigned to each perceptual representation. Third, the predictions from each organization are compared against the incoming acoustic information and deviations from the detected regularities are used to determine which perceptual organization was most predictive, becoming the dominant organization to be perceived. Finally, this repeats as each successive acoustic event is processed.
Neurophysiological evidence from scalp recordings of electroencephalographic or magnetoencephalographic activity supports the premise that brain activity reflects spectrotemporal regularities in acoustic stimuli. In these studies, stimulus-evoked neural responses called event-related potentials (ERPs) are tested in response to a commonly presented (i.e., standard) stimulus versus the rare (i.e., deviant) stimulus (Näätänen et al., 2007; Winkler et al., 2009). The repeated presentation of the standard stimulus induces a spectrotemporal regularity and the occasional deviant stimulus is used to measure whether ERP responses differentially reflect a “deviation” from a standard stimulus. When a deviant stimulus is presented, two differentiable ERP components of the evoked response are elicited. The N1 response reflects any novel acoustic change between the standard and deviant stimulus, and the mismatch negativity (MMN) is specifically generated when a stimulus deviates from a detected regularity (Garrido et al., 2009; Winkler et al., 2009). For example, both the N1 and the MMN are elicited when a stimulus deviates from the standard by a simple change in its acoustic features (Schröger et al., 1992; Alain et al., 1999; Horváth et al., 2001; Kisley et al., 2004). However, the MMN is elicited also by more abstract deviations that do no elicit a differential N1 response, such as when the pattern of tone pairs changes from a standard pattern (Tervaniemi et al., 1994; Korzyukov et al., 2003) or by omissions of an expected standard stimulus (Yabe et al., 1997; 2001).

In further support of this conceptual model for auditory perception, psychophysical and neurophysiological evidence suggest that alternative perceptual organizations are formed and compete for perception. Psychophysically, listeners are more likely to hear a stimulus consisting of tones that alternate in frequency as a single integrated sound early in listening. But with longer listening durations, listeners become more likely hear the stimulus as two segregated sounds consisting of repeats of the same tone as a function of both the timing and spectral separation between the tones (Bregman, 1994; Micheyl et al., 2005). This transition from hearing one to two sounds suggests that evidence accumulates over time favors a switch from an integrated to segregated perceptual organization. Additionally, the percept of alternating tone sequences at
intermediate spectral separations is bi-stable (van Noorden, 1975; Bregman, 1994; Denham and Winkler, 2006), meaning that perceptual reports are capable of spontaneously and repeatedly switching between that of integrated and segregated organizations. Therefore, both of these alternative organizations must exist and continually compete for perception.

Neurophysiologically, Winkler et al. (Winkler et al., 2005) showed that two differentiable ERP components correlate with different aspects of perceptual organization. In their study, the authors asked participants to listen to an alternating-tone sequence that elicited a bi-stable percept and to continuously report their percept as it alternated between integrated and segregated sounds. Occasionally, one of the lower-frequency tones was omitted. The tone sequence was designed in such a way that a differential ERP response was only expected when a listener perceived an integrated sound (Bregman et al., 2000). The results, however, showed that two distinct ERP components were differentially elicited by the stimulus omissions. First, an early ERP component was elicited to stimulus omissions independent of which percept was reported (integrated or segregated). In contrast, a later ERP component was elicited later only when listeners reported an integrated percept. The early ERP response reflected the prediction mismatch for the integrated organization regardless of whether it was perceived, suggesting that it is always formed, whereas the later response reflected the integrated organization only when it was selected to be perceived.

In summary, multiple lines of evidence support a conceptual model for auditory perception as a predictive process with a prominent role for spectrotemporal regularity representation in forming alternative perceptual organizations and selecting the dominant organization to be perceived based on its predictions of future acoustic events. Consequently, substantial work has been conducted to understand the mechanisms underlying sound segregation and spectrotemporal regularity representation, which I describe in more detail below.
The ventral pathway for auditory perception

In order to study the underlying cortical mechanisms of regularity representation, it is critical to know where neural activity reflects the processes related to auditory perception. At the level of cortex, auditory processing begins in the core auditory fields A1 and R in non-human primates and the homologous regions in the transverse temporal gyrus in humans (Kaas and Hackett, 2000; Hackett, 2008). From these core fields, auditory information is thought to be processed primarily in two major pathways: (1) a dorsal pathway that includes middle and posterior belt regions of auditory cortex with connections first to intraparietal regions and ultimately to non-human primate dorsolateral prefrontal cortex or human premotor cortex; and (2) a ventral pathway that includes middle- and antero-lateral belt regions of auditory cortex and further connections to the primate ventrolateral prefrontal cortex (vIPFC) or human inferior frontal cortex (Kaas et al., 1999; Romanski et al., 1999; Rauschecker and Scott, 2009). Whereas the dorsal pathway is thought to contribute primarily to audiomotor processing (Rauschecker, 2011), the ventral pathway is typically considered the primary pathway for processing related to auditory perception.

Neural rate-place codes for sound segregation

The responsivity of A1 neurons to repeating tone stimuli provides the strongest evidence for a role for neural-rate place codes in sound segregation: that is, that different sounds are reflected by topologically separable populations of active neurons (Micheyl et al., 2005; Bidet-Caulet and Bertrand, 2009). In A1, neurons exhibit a systematic organization of acoustic-frequency sensitivity: each area contains a topographic representation in which neurons systematically ‘prefer’ increasing acoustic frequency (i.e., tonotopy) (Steinschneider et al., 1990; Eggermont, 2001). And when presented with repeated presentations of identical tone bursts, an A1 neuron increasingly adapts as a function of increasing (1) tone-repetition rate and (2) the spectral separation between the tone and the neuron's preferred frequency (Fishman et al., 2001; Ulanovsky et al., 2003; Fishman et al., 2004; Micheyl et al., 2005). Thus, A1 topographically
organizes neural responses to tone stimuli as a function of acoustic frequency, with adaptation modulating the extent of separability with increasing presentation and presentation rate.

Consequently, alternating tone sequences of different frequencies would induce adaptation to both tone frequencies such that only neurons whose preferred frequencies are close to either tone frequency would not adapt. When the frequency separation between the tones is small, the active neurons would occupy a single topographic location in A1. In contrast, when the frequency separation between the tones is large, the active neurons would occupy relatively separable locations. This neurophysiological effect mirrors the psychophysical characteristics of alternating-tone segregation, which show that listeners are more likely to report hearing a single sound when the frequency separation between the tone frequencies is small and two sounds when the frequency separation is large (van Noorden, 1975; Bregman, 1994; Micheyl et al., 2005).

A neural rate-place code is also sufficient to explain the perceptual switch between hearing one to two sounds over time in alternating tone stimuli (Bregman, 1994; Micheyl et al., 2005; 2007). Following the onset of the stimulus, A1 neural responses to either tone frequency are relatively strong because adaptation has yet to take effect, resulting in one large active population that would favor the percept of a single sound. But with longer listening times, neural adaptation minimizes the responses of neurons that prefer other tone frequencies, thus increasing the separability of active populations in favor of two sounds.

Finally, a rate-place code can also account for streaming based upon differences in spatial location (Middlebrooks and Bremen, 2013). In a manner similar to stream segregation by acoustic frequency, neurons in A1 adapt differentially to repeated sounds arising from the same or different spatial locations, resulting in topographically similar or distinct active regions that would be read out as one or two auditory streams. Thus, it appears that rate-place codes may play a fundamental role in sound segregation.
Complementary role of temporal coherence in sound segregation

In addition to a neural rate-place code, more recent studies suggest that the timing of neural activity must also play a role in sound segregation. Elhilali et al. (Elhilali et al., 2009) found that when two tone bursts of different frequencies were presented synchronously rather than alternating, listeners were more likely to report hearing one sound, even for tone bursts with large frequency separations. If a neural rate-place code were sufficient to explain these results, then for large frequency separations, synchronous or alternating tone sequences should be represented in topographically similar or distinct A1 regions, respectively. However, neural responses in A1 did not differ between the synchronous or alternating conditions at any frequency separation. Therefore, a neural rate-place code could not differentiate between the synchronous or alternating condition, suggesting it is insufficient to fully describe sound segregation.

To reconcile this paradox, Elhilali et al. (Elhilali et al., 2009) proposed a temporal-coherence model of stream segregation: streams are formed on the basis of the detection of neural populations with temporally coherent activity. This model specifically includes a process of temporal integration, whereby activity from distinct acoustic-frequency channels is integrated simultaneously over multiple timescales by temporal-rate filters with different time constants. Thus, for synchronous tone sequences or alternating sequences with small frequency separations, the active neural population(s) would respond in a temporally coherent manner, which could be read out downstream as evidence for a single stream. On the other hand, alternating tone sequences with large frequency separations produce two neural populations responding in an anti-coherent manner and would be interpreted as two distinct auditory streams.

Ultimately, it is likely that both neural topography and temporal coherence play complementary roles in stream formation. Although neural rate-place codes can explain certain aspects of sound segregation, there is no clear explanation for the perceptual bi-stability of certain stimuli (Denham and Winkler, 2006). Alternatively, a strict interpretation of temporal coherence is also likely insufficient, as more recent studies have found that sounds that would elicit temporally
coherent activity can, in fact, be segregated into separate sounds under certain conditions (Micheyl et al., 2010; 2013a; 2013b).

Currently, the underlying mechanistic explanation for the detection of temporal coherence in neural populations is under debate. Because single neurons have been shown to act as coincidence detectors (Yin and Chan, 1990) and information integrators (Huk and Shadlen, 2005), it is possible that multi-timescale temporal integration could be computed explicitly by the firing activity of individual neurons. However, Elhilali et al. (Elhilali et al., 2009) could not identify any neurons in A1 with reliable firing patterns necessary for these computations. Alternatively, neural oscillatory activity could underlie temporal integration processing. Indeed, neural oscillations have received ever-increasing scientific interest due to their prevalence in cortex and potential for explaining various aspects neural communication and sensory-feature binding for perception (Brown et al., 1996; Traub et al., 1996; Engel and Singer, 2001; Engel et al., 2001; Meador et al., 2002; Ward, 2003; Buzsáki and Draguhn, 2004). Below, I briefly describe neural oscillations and their relation specifically to aspects of spectrotemporal-regularity representation and auditory perception.

**Mechanisms of temporal coherence and neural oscillations**

Neural oscillations reflect the large-scale coordinated activity of neural populations over time scales ranging from as short as 2-15 ms (70–500 Hz) to >10 s (<0.1 Hz) (Penttonen and Buzsáki, 2003; Buzsáki and Draguhn, 2004; Jensen and Colgin, 2007; Lőrincz et al., 2009; Zuo et al., 2010). Neural oscillatory activity arises from the interaction between intrinsic properties of individual neurons and circuit-level dynamics (Hutcheon and Yarom, 2000; Destexhe and Sejnowski, 2003; Whittington and Traub, 2003; Buzsáki and Draguhn, 2004). Individual neurons exhibit resonance in their membrane-potential activity via intrinsic cellular properties that allows them to select for inputs with particular frequency characteristics (Gupta et al., 2000; Hutcheon and Yarom, 2000; Marshall et al., 2002; Thomson and West, 2003; Buzsáki and Draguhn, 2004; Gai et
al., 2014). The complex interactions between neurons with different or varying degrees of resonance properties ultimately gives rise to the macro-scale presence of neural oscillatory activity. For instance, a model of synaptically coupled interneuron populations is sufficient to produce gamma-frequency (~30-70-Hz) oscillations when provided sufficient excitatory drive (Wang and Rinzel, 1992; Traub et al., 1996; Buzsáki and Wang, 2012). Under this regime, oscillatory activity is induced when the activity of one sub-population of interneurons begins to synchronize, sending temporally aligned inhibitory post-synaptic potentials (IPSPs) to another interneuron sub-population. In turn, these inhibited interneurons fire synchronously due the excitatory drive after the decay of hyperpolarization, sending IPSPs back to the other sub-population and leading to a repeat of the cycle (Wang and Rinzel, 1992; Traub et al., 1996; Buzsáki and Wang, 2012). The frequency of this cycle of mutual inhibition is determined largely by the level of excitation and the kinetics of the IPSP decay (Whittington et al., 1995; Wang and Buzsáki, 1996). Adding pyramidal neurons to this model network induces excitatory responses that are time-locked to the gamma cycle.

A variety of other similar interactions within and among neural populations account for the generation of oscillations in other frequency ranges (Whittington et al., 2000; Buzsáki et al., 2003; Destexhe and Sejnowski, 2003; Whittington and Traub, 2003), with both neuron class (Whittington et al., 2000; Whittington and Traub, 2003; Buzsáki et al., 2004) and neuromodulation (Destexhe et al., 1994; Destexhe and Sejnowski, 2003; Bauer et al., 2012; Neymotin et al., 2013) known to influence the frequency and strength of oscillation. Additionally, the size of the neural populations in both a spatial and numerical sense also influences the frequency of oscillations: smaller neural populations are capable of higher-frequency oscillations, whereas larger populations oscillate at lower frequencies (Steriade, 2001; Csicsvari et al., 2003; Buzsáki and Draguhn, 2004).

**Neural oscillations in relation to audition**

Several key characteristics about the nature of oscillatory activity in cortex implicate a functional role for neural oscillations generally in neural information processing. Here, I describe
these characteristics as they relate to auditory perception and summarize the current understanding of the relations between oscillatory activity and audition.

First, oscillatory activity is uniquely positioned to process natural auditory stimuli, which are characterized by a complex set of acoustic features and regularities that are organized across multiple time scales (see Chapter 3). Neural oscillatory activity is hierarchically organized (Bak et al., 1987; Steriade, 2001; Csicsvari et al., 2003; Sirota et al., 2003; Buzsáki and Draguhn, 2004; Lakatos et al., 2005), with the power or phase of slower oscillations capable of modulating activity in faster oscillations. Additionally, the phase of an oscillation itself is correlated with the probability of neural firing activity (Lakatos et al., 2005): neurons are more or less likely to fire action potentials at ‘high’- or ‘low’-excitability phases of an oscillation, respectively. These findings, combined with the fact that multiple oscillatory rhythms can occur simultaneously and interact within and across regions (Steriade, 2001; Varela et al., 2001; Csicsvari et al., 2003; Buzsáki and Draguhn, 2004; Lakatos et al., 2005; Canolty et al., 2006; 2010), suggest that neural oscillations could be a potential mechanism to simultaneously integrate information across multiple temporal scales and brain regions in order to form perceptual representations of sounds (Varela et al., 2001; Buzsáki and Draguhn, 2004; Sejnowski and Paulsen, 2006; Canolty and Knight, 2010).

Second, neural oscillations reflect spectrotemporal regularity by oscillatory ‘entrainment’. When stimulated with a sequence of identical tones at a constant repetition rate, a neural oscillation at the frequency of the repetition rate reliably phase-aligns to each tone onset (Lakatos et al., 2008; Besle et al., 2010; Lakatos et al., 2013). Additionally, oscillatory entrainment occurs for more complex patterns of spectral or amplitude modulations (Patel and Balaban, 2000; Ross et al., 2000; Liégeois-Chauvel et al., 2004; Luo et al., 2006; Bidet-Caulet et al., 2007; Henry and Oblés, 2012; Henry et al., 2014), and multiple oscillatory frequencies have been shown to entrain simultaneously to concurrent regularities with different time scales (Henry et al., 2014). Finally, differential phase entrainment is thought to at least partly underlie the MMN signal that reflects deviations in spectrotemporal regularity (Fell et al., 2004; Klimesch et al., 2004; Fuentemilla et al., 2006;
Hanslmayr et al., 2007; Klimesch et al., 2007; Sauseng et al., 2007; Fuentemilla et al., 2008; Hsiao et al., 2009).

Third, oscillatory entrainment can be modulated by stimulus features and attention (Patel and Balaban, 2000; Lakatos et al., 2008; Besle et al., 2010; Lakatos et al., 2013), potentially allowing for stimulus-dependent and flexible control of neural processing related to sensory and perceptual selection. For example, an A1 site will align its high-excitability phase to a temporally regular sequence of tone bursts when the tone-burst frequency matches the site’s preferred frequency (Lakatos et al., 2013). In contrast, that same site will align its low-excitability phase to a sequence when the tone-burst frequency is far from the site’s preferred frequency. In a similar manner, oscillations at a cortical site tend to differentially align when a stimulus sequence is attended versus ignored (Lakatos et al., 2008; Besle et al., 2010; Lakatos et al., 2013).

Finally, neural oscillatory activity has been shown to correlate with various aspects of perception. Both behavioral performance and reaction times in detection tasks are modulated by the phase of particular low-frequency oscillations that are reliably modulated by regularities in the tasks (Stefanics et al., 2010; Henry and Obleser, 2012; Henry et al., 2014). Thus, it appears that neural oscillations have the potential to be fundamentally important for auditory perception specifically, and perhaps neural processing in general.

**Missing gaps in the neural-oscillatory correlates of audition**

Despite the current evidence in support of a role for neural oscillations in audition, critical aspects of our understanding of the relationships between oscillations and auditory perception are still poorly understood. Whereas numerous studies have focused on how oscillatory activity reflects simple stimulus regularities (Dimitrijevic et al., 2001; Lakatos et al., 2005; Luo et al., 2006; Fuentemilla et al., 2008; Hsiao et al., 2009; Lakatos et al., 2013), little is known about how oscillatory activity tracks regularity dynamically in an ongoing stimulus (Patel and Balaban, 2000; Bendixen et al., 2007; Barascud et al., 2016), as would be necessary in natural settings. Moreover,
oscillatory activity in response stimuli with multiple concurrent regularities exhibit a complex pattern of phase and amplitude modulations in population-level activity that is not well understood (Patel and Balaban, 2000; Luo et al., 2006; Henry et al., 2014).

It is also unclear how the oscillatory representation of spectrotemporal regularity relates to the representation of regularity deviations (Pannese et al., 2015), and whether changes in neural oscillatory activity correlate specifically with behavioral reports of deviance detection. If neural oscillatory activity is causally related to the representation of spectrotemporal regularity, then the ability to detect deviations in spectrotemporal regularities should be reflected in changes in neural oscillatory activity in a manner that relates directly to the timescale of the detected deviation.

Finally, most studies have focused primarily on the oscillatory contributions of the core auditory cortex (Lakatos et al., 2005; 2013) or have studied these contributions with EEG (Dimitrijevic et al., 2001; Bendixen et al., 2007; Fuentemilla et al., 2008; Stefanics et al., 2010; Henry and Obleser, 2012; Henry et al., 2014), for which spatial resolution is poor. However, multiple lines of evidence suggest sensory and perceptual processes related to audition are distributed across multiple regions of cortex, including downstream regions of the ventral pathway (Belin et al., 2000; Rauschecker and Tian, 2000; Romanski and Goldman-Rakic, 2002; Halpern et al., 2004; Romanski et al., 2004; Gifford et al., 2005; Petkov et al., 2008; Russ et al., 2008a; 2008b; Bizley et al., 2009; Hall and Plack, 2009; Lee et al., 2009; Bizley et al., 2010; Tsunada et al., 2011; Niwa et al., 2012; Plakke et al., 2012; Bizley and Cohen, 2013; Niwa et al., 2013; Tsunada et al., 2015), regions along the dorsal pathway (Belin et al., 2000; Cusack, 2005; Ghazanfar and Schroeder, 2006; Hill et al., 2011; Teki et al., 2011; Rauschecker, 2012; Teki et al., 2016), and even other regions not considered to be part of either pathway (Belin et al., 2000; Poremba et al., 2004; Ghazanfar and Schroeder, 2006; Petkov et al., 2008; Teki et al., 2016). Therefore, extent to which oscillatory correlates of auditory perception are distributed along the auditory cortical pathway and beyond into multisensory brain regions remains unclear. Chapters 2 and 3 of this dissertation aim to address these outstanding issues.
Auditory categorization

Although auditory-scene analysis provides valuable information about distinguishable putative sound sources in the environment, additional information is often required to make use of this perceptual information to guide decision-making and behavior. For example, simply knowing that there are distinct putative sound sources in the environment does not necessarily tell the listener about the identity of the sound source (e.g., a trumpet) or how to respond. One can readily imagine being able to segregate sounds and even being able to describe their qualities but not knowing their identity. Indeed, a person’s speech is capable of providing information regarding approximate age, gender, country of origin, and affect but we may not know the speaker’s identity. Finally, it would be infeasible for a listener to map behaviors to every possible auditory perception in the high-dimensional and continuous perceptual space (Seger and Miller, 2010). Therefore, it is critical that the brain has a process for flexibly organizing the perceptual space into robust hierarchical and discrete representations that provide understanding and a practical means with which to respond adaptively to the environment. This fundamental process is known as categorization.

Categorization is a natural and adaptive process that allows a listener to flexibly ignore (or treat equivalently) certain kinds of variability in acoustic stimuli while simultaneously utilizing other kinds of variability that might be important (Russ et al., 2007). When someone yells “fire!” in a crowded movie theater, categorizing his or her age or gender may not be as important as the underlying meaning of the speech signal. However, there may be other times when age and gender categories do provide useful information. A person may change the content and tone of a conversation depending on whether he or she is speaking with a young female versus an adult male. In general, the ability to flexibly categorize perceptual representations allows for flexible behaviors depending on the context.
The perceptual ease with which sounds can be categorized belies the complex computations underlying this ability. One reason categorization is complex is that a sensory property (e.g., harmonicity) may be ambiguous with respect to the stimulus' category membership. For example, because both dogs and wolves can produce howls, the harmonic structure of the howl by itself may not provide enough information to the listener for proper identification of the caller. In such cases, and in the absence of other sensory information, the listener needs to rely on other sources of information to correctly categorize a sound and identify whether the howl came from a dog or a wolf. This information can be prior knowledge such as knowing that the probability of encountering a wolf is low. Since prior information is subjective, it is of fundamental interest to understand how an observer (1) acquires prior information and (2) then uses this subjective information together with the sensory signal to perform categorical judgments.

Using novel stimuli, experimenter-defined categories, and category priors (i.e., the probability of encountering a stimulus from a given category), one can study how observers learn prior information and use this information to perform categorical judgments. The utility of prior information and the strategies employed during categorization judgments have been best studied in the vision and decision-making literature, which is a general form of categorization (Ashby and Berretty, 1997). One common property of categorical judgments is known as probability matching, whereby the probability of an observer’s choice of a particular category for an ambiguous stimulus matches the underlying prior probability of encountering a stimulus from that category during the experiment (Thomas and Legge, 1970; Healy and Kubovy, 1981; Vulkan, 2000). This type of behavior is generally sub-optimal with respect to minimizing categorization (or decision) errors (Ashby and Berretty, 1997; Vulkan, 2000; Gifford et al., 2014). However, it has been argued that probability matching is not a decision strategy per se (Healy and Kubovy, 1981; Ashby and Berretty, 1997; Gifford et al., 2014) and that, indeed, sub-optimal behavioral performance is actually consistent with an optimal decision strategy employed under various degrees of perceptual and categorical uncertainty (Ashby and Maddox, 1993; Ashby and Alfonso-Reese, 1995; Ashby and
Berretty, 1997; Gifford et al., 2014). Alternatively, it is possible that probability matching is an implicit strategy that could reflect an observer’s tendency to search for patterns in random environments (Ayton et al., 1989; Wolford et al., 2004; Wozny et al., 2010).

**Missing gaps in understanding auditory categorization**

Whereas the utility of prior information to inform categorical judgments has been well studied in vision (Lee, 1963; Lee and Janke, 1964; 1965; Ulehla, 1966; Healy and Kubovy, 1981; Ashby and Berretty, 1997; Bohil and Maddox, 2001; Hansen et al., 2012a; 2012b), our understanding of how prior information informs categorical judgments in audition is relatively limited and has only more recently become an active area of research (Sullivan et al., 2005a; 2005b; Holt and Lotto, 2006; Ley et al., 2012; Scharinger et al., 2013). More importantly, auditory categorization has not been tested in situations in which the auditory stimulus is ambiguous with regard to its category membership. Understanding these aspects of auditory categorization are important for determining modality-specific versus more general strategies involved in the categorization process, which can provide insights into the types of neural computations required to perform these categorizations. The final part of this dissertation addresses this outstanding issue by (1) testing whether human listeners use prior information to inform categorical judgments when category identity is uncertain and (2) determining the computational strategy that human listeners employ to make their categorical judgments.
CHAPTER 2

2. Neural-phase alignment is a mechanism for tracking dynamic changes in acoustic spectral regularity


ABSTRACT

A fundamental goal of the auditory system is to transform auditory stimuli from low-level representations of a stimulus’ acoustic features into perceptual representations (i.e., sounds). These perceptual representations are the computational result of the brain’s ability to dynamically track and then group or segregate these acoustic features based on their shared or different spectrotemporal regularities. Here, we identified the mechanisms by which the brain tracks and encodes changes in the spectral regularity of an ongoing acoustic stimulus. We identified these
mechanisms by recording electrocorticographic activity in humans in response to pseudorandom sequences of tone bursts. These sequences had a constant tone repetition rate and varied only in the pattern of tone frequencies (i.e., spectral regularity) over time. We found that the degree of oscillatory-phase alignment in multiple neural-frequency bands dynamically tracked spectral regularity, whereas the amplitude of the neural oscillations did not. Moreover, we identified a complex relationship between these phase-alignment modulations and neural-frequency band. Some neural-frequency bands—both harmonically related and unrelated to the tone repetition rate—were positively modulated by spectral regularity, whereas others were negatively modulated. In particular, phase alignment in the delta frequency band seemed to be the best indicator of spectral regularity. Finally, we found that these regularity representations existed throughout cortex. This widespread reliable modulation in phase alignment—both in neural-frequency space and in cortical space—suggests that phase-based modulations may be a general mechanism for tracking regularity in the auditory system specifically, and perhaps other sensory systems more generally. Our findings also support a general role for the delta-frequency band in processing the regularity of auditory stimuli.

INTRODUCTION

A fundamental goal of the auditory system is to transform auditory stimuli from low-level sensory representations of a stimulus’ acoustic features into perceptual representations (i.e., sounds or ‘auditory streams’) (van Noorden, 1975; Bregman, 1994; Cusack, 2005). Auditory streams are the result of the brain’s ability to group auditory stimuli with similar acoustic spectrotemporal regularities into one auditory stream. Stimuli with different regularities are segregated into different auditory streams (Bregman, 1994; Shinn-Cunningham, 2008; McDermott, 2009; Winkler et al., 2009).
Consider, for example, our ability to detect and follow a friend’s speech sounds (voice) in a noisy party. Natural sounds, like speech, are often harmonic (i.e., at any instant of time, the spectral content of speech occurs at integer multiples of the lowest [fundamental] frequency). Because neural representations of these multiple frequency bands occur simultaneously, the auditory system tends to group this information together into a single stream (i.e., ‘your friend’s voice’). Further, because changes in harmonic structure occur slowly and smoothly over time (i.e., sequentially), we can follow his/her voice throughout the conversation. What happens, though, when a rude person interrupts the conversation? We can segregate this person’s voice from our friend’s due, in part, to differences in the harmonic structure (i.e., the spectral regularity) of their voices (van Noorden, 1975; Bregman, 1994). This segregation occurs despite the fact that there is not an explicit distinction in the spectral content of the two voices or, more generally, between the acoustic stimuli of different sound sources (Cherry, 1953; Bregman, 1994).

Although a large literature has examined the neural mechanisms that contribute to a listener’s ability to detect and track these spectral regularities over time, several open issues still remain. (1) Numerous studies, both at the single-neuron and population level (i.e., oscillatory activity), have identified neural correlates reflecting simple spectral regularities (Dimitrijevic et al., 2001; Ulanovsky et al., 2003; Fishman et al., 2004; Micheyl et al., 2005; Luo et al., 2006; Fuentemilla et al., 2008; Hsiao et al., 2009; Lakatos et al., 2013) and simple changes from regularity (i.e., the appearance of a novel stimulus in the midst of a stream otherwise identical stimuli; for reviews, see (Näätänen et al., 2007; Escera et al., 2013)). However, these studies focused exclusively on stimuli with spectral regularities that remained constant over time. As such, although these studies describe potential neural mechanisms for certain types of regularities, they did not specifically address how the auditory system can dynamically track changes in regularities. (2) Single neurons have been shown to adapt differentially to regularities on multiple timescales (Ulanovsky et al., 2004), but it is unclear if firing rates alone can fully account for dynamic regularity tracking. Electro- and magneto-encephalographic studies have found that population-level activity
is modulated by changes in spectral regularity (Winkler et al., 1996; Ulanovsky et al., 2004; Bendixen et al., 2007; Barascud et al., 2016). However, the findings from these studies are also somewhat limited because they did not systematically test the relationship between neural activity and spectral regularity. (3) Finally, because most studies have focused on the contribution of the core auditory cortex (Ulanovsky et al., 2003; Fishman et al., 2004; Lakatos et al., 2005; Micheyl et al., 2005; Lakatos et al., 2013), the contributions of other regions of the auditory cortex and other cortical regions have yet to be fully elucidated. Thus, the goal of this study was identify the mechanism by which the cortex (both auditory and non-auditory areas) dynamically encodes the degree spectral regularity of an acoustic stimulus.

To achieve this goal, we recorded electrocorticographic (ECoG) activity from electrodes that were distributed across the human cortex while patients listened passively to pseudorandom sequences of alternating tone bursts. The sequences were designed to have dynamical changes in their degree of spectral regularity over short time scales (200-700 ms). We quantified spectral regularity based on the temporal progression of acoustic frequencies in short subsequences (2-7 tone bursts) within the larger stimulus sequence. Specifically, we tested how ECoG activity during the final tone in a subsequence was modulated (i.e., conditioned) by the spectral regularity of the preceding tones in the subsequence. We found that ECoG phase alignment—but not power—correlated with the spectral regularity. Specifically, phase alignment in the delta (<3 Hz) frequency band seemed to be the best indicator of spectral regularity. This finding is consistent with the hypothesis that this frequency band may have a general role in acoustic scene analysis (Giraud and Poeppel, 2012; Doelling et al., 2014; Riecke et al., 2015). Spectral regularity also correlated, to a lesser extent, with phase alignment at the fundamental frequency of the tone-burst repetition rate, its first harmonic, and off-harmonic frequencies. We found these relationships throughout cortex. Together, these results suggest that the degree of phase alignment is a mechanism that can track spectral regularity—and hence, the segregation of stimuli into discrete auditory streams. Finally, because many of these frequency bands were not related to the temporal features of the
acoustic stimulus, it is possible that phased-based modulations are a general mechanism by which the brain tracks regularity across stimulus modalities.

METHODS

Participants

Eleven participants (5 females, 4 left-handed, mean age: 30.1±12.8 years) with medically intractable epilepsy underwent surgery to implant subdurally platinum recording electrodes on the cortical surface and into the brain parenchyma. In each case, clinical teams (either at the Hospital of the University of Pennsylvania or Thomas Jefferson University Hospital) determined electrode placement in order to localize epileptogenic brain regions. Institutional review boards at each hospital approved the research protocol. Informed consent was obtained from each participant prior to their participation in this study.

One participant was implanted twice (time between implantations: ~1 year) and participated in our experiment on both occasions. Because of the time between implantations, differences in electrode placement (i.e., the surgical team targeted different brain regions in each surgery), and because the patient was presented with unique tone-burst sequences (see below) during each experimental session, we treated the data obtained from the two implantations as independent data sets. Thus, a total of twelve subjects completed the task.

Auditory stimuli and task design

We designed the auditory stimuli and task to test how the power and phase alignment of electrocorticographic (ECoG) signals were modulated by local dynamic changes in the spectral regularity of an auditory stimulus. Specifically, we tested how ECoG activity was modulated (i.e., conditioned) by the preceding degree of spectral regularity. Our analyses focused on testing how
ongoing ECoG activity was modulated by regularity changes that occurred in the preceding 200-700 ms time interval.

**Stimulus.** The acoustic stimuli were 48 unique tone-burst sequences (65 dB SPL; 50-ms duration; 5-ms \(\cos^2\) ramps; 50-ms inter-tone interval [10 Hz onset-to-onset interval]). Tone-burst frequencies were either 1000 Hz (\(F_1\)) or 1029 Hz (\(F_2\); \(\frac{1}{2}\) semitone above \(F_1\)). Each tone-burst sequence was constructed by concatenating 4 m-sequences together (Golomb, 1982; Kvale and Schreiner, 1995; Buračas and Boynton, 2002). Because of this design, the temporal progression of frequencies \(F_1\) and \(F_2\) in each sequence was stochastic. From this sequence, we could identify different length subsequences with different spectral regularities. To be clear, because the tone bursts were presented at a constant rate, spectral regularity was manipulated independent of temporal regularity. The tone-burst frequencies and presentation rate were chosen to minimize the possibility that subjects could segregate the sequence into two separate auditory streams (van Noorden, 1975; Bregman, 1994; Cusack, 2005). Each sequence contained 476 tone bursts (for a stimulus duration of 47.6 s per sequence), which was preceded by 22.4 s of silence (70-s total duration). The tone-burst sequences were delivered via calibrated insert-ear buds (ER-MC5, Etymotic) that were connected to a laptop (either a 15-inch MacBook Pro or a 13-inch MacBook Air, Apple).

The spectral regularity of a subsequence was characterized by its specific local configurations of \(F_1\) and \(F_2\). For example, consider the subsequence of ‘length’ = 3: \(F_1\)→\(F_1\)→\(F_1\). This subsequence is ‘perfectly’ regular (and, hence, predictable) because it consists of three presentations of the same frequency. In contrast, this equally long subsequence \(F_1\)→\(F_1\)→\(F_2\) is less regular because the third tone burst is \(F_2\) and not \(F_1\). Consider, also, the following two subsequences with length = 4: \(F_1\)→\(F_1\)→\(F_1\)→\(F_1\) and \(F_1\)→\(F_2\)→\(F_1\)→\(F_2\). Both subsequences are regular in that the current tone-burst can be predicted based on its prior history. However, because their regularities occurred over different time scales (1-back versus 2-back, respectively), they have different degrees of spectral regularity. Below, we discuss a metric that quantifies the degree of
spectral regularity in each subsequence (see BAND-SPECIFIC ANALYSES MEASURING MODULATIONS TO SPECTRAL REGULARITY).

Together, this m-sequence algorithm was advantageous because it helped to ensure that 
(1) \( F_1 \) and \( F_2 \) occurred with approximately equal probability across an entire tone-burst sequence; 
and (2) specific local configurations of \( F_1 \) and \( F_2 \) occurred a predictable number of times within a sequence (Fig. 2.1a and 2.1b). Consequently, each sequence contained \( \geq 1 \) instances of all possible local configurations up a subsequence length of 7 (i.e., a timescale between 200-700 ms or 2 to 7 tone bursts; Fig. 2.1b). However, because each of the 4 m-sequences that constituted each tone-burst sequence (see above) had an odd number of tone bursts, there was a slight imbalance in the number of tone bursts with frequency \( F_1 \) and \( F_2 \). Thus, there was a slight imbalance in the number of instances of local configurations that had opposite temporal progressions (e.g., \( F_1 \rightarrow F_1 \) versus \( F_2 \rightarrow F_2 \)); this imbalance increased with subsequence length (Fig. 2.1b).

Task design. Subjects rested comfortably in their hospital beds and took part in a ‘passive-listening’ task, during which they listened quietly to the tone-burst sequences. Subjects could read but were asked to refrain from speaking. Subjects completed 2-6 sessions of the task. During each session, participants listened to 10 unique tone-burst sequences that were chosen randomly from the test bank of 48 sequences. Sessions were separated by at least 1 minute and at most 12 days.

Data acquisition and preprocessing

Subdural electrodes were arranged in either grids or strips; each electrode contact was separated by 10 mm. Depth electrodes contained 6-8 contacts that were separated by 8 mm; the depth electrodes were located primarily in the medial temporal lobes. Electrodes were localized by co-registering post-operative computed-tomography scans with post-operative MRI scans using the FSL (FMRIB [Functional MRI of the Brain] Software Library), BET (Brain Extraction Tool), and
FLIRT (FMRIB Linear Image Registration Tool) software packages. These electrode locations were then mapped to Talairach space using indirect stereotactic techniques and the OsiriX Imaging Software DICOM viewer package (Burke et al., 2013).

![Figure 2.1: Stimulus information and calculation of regularity metric for local configurations. (a) Schematic portion of a tone-burst sequence depicting the pseudo-random structure of the tone-frequency progression. (b) Probabilities of example local configurations as a function of subsequence length (SL). For subsequence lengths of 3–7, we list four example configurations and then the mean probability of all other local configurations. The standard deviation for all local configurations is ≤0.003. (c) Summary table of example local configurations and their respective Kolmogorov complexity (C_K) values. First row: pattern of tone-frequency progression. Three different subsequences of local configurations of F_1 and F_2 are highlighted in red, green, and blue from left to right. Each configuration has 7 tone bursts. Second row: simplification of highlighted tone-frequency progressions into shorter, repeated patterns. The red-highlighted local configuration is the “most regular” because it can by simplified into seven repeats of the shorter pattern (F_1). The green-highlighted local configuration is less regular because it can be simplified into 3.5 repeats of the shorter pattern (F_1–F_2). The blue-highlighted local configuration is the least regular because it cannot be simplified into a pattern shorter than its entire length. Fourth row: quantitative measure of regularity using C_K metric. The regularity (1/C_K) values for the highlighted local configurations correlate with regularity. The C_K metric can be quantitatively compared across all local configurations that have the same length.

We recorded ECoG signals either with a Nicolet or a Nihon Kohden electroencephalogram system (Burke et al., 2013). ECoG signals were sampled at 1000 Hz. A testing laptop sent ±5-V analog pulses, via an optical isolator, to open lines in the clinical-recording system to align the stimulus- and task-related events with the ECoG recordings.

To minimize reference-line and volume-conduction confounds, we used a bipolar-referencing scheme (Nunez and Srinivasan, 2006; Burke et al., 2013) in which we subtracted the signals from each pair of immediately adjacent electrode contacts on the same grid, strip, or depth.
electrode (Anderson et al., 2010; Burke et al., 2013). We assumed that these bipolar signals were located midway between each electrode-contact pair.

When we measured the event-related potentials that were elicited by stimulus onsets (see Approach), ECoG activity was down-sampled to 250 Hz and then low-pass filtered ($2^{nd}$ order, zero-phase-shift Butterworth filter; pass-band 0-50 Hz). When we tested high-gamma (HG) activity, we down-sampled ECoG activity to 500 Hz and band-pass filtered it between 70–200 Hz ($2^{nd}$-order zero-phase-shift Butterworth filters). When we tested ECoG sensitivity to stimulus regularities (see Approach), we down-sampled ECoG activity to 500 Hz and either notch-filtered ($4^{th}$-order zero-phase-shift Butterworth filter; stop-band: 58-62 Hz) it to remove power-line noise for wideband analyses or band-pass filtered it for frequency-band-specific analyses ($4^{th}$-order zero-phase-shift Butterworth filters with $\sim$3–4-Hz pass bands).

**Approach**

Our general approach to data analysis was to quantify ECoG sensitivity to increasingly more complex components of the tone-burst sequence. **First**, we identified ECoG signals that were modulated by the onset of a tone-burst sequence, independent of its temporal and spectral regularities. **Second**, we tested the sensitivity of the ECoG signal to the temporal regularity of the sequence, independent its spectral regularity. **Finally**, we tested the sensitivity of the ECoG signal to changes in spectral regularity. In particular, we designed our analyses to evaluate the hypothesis that ECoG activity at the time of the *current* tone was *conditioned* on the spectral regularity of the *previous* tone bursts (Patel and Balaban, 2000; Denham and Winkler, 2006; Fuentemilla et al., 2006; 2008; Hsiao et al., 2009; Winkler et al., 2009; Lakatos et al., 2013). For example, consider the last $F_2$ in these two subsequences:

1. $F_2$—$F_1$—$F_2$—$F_1$—$F_2$
2. $F_2$—$F_1$—$F_1$—$F_2$—$F_2$

The first subsequence’s local configuration of $F_1$ and $F_2$ is more regular (and, hence, more predictable) than the second’s local configuration. Because of this difference in regularity, we
hypothesized that the ECoG signal in response to the last $F_2$ (in bold font) in regular subsequence (1) should be different than the ECoG signal in response to the less regular subsequence (2). The following sections describe each of these three analyses in detail.

(1) **Identifying ECoG signals that are reliably modulated by the tone-burst sequence.** For each electrode, we conducted three separate analyses to test for modulations in ECoG activity due to the tone-burst sequences.

First, we measured event-related potentials (ERPs). We isolated 1-s segments of ECoG activity following sequence onset and, on a trial-by-trial basis ($N = 10 \times$ number of sessions for each subject), normalized this activity by the mean and standard deviation of the immediately preceding 1-s period of ‘baseline’ activity (i.e., the silent period prior to sequence onset). These 1-s segments were then averaged together to form the ERP. Next, we identified the longest contiguous time period after stimulus onset in which this mean z-scored ERP signal was significantly different from zero ($t$-tests, raw $p<0.05$). We implemented a randomization procedure to calculate the false-positive rate. For each randomization, we extracted random 2-s segments of ECoG activity (10 random segments per completed session per subject) and performed the same normalization as described above. Subsequently, we performed $t$-tests to identify the longest contiguous time period that was significantly different from zero in these random ECoG signals by chance (raw $p<0.05$). This procedure was repeated 1000 times to obtain a null distribution of durations of significant time periods to compare against the observed duration from the actual ERP. An electrode’s ERP was ‘reliable’ if its duration of significant contiguous time periods was at least in the upper 95th percentile (i.e., corrected $p<0.05$) of this null distribution.

Second, we also measured HG activity in response to sequence onsets. This analysis was conducted in an analogous manner to the ERP analysis, with instantaneous HG amplitude extracted from the filtered and onset-aligned ECoG signals using the Hilbert transform (2-s buffers).
Third, we tested for HG sensitivity to tone-burst frequency (i.e., was an electrode’s ECoG activity modulated more by frequency $F_1$ or $F_2$?). For each electrode, ECoG activity was aligned to the tone-burst sequences and instantaneous HG activity across the whole stimulus sequence was extracted from the ECoG signals using the Hilbert transform. HG activity was then averaged into 100-ms bins corresponding to the duration of each tone burst and the inter-tone interval. Averaged HG activity was then sorted into two groups based upon tone-burst frequency and then a grand-averaged response was computed as a function of tone-burst frequency. HG sensitivity to tone-burst frequency was tested across the tone-burst sequences using signed-rank tests, extracting the z-scored test statistic for each electrode. We used a randomization procedure to determine the significance of each z-score and to estimate the false-positive rate. In this procedure, we randomized the relationship between averaged HG activity and tone-burst frequency for each tone-burst sequence prior to computing the grand-average responses signed-rank test. This randomization was repeated 1000 times to create a null distribution of z-scores that was used to compare with the observed z-score calculated from the aligned data. Significant electrodes had z-scores that fell within the tails of the null distribution (2-tail comparisons, $p<0.05$).

For these (and subsequent) analyses, we opted not to perform corrections for multiple comparisons across electrodes for each subject. Instead, because we used randomization procedures to estimate the false-positive rates, we assessed the reliability of our results by testing the proportions of significant electrodes across subjects directly against the false-positive rate (see Identifying significantly modulated brain regions below).

(2) Testing the sensitivity of ECoG signal to the temporal regularity of the sequence. Next, we tested whether the ECoG signals were sensitive to the 10-Hz repetition rate of the tone-burst sequence (Fig. 2.1a). To test this sensitivity, we first extracted the ECoG signals in response to each tone-burst sequence ($N = 10 \times$ number of sessions for each subject) and then performed a wavelet decomposition to extract the instantaneous power and phase as a function of time (wave number
29

6; 22-s ECoG buffers pre- and post-tone-burst sequence; 15 frequencies between 1–200 Hz, evenly spaced on a log₂ scale) (Besle et al., 2010; Lakatos et al., 2013). Next, we computed the sequence-aligned mean log power and pairwise phase consistency (PPC; a bias-free measure of phase alignment) (Vinck et al., 2010) across the tone-burst sequences and then a grand-mean across time. The PPC is defined as $PPC = -\frac{2}{N(N-1)} \sum_{j=1}^{N-1} \sum_{k=(j+1)}^{N} f(\theta_j, \theta_k)$, with $f(\varphi, \omega) = \cos(\varphi) \cos(\omega) + \sin(\varphi) \sin(\omega)$. In these equations, N is the number of instances (i.e., tone-burst sequences) and $\theta_j$ and $\theta_k$ are unique pairs of phase values sampled from the total population of instances. We started this analysis 500 ms after sequence onset (i.e., the first 5 tones) to minimize the effects of sequence onset on our measures of temporal regularity.

For each electrode, a randomization procedure tested the significance of these measurements of power and PPC. First, we extracted random 91.6-s segments of ECoG signal (corresponding to the sequence duration plus the 22-s buffers; 10 random segments per session completed for each subject) and calculated its random grand-mean power and PPC at each wavelet frequency in the exact same manner as described above. This randomization was repeated 1000 times to create null distributions of power and PPC values. Each random power and PPC spectrum was then compared to its respective null distribution to identify the longest length of contiguous frequency bands that exhibited significant enhancements or decrements compared to the distribution by chance (raw $p<0.05$). This produced a null distribution of lengths of contiguous frequency bands. The raw significance levels of the actual measurements of power and PPC (2-tail comparisons) were then calculated relative to their respective random distributions (raw $p<0.05$), and the actual lengths of significant contiguous frequency bands was determined. Reliable electrodes had lengths of significant activity (either power or PPC) in contiguous frequency bands that exceeded chance (corrected $p<0.05$).

(3) Testing the sensitivity of ECoG signal to the spectral regularity of the tone-burst sequence.

Finally, we characterized the sensitivity of the ECoG signal to spectral regularity by evaluating the
relationship between power or PPC and the spectral regularity of a subsequence’s local configuration of $F_1$ and $F_2$. We restricted this analysis to timescales between 200-700 ms (2-7 tones). The number of trials per condition depended on (1) the number of sessions by subject, (2) the number of instances of subsequences in each tone-burst sequence, and (3) the number of subsequence instances by subsequence length: at minimum, there were 20 trials per subsequence.

Further, we only analyzed the ECoG signals that were elicited by the last tone in each subsequence, corresponding to the time period of 0-100 ms following onset of the last tone burst. We chose this analysis approach in order to identify how ECoG signals were conditioned by the spectral regularity of the previous tone bursts; i.e., its local contextual ‘history’. We only analyzed activity during the 100-ms period following onset of the final tone burst in each subsequence in order to avoid any confounding effects due to activity related to the following tone bursts. Like above, we did not test the ECoG signals that were generated by the first 500 ms of every sequence (i.e., the first 5 tones) to minimize potential interactions between our spectral-regularity measures and onset of the stimulus sequence.

IDENTIFYING NEURAL OSCILLATORY FREQUENCIES THAT WERE MODULATED BY SPECTRAL REGULARITY. In order to identify the frequency bands that were modulated by spectral regularity, we compared the ECoG signal that was elicited by ‘regular’ subsequences in which all of the tone frequencies were the same (e.g., $F_1$—$F_1$—$F_1$ and $F_2$—$F_2$—$F_2$—$F_2$) to that elicited by one of two different classes of ‘irregular’ subsequences. The first class were those subsequences in which all of the tone frequencies were the same except for the last one (e.g., $F_2$—$F_2$—$F_2$—$F_1$ and $F_1$—$F_1$—$F_1$—$F_2$). The second class were those subsequences in which tone frequency alternated (e.g., $F_1$—$F_2$—$F_1$—$F_2$). A complete list of subsequences used in these analyses can be found in Figure 2.2.
Figure 2.2: Selected subsequences used to test spectral regularity. Left panel: selected regular subsequences that have identical tone-frequency progressions. Middle panel: selected irregular subsequences in all of the tone frequencies are the same except for the final tone. Right panel: selected irregular subsequences in which the tone progressions alternate between frequencies $F_1$ and $F_2$. Frequencies that differ from the regular subsequences are colored in red.

The first step of the analysis was to align the ECoG signals relative to tone-burst-sequence onset and perform a wavelet decomposition (wave number 6; 22-s ECoG buffers pre- and post-tone-burst sequence; 35 frequencies between 0.5–200 Hz, evenly spaced on a $\log_2$ scale) to yield instantaneous power and phase responses as a function of time. Next, as a function of subsequence length (2-7) and neural frequency, we computed each subsequence-aligned grand-mean power and PPC responses across all of the electrodes for each subject and tested (signed-rank tests) whether the regularity of the subsequences differentially modulated power or PPC; independent analyses were done for each of the two classes of irregular subsequences. Raw $p$-values were false-discovery-rate (FDR) corrected across subsequence length and frequency (Q=0.05) (Benjamini and Hochberg, 1995). We identified those frequency bands that were consistently modulated by spectral regularity across the majority ($\geq$4) of subsequence lengths.

**Band-specific analyses measuring modulations to spectral regularity.** Next, after identifying those frequency bands that were modulated by spectral regularity, we performed a more extensive analysis that utilized the entire data set. This analysis required a quantification of regularity without making assumptions about which subsequences were more ‘regular’ than others (as we did in the
prior analysis). We used a metric based on the Kolmogorov complexity (Kolmogorov, 1963; Lempel and Ziv, 1976; Kaspar and Schuster, 1987), $C_K$. Here, we define ‘regularity’ as $1/C_K$ (see Fig. 2.1c).

The Kolmogorov complexity is a measure of randomness that quantifies the extent to which a subsequence’s local configuration of $F_1$ and $F_2$ can be reduced to repeats of simpler (shorter) configurations. Consider three subsequences:

1. $F_1$–$F_1$–$F_1$–$F_1$–$F_1$–$F_1$–$F_1$
2. $F_1$–$F_2$–$F_1$–$F_2$–$F_1$–$F_2$–$F_1$
3. $F_1$–$F_2$–$F_1$–$F_1$–$F_2$–$F_2$–$F_1$

The first subsequence can be simplified to seven repeats of $F_1$ and has the highest $1/C_K$ (regularity) value (1.66). The second subsequence can also be simplified, but it has a slightly more complex pattern of repeating pairs of $F_1$–$F_2$. Consequently, it has a lower regularity value (1.0). In contrast, the third configuration cannot be simplified any further and, thus, has the lowest regularity value (0.55).

For practical purposes, we focused our subsequent analyses solely on subsequences that had lengths of 7 tone bursts to minimize redundancy in tests across different subsequence lengths and to maximize the number of unique $1/C_K$ values that we could evaluate, as the number of unique values scales linearly with subsequence length. However, similar results were found when we assessed ECoG responses to shorter subsequences. Similarly, because of our stimulus design, we did not have the statistical power to sample all possible subsequences for lengths > 7.

For each frequency band that was modulated by spectral regularity (either in phase or power; see section IDENTIFYING NEURAL OSCILLATORY FREQUENCIES THAT WERE MODULATED BY SPECTRAL REGULARITY), we correlated ECoG activity with a subsequence’s regularity value. We also conducted an analogous correlation using data from the HG band (70-200 Hz) due to its purported relationship with neural-spiking activity (Mukamel et al., 2005; Ray et al., 2008; Ray and Maunsell, 2011).

Instantaneous phase and amplitude responses in each frequency band as a function of time were computed by first band-pass filtering the ECoG signals and computing the Hilbert
transform (2-s ECoG buffers). To facilitate across-subject comparisons, we first computed the subsequence-aligned average PPC responses across electrodes for each subject. Next, we z-scored the averaged responses after applying the Fisher z-transform using the mean and standard deviation of the population of PPC responses to all local configurations of F₁ and F₂. Finally, across subjects, we tested the relationship between the z-scored PPC responses and spectral regularity with a Spearman correlation (ρ). An analogous analysis was conducted to test the correlation between spectral regularity and z-scored log amplitude. For amplitude, we did not apply the transformation prior to z-scoring.

To identify individual electrodes with PPC or amplitude responses that were significantly correlated with spectral regularity, we calculated the Spearman correlation between the (raw) PPC or amplitude values for each subsequence’s local configuration of F₁ and F₂ and their regularity values. We estimated the false-positive rate by computing a null distribution of 1000 Spearman-correlations by randomizing the relationship between response (PPC or amplitude) and local configuration. An electrode was ‘significant’ if the absolute value of its correlation was greater than random chance (false-positive rate=0.05).

**Identifying significantly modulated brain regions**

We used a “counts t-test” analysis (Ramayya et al., 2015) to test whether the proportions of modulated electrodes in a particular brain region was significantly greater than chance. For stimulus-onset and temporal-regularity analyses, we first converted the number of significant electrodes that were modulated by regularity into z-scores using a binomial null distribution. The null distribution was based on the total number of electrodes and a false-positive rate=0.05 (determined from the randomization analyses). We then tested whether the population of z-scores across subjects differed significantly from zero using a one-sampled t-test for each brain region. We corrected for multiple comparisons across brain regions using FDR correction (Q=0.05).
An analogous procedure determined which brain regions were modulated by spectral regularity. However, for this analysis, we computed z-scores separately for positively and negatively modulated electrodes and used a false-positive rate=0.025 for each modulation direction (total false-positive rate=0.05).

For all brain-region-specific analyses, electrodes were categorized into separate brain regions based on their associated anatomical labels (Table 2.1).

**Table 2.1: Regions of interest.** Anatomical labels used to define regions of interest.

<table>
<thead>
<tr>
<th>Lobe</th>
<th>Region of interest</th>
<th>Desikan-Killiany atlas labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td>Orbitofrontal</td>
<td>medialorbitofrontal, lateralorbitofrontal</td>
</tr>
<tr>
<td></td>
<td>DorsolateralPrefrontal</td>
<td>rostralmiddlefrontal, caudalmiddlefrontal</td>
</tr>
<tr>
<td></td>
<td>VentrolateralPrefrontal</td>
<td>parstriangularis, parsopercularis, parsorbitalis</td>
</tr>
<tr>
<td></td>
<td>AnteriorMedialFrontal</td>
<td>superiorfrontal, rostralanteriorcingulate, caudalanteriorcingulate</td>
</tr>
<tr>
<td></td>
<td>PosteriorMedialFrontal</td>
<td>paracentral, posteriorcingulate, isthmuscingulate</td>
</tr>
<tr>
<td>Frontal/Parietal</td>
<td>Sensorimotor</td>
<td>precentral, postcentral</td>
</tr>
<tr>
<td></td>
<td>SuperiorParietal</td>
<td>superiorparietal</td>
</tr>
<tr>
<td></td>
<td>InferiorParietal</td>
<td>inferiorparietal</td>
</tr>
<tr>
<td></td>
<td>Supramarginal</td>
<td>supramarginal</td>
</tr>
<tr>
<td>Occipital</td>
<td>Occipital</td>
<td>cuneus, lateraloccipital, lingual, pericalcarine</td>
</tr>
<tr>
<td>Temporal</td>
<td>SuperiorTemporal</td>
<td>superiortemporal</td>
</tr>
<tr>
<td></td>
<td>OtherTemporal</td>
<td>banksts, middletemporal, inferiortemporal, fusiform</td>
</tr>
<tr>
<td></td>
<td>MedialTemporallobe</td>
<td>entorhinal, parahippocampal; depth contacts labeled as hippocampal, entorhinal, perirhinal, or parahippocampal by neuroradiologist</td>
</tr>
</tbody>
</table>
RESULTS

We recorded ECoG activity from subdural surface and depth electrodes across the cortex while human subjects listened passively to a sequence of tone bursts. The frequency of the tone bursts was either 1000 Hz ($F_1$) or 1029 Hz ($F_2$; ½ semitone above $F_1$). Because the temporal progression of frequencies $F_1$ and $F_2$ in each sequence was stochastic, the local spectral regularity of the sequence changed dynamically over time.

We performed three sets of analyses to determine the extent to which neural activity throughout the cortex is sensitive to increasingly complex characteristics of auditory stimulation. First, we tested whether ECoG activity was modulated by the onsets of these auditory sequences, independent of its temporal and spectral structure. Second, we asked whether ECoG activity was modulated by the temporal regularity (i.e., the 10 Hz onset-to-onset interval) of these sequences, independent of its spectral structure. Finally, we asked whether ECoG activity was modulated by the spectral regularity (i.e., the local configuration of the $F_1$ and $F_2$ tone bursts).

The onset and temporal regularity of the tone-burst sequence modulates ECoG activity throughout the cortex

In our first set of analyses, we identified (1) those cortical regions that were modulated by the onsets of the tone-burst sequences; (2) cortical regions that were sensitive to particular tone-burst frequencies; and (3) cortical regions that were modulated by the temporal regularity of the sequence.

As expected, we found significant event-related potentials (ERPs) in the temporal lobe (Fig. 2.3a; see Fig. 2.3b for example ERPs). Additionally, we found significant ERPs throughout all regions of the cortex, including the parietal, frontal, and occipital cortices (Fig. 2.3a and 2.3b). ERPs from electrodes near the primary and secondary auditory cortices (in the posterior-superior temporal gyrus) generally had shorter latencies (i.e., the time to a significant response) than those
from other cortical regions (see Fig. 2.3a). ERPs from electrodes near these auditory cortices had an average latency of 0.13±0.029 s (mean ± standard error of the mean [SEM]), whereas those outside of these regions had significantly longer average latency (0.29±0.012 s; two-sample t-test t(156)=−2.5, p=0.014). These relatively long latencies were primarily the result of both our conservative method of measuring reliable ERP responses and the inclusion of electrodes from both primary and secondary auditory cortices. Indeed, the earliest reliable ERP responses occurred began around ~0.056 s, which is consistent with previous work (Edwards et al., 2005). With a less conservative significance criterion of p<0.05 for at least four contiguous time points (~0.016 s), electrodes near the auditory cortices had an average latency of 0.073 ± 0.014 s; whereas those from other cortical regions had an average latency of 0.18 ± 0.0072 s. This difference in latency was still significantly different (two-sample t-test t(156)=−2.6, p=0.011).

Figure 2.3: **Identification of electrodes with significant event-related potentials.** (a) Brain plots depict the locations of electrodes across subjects on an Average-Subject brain. Electrodes with significant ERP activity are color-coded based on the timing of the earliest significant activity. Non-significant electrodes are plotted in gray. (b) Example ERPs from different locations across cortex. Gray shading denotes standard error of the mean (SEM). Horizontal black bars above traces denote significant modulations from baseline. Green inset numbers at top-right of each panel correspond to numbered locations in (a).
A similar pattern of results was found for HG activity in response to sequence onsets (Fig. 2.4). Again, significant onset responses were observed across the entire cortex. Average HG latency near the primary and secondary auditory cortices was $0.11 \pm 0.031$ s, compared with $0.27 \pm 0.017$ s outside of this region (two-sample $t$-test $t(117)=-2.0$, $p=0.053$). With our less-conservative approach, the average latencies were $0.11 \pm 0.031$ s and $0.16 \pm 0.013$ s, respectively (two-sample $t$-test $t(117)=-0.69$, $p=0.49$).

Figure 2.4: Identification of electrodes with significant high-gamma (HG) activity to sequence onsets. (a) Brain plots depict the locations of electrodes across subjects on an Average-Subject brain. Electrodes with significant HG activity are color-coded based on the timing of the earliest significant activity. Non-significant electrodes are plotted in gray. (b) Example HG traces from different locations across cortex. Gray shading denotes standard error of the mean (SEM). Horizontal black bars above traces denote significant modulations from baseline activity. Green inset numbers at top-right of each panel correspond to numbered locations in (a).
To further quantify the distribution of significant modulations to sequence onsets across cortex, we conducted a counts t-test analysis (see METHODS). Consistent with the results from Figures 2.3 and 2.4, we found that both ERP responses and HG activity in all cortical lobes were reliably modulated by sequence onsets (Table 2.2).

Table 2.2: Summary statistics for event-related potentials and high-gamma activity to sequence onsets.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of Electrodes</th>
<th>Number of Subjects</th>
<th>Proportion of electrodes with reliable ERPs; counts t-test results</th>
<th>Proportion of electrodes with reliable evoked HG; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td>0.51; t(10) = 5.8, p = 9.0 x 10^-6</td>
<td>0.22; t(10) = 3.2, p = 0.0044</td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td>0.23; t(6) = 2.5, p = 0.025</td>
<td>0.20; t(6) = 2.0, p = 0.049</td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td>0.52; t(11) = 5.2, p = 1.0 x 10^-4</td>
<td>0.22; t(11) = 3.0, p = 0.0064</td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td>0.41; t(10) = 5.6, p = 1.1 x 10^-4</td>
<td>0.25; t(10) = 4.4, p = 6.2 x 10^-4</td>
</tr>
</tbody>
</table>

For each cortical region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of electrodes with reliable event-related potentials (ERPs; column 4) or with reliable HG activity (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate cortical regions that had onset-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05).

Next, we tested whether electrodes were sensitive to a particular acoustic frequency. Across cortex, we could not identify any differences in HG responses as a function of tone-burst frequency (signed-rank test, p=0.27). Similarly, this was also the case when we tested differences separately by cortical lobe (signed-rank tests, all ps>0.1). Finally, we could not identify a cortical lobe that had a reliable proportion of significantly modulated electrodes (Table 2.3).

Table 2.3: Summary statistics for high-gamma sensitivity to tone-burst frequency.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of Electrodes</th>
<th>Number of Subjects</th>
<th>Proportion of electrodes with reliable preference for F1; counts t-test results</th>
<th>Proportion of electrodes with reliable preference for F2; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td>0.020; t(10) = -0.64, p = 0.73</td>
<td>0.045; t(10) = 1.2, p = 0.12</td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td>0; t(6) = -6.7, p = 1.0</td>
<td>0.11; t(6) = 1.2, p = 0.14</td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td>0.056; t(11) = 0.85, p = 0.21</td>
<td>0.033; t(11) = 0.81, p = 0.22</td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td>0.024; t(10) = -0.11, p = 0.54</td>
<td>0.022; t(10) = -0.17, p = 0.56</td>
</tr>
</tbody>
</table>

For each cortical region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of electrodes with HG activity that was preferentially sensitive to frequency F1 (column 4) or frequency F2 (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate regions that showed onset-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05).
Finally, we tested the sensitivity of the ECoG signal to the 10-Hz temporal regularity of our tone-burst sequences. By aligning the ECoG signals in response to each tone-burst sequence, we could test whether ECoG activity was modulated by the temporal regularity of the sequences independent of the spectral regularity (which varied tone-to-tone and across sequences). We found that the ECoG signal of a significant number of electrodes had significant power modulations, relative to random ECoG activity (grand-mean proportion of significant electrodes across subjects = 0.36, false-positive rate = 0.05). These modulations were observed across a broad range of frequencies and across a broad region of cortex (Table 2.4, Fig. 2.5a, bottom and Fig. 2.6, right). The across-electrode averages identified significant decreases in both low-frequency (~1 Hz) and high-frequency (~90-130 Hz) power (Fig. 2.5b, bottom; signed-rank tests, FDR-corrected p<0.05).

Despite the fact that across-electrode averages did not identify significant enhancements of 10-Hz power (corresponding to the tone-burst-repetition rate), the distribution of significant electrodes with increases in power modulations did peak near 10 Hz. These positive modulations at 10-Hz power occurred most reliably in temporal cortex (counts t-test, t(10) = 3.6, FDR-corrected p < 0.05), with less reliable modulations occurring in frontal and parietal cortices (counts t-tests, ps < 0.05, uncorrected). We also found that a reliable population of electrodes in the temporal cortex had negative modulations in 10-Hz power (counts t-test, t(10) = 3.1, FDR-corrected p < 0.05), along with less reliable negative modulations in occipital cortex (counts t-test, p = 0.036, uncorrected).
Figure 2.5: Wide-band neural frequency response to tone-burst sequences. (a) Proportion of electrodes with significant modulations (increases in red, decreases in blue) in each frequency band. Dotted lines depict the false-positive rate=0.05 (0.025 for each direction of modulation) for randomization test. (b) pairwise phase consistency (PPC; top) and power (bottom) spectra for individual subjects (grey traces) in response to the entire tone-burst sequence. The thick black trace is the across-subject mean response. Data is z-scored for each electrode relative to its noise distribution. Asterisks identify frequencies with significant modulations across subjects (signed-rank tests, FDR-corrected p<0.05).

Figure 2.6: Identification of electrodes with modulations to temporal regularity. Brain plots depict the locations of electrodes across subjects on an Average-Subject brain. Electrodes with significant modulations in PPC (left) or power (right) color-coded based on the neural frequency band modulated and direction of modulation. Non-significant electrodes are plotted in gray.
Table 2.4: Summary statistics for power and phase-alignment modulations to temporal regularity.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of Electrodes</th>
<th>Number of Subjects</th>
<th>Proportion of electrodes with reliable positive modulations; counts t-test results</th>
<th>Proportion of electrodes with reliable negative modulations; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td>0.25; t(10) = 3.4, p = 0.0034</td>
<td>0.39; t(10) = 4.0, p = 0.0014</td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td>0.33; t(6) = 2.3, p = 0.031</td>
<td>0.39; t(6) = 2.9, p = 0.013</td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td>0.22; t(11) = 2.3, p = 0.020</td>
<td>0.38; t(11) = 4.3, p = 0.00068</td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td>0.25; t(10) = 4.4, p = 0.00067</td>
<td>0.33; t(10) = 4.4, p = 0.00065</td>
</tr>
<tr>
<td>Phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td>0.041; t(10) = −0.078, p = 0.53</td>
<td>0.0073; t(10) = −2.7, p = 0.99</td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td>0.014; t(6) = −0.76, p = 0.76</td>
<td>0; t(6) = −6.7, p = 1.0</td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td>0.038; t(11) = 0.034, p = 0.49</td>
<td>0.0025; t(11) = −5.9, p = 1.0</td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td>0.060; t(10) = 1.3, p = 0.11</td>
<td>0.023; t(10) = −1.1, p = 0.85</td>
</tr>
</tbody>
</table>

For each cortical region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of electrodes with reliable positive modulations (column 4) or with reliable negative modulations (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate regions that showed onset-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05). These analyses were done independently for both power and phase (PPC).

In contrast, PPC was not reliably modulated by the temporal regularity (Table 2.4, Fig. 2.5a and 2.5b, top and Fig. 2.6, left). We could not identify any significant PPC modulations in the average phase spectra across electrodes and subjects (Fig. 2.5b, top). This was largely due to the fact that (1) only a small proportion of electrodes was significantly modulated, and (2) we could not identify an individual frequency band that was significantly modulated (either increased or decreased) above the false-positive rate of 0.05.

Overall, this pattern of widespread cortical activation (in onset-induced ERP and HG activity, and in power-modulation to temporal regularity) is consistent with previous whole-brain ECoG studies (Besle et al., 2010; Burke et al., 2013; Ramayya et al., 2015). Because we found that sequence onset and temporal regularity modulated electrodes reliably across the cortex, our subsequent spectral-regularity analyses were conducted using our entire electrode dataset.

Phase—but not envelope—is modulated by local spectral regularity

In our critical set of analyses, we tested whether the power and phase alignment of the ECoG signal was modulated by spectral regularity. First, we compared ECoG activity in response to ‘regular’ subsequences with that elicited by two different classes of ‘irregular’ subsequences. The regular subsequence consisted a single-frequency tone progression (e.g., F₁—F₁—F₁—F₁).
The first class of irregular subsequences had the same frequency until the final tone (e.g., $F_1—F_1—F_1—F_2$). This comparison is somewhat analogous to those that have examined mismatch negativity, stimulus-specific adaptation, and other deviance-detection paradigms (Fuentemilla et al., 2006; Näätänen et al., 2007; Hsiao et al., 2009; Escera et al., 2013). The second class of irregular subsequences alternated on every tone burst (e.g., $F_1—F_2—F_1—F_2$).

The results from these two comparisons are shown in Figures 2.7 and 2.8, respectively. As a reminder, our goal was to test how ECoG activity in response to the last tone burst (highlighted in bold above) of a subsequence was modulated (conditioned) by the spectral regularity of the previous tone bursts.

In Figures 2.7a and 2.8a, we plot the differences in phase alignment between the regular subsequences (e.g., $F_1—F_1—F_1—F_1$) and the irregular subsequences (e.g., $F_1—F_1—F_1—F_2$ or $F_1—F_2—F_1—F_2$). These differences are plotted as a function of subsequence length (left panels) or combined across subsequence lengths (right panels). Analogous plots for differences in power are shown in Figures 2.7b and 2.8b.

We found that spectral regularity across all tested subsequence lengths (2-7) consistently modulated phase alignment in distinct frequency bands (signed-rank tests, $p<0.05$ for all, FDR-corrected across all tests). In the frequency band corresponding to the temporal regularity of the tone-burst sequences (10 Hz), phase alignment was modulated by spectral regularity (green-shaded regions in Figs. 2.7 and 2.8, top): phase alignment was greater for the regular subsequences than for the irregular subsequences. We also found that harmonic (20 Hz; red-shaded regions in Figs. 2.7a and 2.8a) and inharmonic frequencies (5 and 15 Hz; blue- and orange-shaded, respectively, regions in Figs. 2.7a and 2.8a) of this 10-Hz band were also modulated by spectral regularity. For the 20-Hz band, phase alignment was positively modulated by spectral regularity. But for the 5- and 15-Hz bands, it was negatively modulated. Finally, we also identified significant modulation in the delta-frequency band (<3 Hz, purple shaded regions in Figs. 2.7a and
2.8a), a frequency band that is not harmonically related to the sequence’s temporal regularity. In contrast, we could not identify any significant differences in the power spectra (Figs. 2.7b and 2.8b).

Figure 2.7: Wide-band frequency relationships between pairwise phase consistency or power and spectral regularity in deviant comparison. Difference in raw pairwise phase consistency (PPC) (a) and log-power (b) values between the regular and irregular sequences for each subsequence length (SL). In left panels of (a) and (b), black traces and shaded regions depict mean ± SEM across subjects, respectively. Horizontal black bars above subplots depict neural frequencies for which the difference value is significantly different from zero (signed-rank tests, all p < 0.05 with FDR correction). In right panels of (a) and (b), the across-subject mean spectra are plotted together (color coded by gray shading) for visual clarity. These color-shaded regions identify the significant neural-frequency bands used in subsequent analyses: purple (delta: <3 Hz), blue (5-Hz band: ~3-7 Hz), green (10-Hz band: ~8-12 Hz), orange (15-Hz band: ~13-17 Hz), and red (20-Hz band: ~18-22 Hz).
Single electrode ECoG and PPC traces are shown in Fig. 2.9. These traces were generated in response to the regular and irregular sequences. The example electrode in Figure 2.9a has greater 10-Hz phase alignment during the final tone (see gray shaded region in Fig. 2.9a, bottom) in response to the regular subsequences (red shaded traces in Fig 2.9a, bottom panel) than for the irregular subsequences (blue shaded traces in Fig 2.9a, bottom panel). Similarly, the example electrode in Figure 2.9b has greater delta-band phase alignment in response to the regular subsequences (compare red and blue traces in gray shaded region in Fig. 2.9b, bottom panel). The overall time course of the pairwise-phase-consistency fluctuations with respect to each subsequence was highly variable across individual electrodes. However, the modulation in phase...
alignment during the final tone-burst, as a function of spectral regularity, was consistent with the averaged electrode responses (Figs. 2.7a and 2.8a).

![Figure 2.9](image)

**Figure 2.9:** Example electrode traces aligned to subsequences with different local configurations of $F_1$ and $F_2$. (a) ECoG (top) and PPC (bottom) traces from an example electrode with significant modulations in 10-Hz PPC with respect to spectral regularity. Black solid vertical lines denote tone-burst onsets of aligned subsequences with the same local configuration. Black dashed vertical lines denote tone-burst onsets preceding and trailing tone-bursts in the aligned subsequence. ECoG activity is z-scored by the preceding 1-s of baseline activity prior to subsequence onset. Colors of traces denote different local configurations of $F_1$ and $F_2$. Red-shaded traces reflect more regular configurations. Blue shaded traces reflect more irregular configurations. Gray shaded regions depict where comparisons of PPC measurements were made. (b) Analogous plots as in (a), but for a second electrode with significant modulations in delta-band PPC with respect to spectral regularity. In each panel, the green circle in the brain indicates the electrode location.

**Phase alignment correlates with the degree of spectral regularity**

Together, these results strongly support a role for phase alignment—but not power—in the encoding of acoustic spectral regularity. However, because we focused only on a small subset of all subsequences (e.g., $F_1$—$F_1$—$F_1$—$F_1$ versus $F_1$—$F_1$—$F_1$—$F_2$), we could not fully assess the extent to which phase alignment and power were modulated by spectral regularity. To this end, we used the inverse of the Kolmogorov complexity ($1/C_K$; see METHODS) to quantify the spectral regularity of a subsequence. With this metric, we tested how phase alignment and amplitude were
correlated with the degree of spectral regularity. We focused solely on subsequences of length=7 to minimize redundancy in tests across shorter subsequence lengths and to maximize the number of unique $1/C_K$ values, which scales linearly with subsequence length. However, a similar pattern of results was found for shorter subsequence lengths (data not shown). Further, we restricted this analysis to those frequency bands that we identified previously as being modulated by spectral regularity (i.e., delta, 5 Hz, 10 Hz, 15 Hz, and 20 Hz; see Figs. 2.7 and 2.8).

The results of this analysis are shown in Fig. 2.10. In this Figure, the top two panels plot PPC, normalized for each subject across responses to each local configuration, as a function of regularity ($1/C_K$). Consistent with our previous findings (see Figs. 2.7 and 2.8), PPC was positively correlated with regularity across subjects in the delta, 10-Hz, and 20-Hz frequency bands (Fig. 2.10a; Spearman $\rho=0.47$, 0.28, and 0.34, respectively; all $p<0.05$ with Holm-Bonferroni correction) and negatively correlated with regularity in the 5- and 15-Hz frequency bands (Figs. 2.10b; Spearman $\rho=–0.32$ and –0.52, respectively; Holm-Bonferroni-corrected $p<0.05$ for each). These populations trends were also generally evident at the level of individual subjects (insets in Fig. 2.10a and b): the median individual-subject correlations in the delta, and 10-Hz frequency bands were significantly greater than zero (signed rank tests: $p<0.05$ for each with Holm-Bonferroni correction), whereas the median correlations in the 5- and 15-Hz frequency bands were significantly less than 0 (signed rank tests: $p<0.05$ for each with Holm-Bonferroni correction). We could not identify any differences in the strengths of the correlations among the positively correlated or negatively correlated frequency bands (randomization tests: $p>0.05$). The apparent differences in phase-alignment between the subsequences with the largest regularity values (e.g., compare green data points for $1/C_K= 1.662$ in Fig. 2.10) were not reliable across subjects (paired t-tests, all $p<0.05$ with Holm-Bonferroni correction).
Figure 2.10: Relationship between spectral regularity and pairwise phase consistency. Z-scored PPC as a function of regularity for subsequence length=7. Panels (a) and (b) depict frequency bands with positive and negative correlations with spectral regularity, respectively. Unfilled data points reflect mean responses to each individual local configuration of F1 and F2 across electrodes and subjects. Filled data points with error bars depict across-subject mean and SEM responses to all local configurations with the same spectral regularity value. Filled data points are connected to highlight trends. Insets depict individual Spearman correlation values for each subject, separately for each frequency band. Color conventions follow from Figs. 2.7 and 2.8. Asterisks above data points in insets denote frequency bands with significant individual-subject correlations (signed-rank tests: p<0.05 with Holm-Bonferroni correction).

However, inspection of Fig. 2.10 demonstrates that the largest PPC values occurred at the most regular subsequences (i.e., largest 1/Ck values). This suggests that the significance of the Spearman correlation value might have been driven primarily by these values. Indeed, for the
frequency bands above the delta band, if we exclude the pairwise-phase-consistency values generated from the most regular configurations, the across-subject correlation values were generally not significant (p>0.05 for all except for the 15-Hz frequency band; Spearman ρ=−0.25; p=0.037). For the delta-band data, when we removed these values, the correlation remained highly significant (Spearman ρ=0.37; p=0.0014). Similarly, the correlation in the delta band remained significant when we removed the least regular configurations from the analysis (Spearman ρ=0.53, p<0.001). The correlation in the delta band trended toward significance when the subsequences with the two largest regularity values were removed from the analysis (Spearman ρ=0.21; p=0.11).

However, when computing the correlation without averaging responses as a function of regularity, the correlation remained significant even after removing these subsequences (Spearman ρ=0.056; p=0.031). Finally, the correlations remained significant after removing both the most-regular and least-regular configurations, whether computing the correlation of the averaged responses as a function of regularity (Spearman ρ=0.46, p<0.001) or computing the correlation of the individual responses to each configuration (Spearman ρ=0.056, p=0.031).

We also tested whether our PPC findings could be attributed simply to the number of F₁ and F₂ tone-bursts in a subsequence, regardless of its spectral regularity. To do this, we repeated the correlation analyses for the delta band after sorting PPC responses by the proportion of tones of a single frequency in the local configurations instead of the regularity metric 1/Cₖ. For this analysis, we collapsed PPC responses across opposite tone-burst-frequency progressions that had the same single-frequency proportions (e.g., 7•{F₁} and 7•{F₂} have the same proportion 7/7 = 1). We found that PPC responses were positively correlated with single-frequency proportion (Spearman ρ=0.53, p<0.001). However, this is not surprising considering that 1/Cₖ was also correlated with single-frequency proportion (Spearman ρ=0.41, p<0.001).

In contrast to modulations in phase alignment, we could not identify any reliable correlations between amplitude modulations and spectral regularity, either in the population trends.
or at the level of individual subjects (all $p>0.05$ after Holm-Bonferroni correction; Fig. 2.11a and 2.11b).

**Figure 2.11: Relationship between spectral regularity and amplitude.** Z-scored log-amplitude as a function of regularity for subsequence length=7. Data in (a) follow same format as those in the main panels in Fig. 2.10, with the addition of the high-gamma responses in gray. Data in (b) follow the same format as those in the insets in Fig. 2.10, with the addition of the high-gamma responses in gray.

**Phase alignment reflects spectral regularity in multiple brain regions**

We found neural correlates of spectral regularity in the whole-brain averages of ECoG PPC responses. These results could have been due to consistent and widespread activity across the entire brain or the result of strong responses that originated from specific cortical locations. To differentiate between these two possibilities, we conducted single-electrode analyses to identify regions of cortex that were reliably modulated by spectral regularity. We computed the Spearman correlation for each electrode, this time using the raw PPC values. With these single-electrode measures, we conducted a counts t-test analysis (see METHODS) to localize significant effects across cortex; the full results of this analysis are listed in Table 2.5.
Table 2.5: Summary correlation statistics for pairwise phase consistency in band-specific analyses.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of Electrodes</th>
<th>Number of Subjects</th>
<th>Proportion of positively modulated electrodes; counts t-test results</th>
<th>Proportion of negatively modulated electrodes; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td><strong>0.066; t(10) = 3.5, p = 0.0029</strong></td>
<td><strong>0.027; t(10) = 1.1, p = 0.15</strong></td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td><strong>0.043; t(6) = 1.3, p = 0.13</strong></td>
<td><strong>0.043; t(6) = 1.3, p = 0.13</strong></td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td><strong>0.091; t(11) = 3.0, p = 0.0063</strong></td>
<td><strong>0.054; t(11) = 2.6, p = 0.012</strong></td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td><strong>0.037; t(10) = 3.1, p = 0.0053</strong></td>
<td><strong>0.072; t(10) = 3.3, p = 0.0040</strong></td>
</tr>
<tr>
<td><strong>5 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td><strong>0.024; t(10) = 1.2, p = 0.14</strong></td>
<td><strong>0.025; t(10) = 0.73, p = 0.24</strong></td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td><strong>0.036; t(6) = -0.12, p = 0.55</strong></td>
<td><strong>0.091; t(6) = 1.6, p = 0.075</strong></td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td><strong>0.026; t(11) = 0.71, p = 0.25</strong></td>
<td><strong>0.054; t(11) = 1.9, p = 0.046</strong></td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td><strong>0.049; t(10) = 1.9, p = 0.041</strong></td>
<td><strong>0.057; t(10) = 1.8, p = 0.047</strong></td>
</tr>
<tr>
<td><strong>10 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td><strong>0.029; t(10) = 0.79, p = 0.23</strong></td>
<td><strong>0.029; t(10) = 0.51, p = 0.31</strong></td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td><strong>0.059; t(6) = 1.3, p = 0.13</strong></td>
<td><strong>0.038; t(6) = 0.94, p = 0.19</strong></td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td><strong>0.034; t(11) = 1.3, p = 0.12</strong></td>
<td><strong>0.045; t(11) = 0.63, p = 0.27</strong></td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td><strong>0.093; t(10) = 4.6, p = 0.00038</strong></td>
<td><strong>0.030; t(10) = 0.31, p = 0.38</strong></td>
</tr>
<tr>
<td><strong>15 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td><strong>0.023; t(10) = 0.15, p = 0.44</strong></td>
<td><strong>0.048; t(10) = 2.4, p = 0.019</strong></td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td>0; t(6) = -0.9, p = 1.0</td>
<td>0.11; t(6) = 2.5, p = 0.022</td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td><strong>0.015; t(11) = -1.1, p = 0.85</strong></td>
<td><strong>0.022; t(11) = -0.067, p = 0.53</strong></td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td><strong>0.030; t(10) = 0.65, p = 0.26</strong></td>
<td><strong>0.044; t(10) = 3.1, p = 0.00060</strong></td>
</tr>
<tr>
<td><strong>20 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal Cortex</td>
<td>344</td>
<td>11</td>
<td><strong>0.037; t(10) = 0.92, p = 0.19</strong></td>
<td><strong>0.026; t(10) = 0.47, p = 0.32</strong></td>
</tr>
<tr>
<td>Occipital Cortex</td>
<td>83</td>
<td>7</td>
<td><strong>0.031; t(6) = 0.61, p = 0.29</strong></td>
<td><strong>0.019; t(6) = -0.21, p = 0.58</strong></td>
</tr>
<tr>
<td>Parietal Cortex</td>
<td>266</td>
<td>12</td>
<td><strong>0.067; t(11) = 2.4, p = 0.019</strong></td>
<td><strong>0.065; t(11) = 1.4, p = 0.069</strong></td>
</tr>
<tr>
<td>Temporal Cortex</td>
<td>484</td>
<td>11</td>
<td><strong>0.047; t(10) = 1.9, p = 0.046</strong></td>
<td><strong>0.029; t(10) = 0.21, p = 0.42</strong></td>
</tr>
</tbody>
</table>

For each brain region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of positively-modulated electrodes (column 4), and proportion of negatively-modulated electrodes (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate regions that showed regularity-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05).

In the delta-frequency band, we found that the frontal, temporal, and parietal cortices had significant proportions of electrodes in which the correlation between spectral regularity and PPC was reliably positive (counts t-tests; FDR-corrected p<0.05 for all). Additionally, the temporal cortex had a significant proportion of electrodes that negatively tracked spectral regularity (p=0.0040). In the 10-Hz band and the 15-Hz band, the temporal cortex had a significant proportion of electrodes that positively (p=0.0038) and negatively (p=0.0060), respectively, correlated with spectral regularity. Thus, although spectral-regularity representation is distributed across cortex, activity in the temporal cortex appeared to have a more predominant role in tracking spectral regularity.

To better localize these effects, we repeated our counts t-test with a finer-grained regional analysis for the subset of frequency bands that were reliably modulated in at least one cortical lobe (Table 2.6). We found that reliable positive modulations in the delta-band were present in the
dorsolateral prefrontal cortex, inferior parietal cortex, and “OtherTemporal” cortex (counts t-tests, FDR-corrected ps<0.05 with Q=0.06). In the 10-Hz band, we found that only “OtherTemporal” was reliably modulated by spectral regularity (counts t-tests, FDR-corrected ps<0.05 with Q=0.06).

Table 2.6: Summary correlation statistics for pairwise phase consistency in band-specific analyses.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of electrodes</th>
<th>Number of subjects</th>
<th>Proportion positively modulated electrodes; counts t-test results</th>
<th>Proportion negatively modulated electrodes; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AnteriorMedialFrontal</td>
<td>73</td>
<td>6</td>
<td>0.024; t(5) = −0.22, p = 0.60</td>
<td>0.066; t(5) = 1.4, p = 0.11</td>
</tr>
<tr>
<td>DorsolateralPrefrontal</td>
<td>133</td>
<td>9</td>
<td>0.15; t(8) = 3.9, p = 0.0024</td>
<td>0.027; t(8) = 0.83, p = 0.22</td>
</tr>
<tr>
<td>InferiorParietal</td>
<td>153</td>
<td>10</td>
<td>0.13; t(9) = 3.2, p = 0.0057</td>
<td>0.065; t(9) = 2.1, p = 0.032</td>
</tr>
<tr>
<td>MedialTemporalLobe</td>
<td>36</td>
<td>6</td>
<td>0.055; t(5) = 0.85, p = 0.22</td>
<td>0.042; t(5) = 0.86, p = 0.22</td>
</tr>
<tr>
<td>Orbitofrontal</td>
<td>25</td>
<td>6</td>
<td>0.074; t(5) = 1.1, p = 0.17</td>
<td>0.019; t(5) = 0.15, p = 0.44</td>
</tr>
<tr>
<td>OtherTemporal</td>
<td>136</td>
<td>11</td>
<td><strong>0.097; t(10) = 3.0, p = 0.0063</strong></td>
<td>0.060; t(10) = 1.8, p = 0.052</td>
</tr>
<tr>
<td>Sensorimotor</td>
<td>49</td>
<td>9</td>
<td>0.055; t(8) = 1.4, p = 0.094</td>
<td>0.059; t(8) = 0.67, p = 0.26</td>
</tr>
<tr>
<td>SuperiorParietal</td>
<td>121</td>
<td>11</td>
<td>0.17; t(10) = 2.5, p = 0.018</td>
<td>0.053; t(10) = 0.86, p = 0.21</td>
</tr>
<tr>
<td>SuperiorTemporal</td>
<td>82</td>
<td>11</td>
<td>0.14; t(10) = 2.1, p = 0.034</td>
<td>0.063; t(10) = 2.5, p = 0.015</td>
</tr>
<tr>
<td>Supramarginal</td>
<td>266</td>
<td>11</td>
<td>0.028; t(10) = 0.68, p = 0.26</td>
<td>0.087; t(10) = 1.7, p = 0.064</td>
</tr>
<tr>
<td>VentrolateralPrefrontal</td>
<td>44</td>
<td>8</td>
<td>0.042; t(7) = 0.22, p = 0.42</td>
<td>0.039; t(7) = 0.67, p = 0.26</td>
</tr>
<tr>
<td><strong>10 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MedialTemporalLobe</td>
<td>36</td>
<td>6</td>
<td>0.13; t(5) = 1.5, p = 0.10</td>
<td>0.034; t(5) = 0.79, p = 0.23</td>
</tr>
<tr>
<td>OtherTemporal</td>
<td>136</td>
<td>11</td>
<td><strong>0.092; t(10) = 4.3, p = 7.3 × 10^{-4}</strong></td>
<td>0.034; t(10) = 0.15, p = 0.44</td>
</tr>
<tr>
<td>SuperiorTemporal</td>
<td>82</td>
<td>11</td>
<td>0.17; t(10) = 2.8, p = 0.010</td>
<td>0.062; t(10) = 1.0, p = 0.16</td>
</tr>
<tr>
<td><strong>15 Hz</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MedialTemporalLobe</td>
<td>36</td>
<td>6</td>
<td>0; t(5) = −5.7, p = 1.0</td>
<td>0.034; t(5) = 0.79, p = 0.23</td>
</tr>
<tr>
<td>OtherTemporal</td>
<td>136</td>
<td>11</td>
<td>0.029; t(10) = 0.39, p = 0.35</td>
<td>0.036; t(10) = 1.5, p = 0.088</td>
</tr>
<tr>
<td>SuperiorTemporal</td>
<td>82</td>
<td>11</td>
<td>0.25; t(10) = 1.1, p = 0.15</td>
<td>0.10; t(10) = 1.6, p = 0.070</td>
</tr>
</tbody>
</table>

For each brain region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of positively-modulated electrodes (column 4), and proportion of negatively-modulated electrodes (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate regions that showed regularly-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05, with Q=0.06). Italicized text in columns 4 and 5 indicate regions that showed significance with a less stringent criterion (FDR-corrected p<0.05, with Q=0.1).

In comparison, we did not identify a consistent relationship between amplitude modulations and spectral regularity in the whole-brain averaged responses. Once again, this simply could be due to the fact that the brain regions did not have any reliable modulations. Alternatively, it is possible that, within a brain region, different electrodes had both positive and negative modulations, which effectively canceled out upon averaging. We found that, although we could identify brain regions with reliable modulations in amplitude, we could not identify any consistent relationships between reliable amplitude modulations and spectral regularity in any frequency band or in any brain region (Table 2.7). This is despite the fact that the proportions of electrodes with significant
power modulations were generally larger than the proportions of electrodes with significant pairwise-phase-consistency modulations (compare Tables 2.5 and 2.7). Together, this suggests that our inability to identify correlations between amplitude and spectral regularity in the whole-brain responses (see Fig. 2.11) was largely due to the fact that there was not any consistent relationship between amplitude responses and regularity in any brain region.

Table 2.7: Summary correlation statistics for amplitude in band-specific analyses.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of electrodes</th>
<th>Number of subjects</th>
<th>Proportion positively modulated electrodes; counts t-test results</th>
<th>Proportion negatively modulated electrodes; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AnteriorMedialFrontal</td>
<td>73</td>
<td>6</td>
<td>0.20; t(5) = 2.3, p = 0.036</td>
<td>0.071; t(5) = 2.0, p = 0.049</td>
</tr>
<tr>
<td>DorsolateralPrefrontal</td>
<td>133</td>
<td>9</td>
<td>0.094; t(8) = 2.1, p = 0.037</td>
<td>0.10; t(8) = 1.7, p = 0.062</td>
</tr>
<tr>
<td>InferiorParietal</td>
<td>153</td>
<td>10</td>
<td>0.088; t(9) = 1.8, p = 0.051</td>
<td>0.066; t(9) = 2.5, p = 0.017</td>
</tr>
<tr>
<td>MedialTemporalLobe</td>
<td>36</td>
<td>6</td>
<td>0.21; t(5) = 1.8, p = 0.063</td>
<td>0.32; t(5) = 0.14, p = 0.35</td>
</tr>
<tr>
<td>Orbitalfrontal</td>
<td>25</td>
<td>6</td>
<td>0.13; t(5) = 1.8, p = 0.069</td>
<td>0.32; t(5) = 2.6, p = 0.026</td>
</tr>
<tr>
<td>OtherTemporal</td>
<td>136</td>
<td>11</td>
<td><strong>0.13; t(10) = 3.6, p = 0.0025</strong></td>
<td><strong>0.05; t(10) = 3.6, p = 0.0025</strong></td>
</tr>
<tr>
<td>Sensorimotor</td>
<td>49</td>
<td>9</td>
<td><strong>0.30; t(8) = 5.7, p = 1.0 × 10</strong></td>
<td><strong>0.10; t(8) = 2.5, p = 0.016</strong></td>
</tr>
<tr>
<td>SuperiorParietal</td>
<td>121</td>
<td>11</td>
<td>0.10; t(10) = 1.8, p = 0.052</td>
<td>0.30; t(10) = 0.47, p = 0.33</td>
</tr>
<tr>
<td>SuperiorTemporal</td>
<td>82</td>
<td>11</td>
<td>0.075; t(10) = 1.9, p = 0.047</td>
<td>0.12; t(10) = 2.0, p = 0.039</td>
</tr>
<tr>
<td>Supramarginal</td>
<td>266</td>
<td>11</td>
<td>0.11; t(10) = 1.7, p = 0.058</td>
<td>0.079; t(10) = 1.6, p = 0.066</td>
</tr>
<tr>
<td>VentrolateralPrefrontal</td>
<td>44</td>
<td>8</td>
<td>0.078; t(7) = 1.3, p = 0.11</td>
<td>0.23; t(7) = 2.1, p = 0.038</td>
</tr>
</tbody>
</table>

| 10 Hz                      |                      |                    |                                                                 |                                                                 |
| InferiorParietal           | 153                  | 10                 | 0.13; t(9) = 1.8, p = 0.057                                     | 0.047; t(9) = 1.3, p = 0.12                                     |
| MedialTemporalLobe         | 36                   | 6                  | 0.034; t(5) = 0.79, p = 0.23                                     | 0.047; t(5) = 1.0, p = 0.18                                     |
| OtherTemporal              | 136                  | 11                 | 0.084; t(10) = 3.0, p = 0.0072                                   | 0.048; t(10) = 1.8, p = 0.053                                   |
| Sensorimotor               | 49                   | 9                  | 0.11; t(8) = 1.2, p = 0.11                                     | 0.11; t(8) = 2.0, p = 0.037                                     |
| SuperiorParietal           | 121                  | 11                 | 0.064; t(10) = 1.2, p = 0.14                                     | 0.16; t(10) = 2.3, p = 0.025                                     |
| SuperiorTemporal           | 82                   | 11                 | 0.018; t(10) = −0.15, p = 0.56                                   | 0.13; t(10) = 1.7, p = 0.057                                   |
| Supramarginal              | 266                  | 11                 | 0.19; t(10) = 2.5, p = 0.015                                     | 0.0070; t(10) = −1.6, p = 0.93                                   |

| 15 Hz                      |                      |                    |                                                                 |                                                                 |
| InferiorParietal           | 153                  | 10                 | 0.11; t(9) = 1.8, p = 0.042                                     | 0.013; t(9) = −1.4, p = 0.90                                     |
| MedialTemporalLobe         | 36                   | 6                  | 0.096; t(5) = 1.6, p = 0.087                                     | 0.10; t(5) = 0.97, p = 0.19                                     |
| OtherTemporal              | 136                  | 11                 | 0.053; t(10) = 2.7, p = 0.011                                   | 0.062; t(10) = 2.4, p = 0.019                                   |
| Sensorimotor               | 49                   | 9                  | 0.081; t(8) = 1.2, p = 0.13                                     | 0.077; t(8) = 1.9, p = 0.043                                   |
| SuperiorParietal           | 121                  | 11                 | 0.060; t(10) = 0.97, p = 0.18                                   | 0.11; t(10) = 2.4, p = 0.007                                   |
| SuperiorTemporal           | 82                   | 11                 | 0.074; t(10) = 1.1, p = 0.15                                     | 0.037; t(10) = 1.6, p = 0.070                                   |
| Supramarginal              | 266                  | 11                 | 0.024; t(10) = 0.46, p = 0.33                                   | 0.052; t(10) = 0.62, p = 0.22                                   |

| 20 Hz                      |                      |                    |                                                                 |                                                                 |
| InferiorParietal           | 153                  | 10                 | 0.11; t(9) = 1.8, p = 0.042                                     | 0.013; t(9) = −1.4, p = 0.90                                     |
| MedialTemporalLobe         | 36                   | 6                  | 0.096; t(5) = 1.6, p = 0.087                                     | 0.10; t(5) = 0.97, p = 0.19                                     |
| OtherTemporal              | 136                  | 11                 | 0.053; t(10) = 2.7, p = 0.011                                   | 0.062; t(10) = 2.4, p = 0.019                                   |
| Sensorimotor               | 49                   | 9                  | 0.081; t(8) = 1.2, p = 0.13                                     | 0.077; t(8) = 1.9, p = 0.043                                   |
| SuperiorParietal           | 121                  | 11                 | 0.060; t(10) = 0.97, p = 0.18                                   | 0.11; t(10) = 2.4, p = 0.007                                   |
| SuperiorTemporal           | 82                   | 11                 | 0.074; t(10) = 1.1, p = 0.15                                     | 0.037; t(10) = 1.6, p = 0.070                                   |
| Supramarginal              | 266                  | 11                 | 0.024; t(10) = 0.46, p = 0.33                                   | 0.052; t(10) = 0.62, p = 0.22                                   |

| High Gamma                 |                      |                    |                                                                 |                                                                 |
| MedialTemporalLobe         | 36                   | 6                  | 0; t(5) = −0.57, p = 1.0                                      | 0.021; t(5) = 0.034, p = 0.49                                     |
| OtherTemporal              | 82                   | 11                 | 0.002; t(10) = 2.5, p = 0.016                                   | 0.060; t(10) = 1.1, p = 0.15                                   |
| SuperiorTemporal           | 136                  | 11                 | 0.048; t(10) = 1.3, p = 0.10                                   | 0.032; t(10) = 1.1, p = 0.15                                   |

For each region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of positively-modulated electrodes (column 4), and proportion of negatively-modulated electrodes (column 5). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold text in columns 4 and 5 indicate regions that showed regularly-modulated electrodes more frequently than expected by chance (FDR-corrected p<0.05).
Figure 2.12 plots the reliability of modulated ECoG activity across cortex as a function of our ERP, HG, temporal-regularity, and spectral-regularity analyses. It is clear that simpler components of the tone-burst sequence (i.e., stimulus onset and temporal regularity) elicit the most reliable responses in each brain region. In contrast, phase sensitivity to spectral regularity was more limited. Reliable phase modulations in the delta-frequency band tended to be more prevalent than those of the other frequency bands in frontal, parietal, and temporal cortices.

**Figure 2.12: Electrode sensitivity to acoustic information by brain region and analysis.** Individual bars denote mean (± SEM) proportion of significant electrodes for each analysis in each brain region. Colors reflect the various analyses conducted. Only electrodes with reliable modulations in power are shown for the temporal-regularity analysis, and only electrodes with reliable modulations in PPC are shown for spectral-regularity analyses.
DISCUSSION

Acoustic stimulation elicits widespread brain activation

Our ERP and temporal-regularity analyses indicated that electrodes in each cortical lobe responded to some component of the tone-burst sequences (see Fig. 2.12). It is perhaps not surprising that acoustic stimulation modulates temporal, frontal, and parietal, consistent with multiple lines of previous work (Edwards et al., 2005; 2009; Hsiao et al., 2009; Ishii et al., 2009; Besle et al., 2010; Chennu et al., 2013; Golumbic et al., 2013; Eliades et al., 2014). However, it is interesting to note that even in the occipital lobe, we found reliable auditory-elicited activity (McDonald et al., 2013; Mercier et al., 2013; Brang et al., 2015). This is not to suggest that occipital regions play a necessary or even a supplementary role in processing the acoustic information per se. Instead, these widespread neural modulations in response to new incoming sensory information may be a mechanism by which attention can be redirected to novel events in the environment (Schröger, 1996; Schröger et al., 2000; Parmentier et al., 2008) and/or to facilitate the coupling of multisensory representations; see (Kayser and Logothetis, 2007) and (Foxe and Schroeder, 2005). The fact that a much more limited (though still reliable) proportion of electrodes exhibited modulations to spectral regularity (see Fig. 2.12) suggests that sub-populations of neurons in each cortical lobe may contribute to regularity representations.

Our findings that the temporal cortex, along with the frontal and parietal cortices, track (spectral) regularity is consistent with previous work (Patel and Balaban, 2000; Dimitrijevic et al., 2001; Doeller et al., 2003; Lakatos et al., 2005; Luo et al., 2006; Besle et al., 2010; Hsiao et al., 2010; Garrido et al., 2013; Lakatos et al., 2013; Lappe et al., 2013). Specifically, damage to the dorsolateral prefrontal (dPFC) and parietal cortices reduces the amplitude of the mismatch negativity (MMN), an automatic brain response reflecting a detected stimulus change from a commonly presented stimulus (Alho et al., 1994; Alain et al., 1998). Because the commonly
presented stimulus creates a spectrotemporal regularity, it is unsurprising that these regions involved in change detection are also modulated by spectral regularity. Additionally, a growing body of literature suggests that regions of the inferior parietal cortex, in particular, are critically associated with spectrotemporal processing and perceptual organization (Giraud et al., 2000; Cusack, 2005; Obleser et al., 2007; Dykstra et al., 2011; Rauschecker, 2011; Teki et al., 2011; Bornkessel-Schlesewsky et al., 2015; Teki et al., 2016), which would necessarily require them to process information related to the regularities in acoustic stimuli. The current study expands this previous work by showing that these brain regions may contribute more generally to spectral regularity than previously surmised (Doeller et al., 2003; Hsiao et al., 2010; Garrido et al., 2013; Lappe et al., 2013). These wide-spread cortical responses indicate that regularity identification and representation is fundamentally important across cortex in sensory processing specifically (Ulanovsky et al., 2004; Turk-Browne et al., 2008; Winkler et al., 2009; Schapiro et al., 2012) and in cortical processing more generally through oscillatory coherence (Singer, 1999; Buzsáki and Draguhn, 2004; Fries, 2015), whereby distributed networks of cortical processing become functionally linked by sharing a common temporal regularity in their oscillatory behavior.

**Spectral regularity is represented in phase alignment**

The primary finding in this study was that changes in spectral regularity were reflected only in the degree of phase alignment of ECoG activity (see Figs. 2.7-2.11). In contrast, power was not preferentially positively or negatively correlated with spectral regularity in any neural frequency band in any cortical lobe. One interpretation of these findings is that increases in spectral regularity systematically affects the tendency of endogenous cortical oscillations to align to the tone bursts (Klimesch et al., 2007; Schroeder and Lakatos, 2009; Lakatos et al., 2013). Alternatively, our phase results could reflect non-oscillatory origins related to stimulus-evoked activity that itself is aligned to the tone bursts (Mäkinen et al., 2005). In favor the phase-alignment interpretation, power in the delta-frequency band tended to be negatively modulated by the tone-burst sequences when
compared to random ECoG segments (see Fig. 2.5), which would not be expected if the tone-bursts induced evoked delta-frequency activity. Moreover, if spectral regularity affected the degree of tone-evoked activity, we would have identified a systematic relationship between amplitude and spectral regularity. However, we could not identify such a relationship (see Fig. 2.11).

It is still possible that, whereas on average our results suggest an oscillatory component, individual electrodes may exhibit oscillatory or evoked-type responses that are modulated by the spectral regularity. Indeed, some brain regions exhibited significant proportions of electrodes with reliable correlations between amplitude and spectral regularity (see Table 2.7). Altogether, it is likely that both evoked-type and oscillatory activity are required to fully explain the present results (Ding et al., 2016). Further analyses will be required to fully elucidate the differential contributions of each type of activity to spectral-regularity representation, which will require special models and analytical techniques to distinguish between the two alternatives (Truccolo et al., 2002; Luzhou Xu et al., 2009).

**Regularity representations or neural adaptation**

We have interpreted our results as evidence of neural representations of spectral regularity. However, because our stimuli used only two different frequency tone bursts, an alternate interpretation is that our findings do not reflect regularity but, instead, reflect the effects of neural adaptation on short time scales (Fishman et al., 2001; 2004; Ulanovsky et al., 2004; Micheyl et al., 2005; Eliades et al., 2014). Indeed, there is considerable debate as to the extent to which other common neural signatures of regularity (e.g., mismatch negativity) reflect mechanisms of neural adaptation versus true regularity representations *per se* (Fishman and Steinschneider, 2012; Fishman, 2013). The degree of neural adaptation is modulated by regularity (Todorovic et al., 2011), suggesting that, at worse, our results still partly reflect spectral-regularity representations that may simply not generalize to other conditions where neural adaptation is less likely to play a role. If this is the case, our results are still important because it would extend our understanding of
the nature of neural adaptation beyond HG activity (see Fig. 2.4 and Table 2.2; (Fishman and Steinschneider, 2012; Eliades et al., 2014)). Our findings also implicate a broad cortical circuit that mediates these numerous correlates of neural adaptation (e.g., see Fig. 2.12 and Tables 2.5-2.7).

Nonetheless, several lines of evidence suggest that our findings do, in fact, reflect a true regularity representation. First, whereas a primary feature of neural adaptation is a systematic reduction in HG activity to commonly presented stimuli, we were generally unable to find a reliable effect in HG activity across electrodes. This is consistent with our finding that, overall, there were not any reliable differences in HG activity as a function of tone-burst frequency. Second, previous work has shown that the differential effects of neural adaptation are minimal with similar separations in tone-burst frequency and repetition rate (Fishman et al., 2004), suggesting that the effects we found are not likely to be attributable solely to neural adaptation. Third, neural adaptation itself has been suggested to be a correlate of regularity representations (Ulanovsky et al., 2004; Nelken and Ulanovsky, 2007; Winkler et al., 2009). Fourth, our results are consistent with previous work that demonstrated oscillatory-phase progression is dependent on the statistical structure of the auditory stimuli (Patel and Balaban, 2000), which was also likely to be independent of neural adaptation.

**Comparison with previous studies on regularity representation and deviance detection**

Our current results are, at first glance, inconsistent with the broad MMN literature. In MMN studies, a rare ‘deviant’ stimulus (e.g., 10% of stimulus events) is presented randomly interleaved with a standard stimulus (i.e., the other 90% of events) at a constant rate. In response to this deviant stimulus, there are well-known and well-characterized changes in ECoG and electro- and magnetoencephalographic phase and power (Edwards et al., 2005; Fuentemilla et al., 2008; Hsiao et al., 2009; Ko et al., 2012; Eliades et al., 2014).

This paradigm is similar to the subset of ECoG analyses in which we compared regular subsequence configurations (F₁—F₁—F₁—F₁) and irregular configurations (F₁—F₁—F₁—F₂). However, several factors argue that a direct comparison is not straightforward. First, there is a
difference in the timing of measured activity: we analyzed activity during the time period of 0-100 ms with respect to the last tone in a local configuration, whereas power and phase modulation purportedly related to the MMN typically occur at time periods >50 ms after the deviant tone, with peaks closer to 100-200 ms. Second, there is a difference in the presentation rate of the stimuli: in the MMN literature, stimuli are presented at much slower rates, such that each stimulus elicits an evoked-type response. With the 10-Hz repetition rate of our stimulus, we found that only a small subset of electrodes (~15% of electrodes, see Fig. 2.5) exhibited reliable enhancements in 10-Hz power, suggesting that, at most, evoked activity in response to each tone burst played a minimal role in the present findings. Finally, because of this high presentation rate, any differential effects in the raw ECoG signal beyond 100 ms following the last tone in a local configuration may be influenced by the following tones in the stimulus sequence, thus confounding any analysis of MMN-like activity. In any case, a more appropriate comparison is with the work of Patel and Balaban (Patel and Balaban, 2000), which also showed that phase, rather than power, reflected the structure of acoustic-frequency content in a series of tone-burst sequences.

Regularity, predictability, and phase alignment

The auditory system is designed to segregate or group acoustic information based upon shared or different spectrotemporal regularities in the acoustic environment (Bregman, 1994; Shinn-Cunningham, 2008; McDermott, 2009; Winkler et al., 2009). Spectrotemporal regularities are inherently predictive since they probabilistically define the nature of the acoustic information over time. It has been proposed that auditory perception itself is a process of active prediction (Winkler et al., 2009; Wacongne et al., 2011; Arnal and Giraud, 2012; Bendixen et al., 2012; Sedley et al., 2016): testing predictions based on alternative regularity representations to guide perception. As a consequence, it is critical for the auditory system to develop a process to represent spectrotemporal regularity. Multiple lines of evidence suggest that phase alignment of neural oscillations may contribute to the representation of temporal regularity (Lakatos et al., 2005; 2008; Besle et al.,
A complex relationship between phase alignment and spectral regularity as a function of neural frequency

The second major finding of our study is the nature of the relationships between phase alignment in distinct neural frequency bands and spectral regularity, in particular the delta and 10-Hz frequency bands. We found first that phase alignment in the 10-Hz frequency range (i.e., the tone-repetition rate) was positively correlated with spectral regularity, which is consistent with the fact that phase alignment typically increases in the neural frequency band corresponding to the repetition rate of the stimulus (Lakatos et al., 2005; 2013) and is consistent with previous findings where spectral structure in the acoustic sequence is modulated (Patel and Balaban, 2000). However, activity in the 10-Hz frequency band was only a small portion of the overall picture. Spectral regularities modulated phase alignment in multiple other frequency bands, including not only harmonically related frequency bands (i.e., ~5 Hz and ~20 Hz), but also inharmonic frequency bands (~15 Hz) and frequency bands seemingly unrelated to temporal regularity in the stimulus (delta band).

However, we found that the most prominent frequency band found to track changes in spectral regularity (both in significance of the correlation and proportion of significant electrodes) was the delta band, the one frequency band that is not trivially related to the temporal regularity of the tone-burst sequence. In experiments where a delta-frequency temporal regularity is imposed in either the stimulus or the task, the phase of ongoing delta-band activity differentially aligns to the
stimulus (Lakatos et al., 2005; 2008; 2013) and correlates with both reaction time (Stefanics et al., 2010) and behavioral performance (Henry and Obleser, 2012). It has been proposed that the delta band plays a role in sensory selection in stimulus regimes for which there exists a temporal rhythm in the delta-frequency range, whereby the “excitable” phase of the neural oscillation is appropriately aligned with the onset of the expected incoming stimulus events (Schroeder and Lakatos, 2009). Our findings expand this hypothesis by demonstrating that delta oscillations are modulated in rhythmic regimes even under circumstances where the (primary) spectrotemporal regularity is not in the delta-frequency range. It is possible that the modulations in the delta-frequency range reflect the tracking of regularities on longer time scales (i.e., 300+ ms, or groups of 3+ tone bursts). For instance, local configurations that consist of repeats of triplets of tones (e.g., $F_1$—$F_2$—$F_1$—$F_1$—$F_2$—$F_1$) would have a pattern repetition rate of 3.33 Hz, and configurations with repeats of quadruplets would have a repetition rate of 2.5 Hz. Tracking these longer regularities might require modulations in the delta band. Our findings may be a more general case of the mechanisms that occur in speech, which exhibits regularities on multiple time scales concurrently, and speech processing, which modulates activity in multiple neural frequency bands, including the delta band (Schroeder et al., 2008; Kerlin et al., 2010; Giraud and Poeppel, 2012).

The fact that multiple other neural frequency bands (in addition to the 10-Hz and delta bands) were also reliably modulated by the changes in spectral regularity implies the complex nature of the stimulus representation in the neural signal. Moreover, it strongly suggests that simply selecting a frequency band of interest to analyze a priori based on task or stimulus design, although a perfectly reasonable approach, may not provide a complete picture of the nature of the neural representation. The enhancement in phase alignment to the most regular local configurations in the first harmonic of the repetition rate (i.e., 20 Hz) coincides with the expectation that a temporally regular sequence of events elicits a frequency-following response in neural activity at fundamental and (sub-) harmonic frequencies (Gomez-Ramirez et al., 2011; Henry and Obleser, 2012; Nozaradan et al., 2012; Henry et al., 2014). The activity in these harmonic frequency bands also
correlates with previously identified aspects of task performance (Gomez-Ramirez et al., 2011; Nozaradan et al., 2012; Henry et al., 2014), suggesting that the modulation in the 20-Hz band in our study may be of functional importance. In contrast, to the best of our knowledge, the decrements in phase alignment with increasing spectral regularity in the frequency bands neighboring the 10-Hz band (i.e., the 5- and 15-Hz bands) have not been found previously.

**Applicability to healthy individuals**

One potential concern that affects the current work and indeed all human ECoG work is related to the applicability of the findings to healthy individuals. However, several important factors lend support to the broader applicability of our findings. First, none of our subjects were diagnosed with any hearing impairments, suggesting that our neural findings were not conflated by auditory-perceptual deficits. Second, only electrodes that were deemed free of epileptic activity by the clinical staff were included in the analyses, and subjects were generally not tested within 12 hours of a recorded seizure. Thus, it is unlikely that our results were confounded by long-lasting effects of epileptic activity or interictal events. Third, similar results regarding the representation of spectrotemporal regularity in oscillatory phase have been found in healthy individuals (Patel and Balaban, 2000; Henry and Obleser, 2012; Henry et al., 2014).

**Conclusions**

The current study lends support for the role of neural oscillations in the representation of spectrotemporal regularity and provides greater insight into the complex nature of the representation, in terms of both the distribution of the representation across cortex and across the neural frequency space. Further research is required to elucidate whether neural oscillatory activity plays are causal role in the perceptual representations of spectrotemporal regularities.
CHAPTER 3

3. Neural oscillatory correlates of concurrent spectrotemporal-regularity representation and deviance detection

ABSTRACT

A fundamental goal of the auditory system is to transform auditory stimuli from low-level representations of a stimulus’ acoustic features into perceptual representations (i.e., sounds). These perceptual representations are the result of computational processes that group or segregate acoustic stimuli based on the spectrotemporal regularities that characterize emissions from the same or different sound sources. Here, we identified the mechanisms by which the brain represents stimuli with multiple concurrent regularities—a feature that is characteristic of many natural sounds (e.g., speech)—and reflects a listener’s reports of detected deviations in these spectrotemporal regularities. We identified these mechanisms by recording electrocorticographic
activity in humans while they participated in a deviant-detection task. We found that both ECoG power and phase was modulated across cortex in response to a stimulus with multiple concurrent regularities, but that only phase-alignment was modulated in a neural-frequency-specific manner that reflected the time scales of each regularity. We also found that both ECoG power and phase activity was reliably modulated by both stimulus deviations in spectrotemporal regularity and a listener’s reports of detected deviations in spectrotemporal regularity, but were unable to identify whether neural activity differentially reflected the three different types of deviant stimuli tested. Future work should focus on the contributions of the inferior parietal, ventrolateral and dorsolateral prefrontal, and temporal cortices, which seemed to show the most reliable modulations with respect to stimulus and choice behavior.

INTRODUCTION

A fundamental goal of the auditory system is to parse the auditory scene into distinct perceptual representations (i.e., sounds) that are reflective of the putative sound sources in the environment (van Noorden, 1975; Bregman, 1994; Cusack, 2005). Our ability to parse the auditory scene is the result of computational processes that group or segregate acoustic stimuli based on the spectrotemporal regularities that characterize emissions from the same or different sound sources (Bregman, 1994; Shinn-Cunningham, 2008; McDermott, 2009; Winkler et al., 2009). These spectrotemporal regularities also influence the perception of a stimulus depending on the time scales over which they occur. For example, fluctuations in sound-envelope amplitudes are perceived as changes in loudness on long time scales, flutter on medium time scales, and pitch on short time scales (Joris et al., 2004). Thus, an understanding of auditory perception requires, in part, knowledge of the representation of spectrotemporal regularity and how these representations influence perception.
A complicating factor in parsing the auditory scene is that environmental auditory stimuli often exhibit multiple concurrent spectrotemporal regularities at different timescales that are utilized in conjunction to form perceptual representations. For example, human speech is characterized by regularities reflecting distinct linguistic features that occur concurrently on different time scales (Rosen, 1992; Greenberg et al., 2003; Poeppel, 2003; Golumbic et al., 2012): prosody or phrasal cues in the temporal envelope occur on relatively long time scales (~300–2000 ms); syllabic cues occur on shorter time scales (~100–300 ms); and phonetic cues occur on even shorter time scales (~20–50 ms). Numerous lines of evidence suggest that the auditory system is capable of representing concurrent regularities. First, sound segregation is influenced by conjunctions of multiple regularities (Bendixen et al., 2010; 2013). Second, temporal envelope cues on multiple time scales can act in conjunction to improve speech-sound recognition (Tasell et al., 1987; Drullman et al., 1994a; 1994b; Shannon et al., 1995). And third, the recognition and intelligibility of speech sounds are influenced by the presentation rate, temporal predictability, and contextual information over time scales longer than those of the individual speech sounds (Pollack and Pickett, 1964; Cooper et al., 1978; Ganong, 1980; Mann, 1980; Repp, 1982; Norris et al., 1997; Borsky et al., 1998; Roncaglia-Denissen et al., 2013).

Substantial progress has been made in our understanding of the neural representation of spectrotemporal regularities, including how a variety of neural signatures reflect (1) stimuli characterized by a single spectrotemporal regularity (Ulanovsky et al., 2003; Fishman et al., 2004; Micheyl et al., 2005; Fuentemilla et al., 2006; Hsiao et al., 2009; Lakatos et al., 2013), (2) stimuli that deviate from an established spectrotemporal regularity (Näätänen et al., 2007; Fuentemilla et al., 2008; Hsiao et al., 2009; Escera et al., 2013), and (3) simultaneous sensitivity to a combination of local and global stimulus probabilities (Squires et al., 1976; Ulanovsky et al., 2004). Despite the progress in our understanding of regularity representation, several important issues regarding the nature of concurrent-regularity representation and its relation to auditory perception remain.
Despite the progress in our understanding of regularity representation, several important issues regarding the nature of concurrent-regularity representation and its relation to auditory perception remain. First, although the neural responses to stimuli with a single spectrotemporal regularity have been extensively studied (Ulanovsky et al., 2003; Fishman et al., 2004; Micheyl et al., 2005; Fuentemilla et al., 2006; Näätänen et al., 2007; Fuentemilla et al., 2008; Hsiao et al., 2009; Escera et al., 2013; Lakatos et al., 2013), neural activity in response stimuli with multiple concurrent regularities exhibit a complex pattern of phase and amplitude modulations in population-level activity that is comparatively less well studied (Dimitrijevic et al., 2001; Luo et al., 2006; Sanders and Poeppel, 2007; Henry et al., 2014). Moreover, the relationship between the mechanisms that reflect spectrotemporal regularity representation and those that reflect changes in spectrotemporal regularity have only begun to be addressed (Pannese et al., 2015). Second, because the overwhelming majority of previous research related to spectrotemporal-regularity representation has relied on paradigms that cannot distinguish between perceptual and sensory representations of auditory stimuli (Dimitrijevic et al., 2001; Ulanovsky et al., 2003; Fishman et al., 2004; Lakatos et al., 2005; Micheyl et al., 2005; Luo et al., 2006; Sanders and Poeppel, 2007; Lakatos et al., 2013), little is known about the extent to which the neural activity that reflects spectrotemporal regularities ultimately relates to perception (Henry et al., 2014). An understanding of this relationship is critical for determining the computational mechanisms that transform acoustic information into auditory percepts. Third, because most studies have focused on the contribution of the core auditory cortex with respect to regularity representation (Ulanovsky et al., 2003; Fishman et al., 2004; Ulanovsky et al., 2004; Lakatos et al., 2005; Micheyl et al., 2005; Lakatos et al., 2013), the potential contributions of other cortical regions shown to exhibit neural correlates of auditory-perceptual processes have yet to be fully elucidated (Poremba et al., 2004; Cusack, 2005; Rauschecker, 2012; Barascud et al., 2016; Teki et al., 2016).

Thus, the goals of this study were to identify the mechanisms by which the cortex (both auditory and non-auditory areas) reflects concurrent spectrotemporal regularities and to localize
the sensory versus perceptual correlates of spectrotemporal regularity representation. To achieve these goals, we recorded electrocorticographic (ECoG) activity from electrodes that were distributed across the human cortex while patients participated in a deviance-detection task in which they had to report whether a repeating tone-burst sequence maintained or deviated from a set of three concurrent spectrotemporal regularities. We found that ECoG power modulations in response to the spectrotemporal regularities were widespread and generally consistent throughout cortex, exhibiting decreases in low- and high-frequency power and increases in mid-frequency power that peaked near the oscillatory frequency corresponding to the tone repetition rate. In contrast, ECoG modulations in phase alignment were restricted to a much more limited set of brain regions in each cortical lobe and exhibited peaks in the phase-alignment spectra that corresponded to the time scales of each of concurrent spectrotemporal regularities. Preliminary analyses suggest that power and phase modulations reliably distinguished between stimuli that maintained or deviated from spectrotemporal regularity and between a listener’s reports of hearing a deviant stimulus or not. However, we were unable to identify whether the time scale over which a deviation from regularity occurred affected oscillatory activity in a neural-frequency specific manner. Completion of this work should focus on testing subjects with electrode coverage in the inferior parietal, ventrolateral and dorsolateral prefrontal, and temporal cortices, which seemed to show the most reliable modulations with respect to stimulus and choice behavior.

METHODS

Subjects

17 subjects (7 females, 3 left-handed and 1 ambidextrous, mean age: 37 ± 12 years) with medically intractable epilepsy underwent surgery to implant subdurally platinum recording electrodes on the cortical surface and into the brain parenchyma. In each case, clinical teams
determined electrode placement in order to localize epileptogenic brain regions. Institutional review boards at each hospital approved the research protocol, and informed consent was obtained from each subject prior to their participation.

**Auditory stimuli and task design**

We designed the auditory stimuli and task to test two main questions regarding spectrotemporal-regularity representation. First, how does electrocorticographic (ECoG) activity reflect an acoustic stimulus with multiple concurrent spectrotemporal regularities? And second, how are changes in ECoG activity correlated with behavioral reports of detected deviations in spectrotemporal regularity?

*Task design and stimuli.* Subjects rested comfortably in their hospital beds and took part in a ‘deviant-detection’ task, during which they reported whether or not they detected a change in a stimulus’ spectrotemporal regularity.

The acoustic stimuli were repetitions of a sequence of tone bursts (70 dB SPL; 100-ms duration, gated by 10-ms squared cosine ramps with 100-ms inter-tone interval [5 Hz onset-to-onset interval]). All acoustic stimuli consisted of repetitions of a ‘standard’ sequence, followed by either another standard sequence or one of three ‘deviant’ sequences (described below).

A ‘standard’ sequence consisted of three tone-burst triplets (Fig. 3.1a). Each tone burst within an individual triplet had the same frequency (800, 1040, or 1280 Hz). In this sequence of tone bursts, there are two spectrotemporal regularities: (1) a ‘local’ regularity (periodicity $T_L=200$ ms), which defines the frequency of the tone bursts within a triplet; and (2) a more ‘global’ regularity (periodicity $T_G=600$ ms), which defines the frequency transition between the triplets. In any acoustic stimulus, this standard sequence could be repeated 2-5 times, creating (3) a third, sequence-level, regularity (periodicity $T_{SS}=1800$ ms).
Figure 3.1: Stimulus design and trial progression. (a) Schematic of the standard sequence (SS), consisting of three sets of tone triplets. Information in red describes the tone repetition rate (the ‘local’ regularity), information in blue describes the repetition rate of the triplet (the ‘global’ regularity), and information in green describes the repetition rate of the SS (the ‘sequence-level’ regularity). (b) Schematic of the local deviant ($D_L$), which corresponded to a frequency increase in the middle tone of the final triplet (red-colored tone burst). (c) Schematic of the global deviant ($D_G$), which corresponded to a frequency increase of each tone in the final triplet (blue-colored tone bursts). (d) Schematic of the local+global deviant ($D_{LG}$), which corresponded to a combination of the local and global deviants (purple-colored tone bursts). Note: in (b)-(d), gray-shaded and dashed tone bursts in the final triplet correspond to the final-triplet tone frequency in the standard sequence. (e) Depiction of a single-trial progression. On a standard trial (left), the stimulus consisted of 3-5 repeats of the standard sequence only. On a deviant trial (right), the stimulus consisted of 2-4 repeats of the standard sequence, followed immediately by one of the three deviant sequences. The duration of each trial depended on the number of repeats of standard and/or deviant sequences (bottom).

A ‘deviant’ sequence was one that began as a standard sequence but with a frequency change in the third triplet that disrupted the established spectrotemporal regularity of the standard sequence. We constructed three types of deviant sequences. (1) A ‘local’ deviant ($D_L$) occurred when the frequency of the middle tone in the final triplet was increased compared to the others in the same triplet (Fig. 3.1b). (2) A ‘global’ deviant ($D_G$) occurred when the frequency of each tone in
the final triplet was increased by more than the standard 30% relative to the previous triplet (Fig. 3.1c). (3) A 'local+global' deviant \( (D_{LG}) \) contained both the local and global deviations (Fig. 3.1d). In any acoustic stimulus with a deviant sequence, the deviant sequence was always the last sequence in the stimulus.

In the detection task, trials were categorized as either standard or deviant. In a standard trial, listeners heard 3-5 repeats of the standard sequence only (Fig. 3.1e, left). In a deviant trial, the stimulus consisted of 2-4 repeats of the standard sequence, immediately followed by a deviant sequence (Fig. 3.1e, right). Multiple repetitions of the standard sequence were included in each stimulus to ensure that the three regularities were present in each trial. However, the number of repetitions varied across stimuli to minimize the possibility that subjects could predict when the stimulus would end, forcing them to attend to the entire stimulus.

The timing of the task is depicted in Figure 3.2, which was the same for both standard and deviant trials. After offset of the last tone burst, subjects had 3000 ms to report their response and received immediate visual feedback on their report. They pressed either 'I' or 'Z' keys to report whether they heard or did not hear a deviant sequence, respectively. The inter-trial interval was jittered randomly between 1000 and 1500 ms. On a trial-by-trial basis, we randomly selected whether the trial was standard or deviant. On deviant trials, we randomly selected the type of deviant sequence (i.e., \( D_L, D_G, \) or \( D_{LG} \)).

**Trial outcomes.** Trials fell into four categories based on trial type and the subject's behavioral report. (1) A hit (H) was a correct report of a deviant trial. (2) A correct rejection (CR) was a correct report of a standard trial. (3) A miss (M) was an incorrect report of a standard trial. Finally, (4) a false alarm (FA) was an incorrect report of a deviant trial.
Task training and testing. In the first session, subjects completed a series of 15 practice trials (7 standard trials) to familiarize themselves with the stimuli and the timing of the task. After the first session, subjects had the option of forgoing the practice session. Subjects completed between 1-8 sessions of the task, with each session containing 54 trials (27 standard trials and 9 trials of each deviant type). The time between sessions varied across subjects from 1 minute and 5 days.

A pilot study with four healthy subjects using the same task design tested how deviant-detection performance varied as the difference between the standard and deviant tone-burst frequencies increased. In the pilot study, we measured detection performance at 8 different levels of frequency deviation for each type of deviant sequence (Fig. 3.3). This let us determine an appropriate range of frequency deviations for the neurophysiological experiment.

Based on these results, subjects participated in one of two versions of the deviant-detection task. In version 1 (for the first six subjects), a single frequency deviation was used for each type of deviant (indicated by black vertical dashed lines in Fig. 3.3). In version 2 (for the last 11 subjects), for each subject, we adjusted the frequency increase between the standard and deviant tone bursts across testing sessions. We adjusted this frequency difference to help to ensure that subjects performance on the deviant trials was ~50%. For example, if a subject’s hit rate was 75% (or 25%) for deviant $D_L$, on the subsequent session, we decreased the frequency difference to make the task...
more difficult. Similarly, if a subject’s hit rate was 25%, we increased the frequency difference to make the task easier. Because of this procedure, we could sample both hit and miss trials to test how neural activity reflected reports of perceived deviant detections independent of changes in the actual stimulus.

**Behavioral analyses**

We tested the significance of each subject’s performance against chance (50%), first across all trials, then separately for standard and deviant trials (one-sample z-tests, \( p < 0.05 \)). We also conducted a \( d' \) analysis to determine an unbounded measure of behavioral performance for each subject (two-sample z-tests, \( p < 0.05 \)).

**Data acquisition and preprocessing**

Subdural electrodes were arranged in either grids or strips; each electrode contact was separated by 10 mm. Depth electrodes contained 6-8 contacts that were separated by 8 mm; the depth electrodes were located primarily in the medial temporal lobes. Electrodes were localized by co-registering post-operative computed-tomography scans with post-operative MRI scans using the FSL (FMRIB [Functional MRI of the Brain] Software Library), BET (Brain Extraction Tool), and FLIRT (FMRIB Linear Image Registration Tool) software packages. These electrode locations were then mapped to Talairach space using indirect stereotactic techniques and the OsiriX Imaging Software DICOM viewer package (Burke et al., 2013).

We recorded ECoG signals either with a Nicolet or a Nihon Kohden electroencephalogram system (Burke et al., 2013). ECoG signals were sampled at 1000 Hz. A testing laptop sent ±5-V analog pulses, via an optical isolator, to open lines in the clinical-recording system to align the stimulus- and task-related events with the ECoG recordings.

To minimize reference-line and volume-conduction confounds, we used a bipolar-referencing scheme (Nunez and Srinivasan, 2006; Burke et al., 2013) in which we subtracted the
signals from each pair of immediately adjacent electrode contacts on the same grid, strip, or depth electrode (Anderson et al., 2010; Burke et al., 2013). We assumed that these bipolar signals were located midway between each electrode-contact pair.

To test ECoG sensitivity to stimulus regularities (see Electrophysiological analyses), we down-sampled ECoG activity to 500 Hz and either notch-filtered ($4^{th}$-order zero-phase-shift Butterworth filter; stop-band: 58-62 Hz) to remove power-line noise for wideband analyses or low- or band-pass filtered for frequency-band-specific analyses ($2^{nd}$ or $4^{th}$-order zero-phase-shift Butterworth filters for low- and band-pass, respectively, with pass bands indicated in Table 3.1).

<table>
<thead>
<tr>
<th>Band</th>
<th>Passband</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low delta</td>
<td>&lt;1 Hz</td>
</tr>
<tr>
<td>Delta</td>
<td>1–4 Hz</td>
</tr>
<tr>
<td>Theta</td>
<td>4–8 Hz</td>
</tr>
<tr>
<td>Alpha</td>
<td>8–12 Hz</td>
</tr>
<tr>
<td>Beta</td>
<td>12–25 Hz</td>
</tr>
<tr>
<td>Low gamma</td>
<td>25–58 Hz</td>
</tr>
<tr>
<td>High gamma</td>
<td>70–200 Hz</td>
</tr>
</tbody>
</table>

**Electrophysiological analyses**

We designed our analyses to test two main hypotheses: (1) multiple spectrotemporal regularities were represented by increased power and phase alignment in neural frequencies that related to the timescales of the stimulus regularities; and (2) changes in ECoG activity, reflected both in the raw ECoG signal and in power and phase modulations, correlated with the subject’s behavioral reports. For example, we hypothesized that ECoG activity in response to hits (i.e., correct reports of deviant trials) should be different than correct rejections (i.e., correctly reports of standard trials). Similarly, we hypothesized that ECoG activity in response to hits and false alarms (i.e., reports of deviant trials) should be similarly different than misses and correct rejections (i.e., reports of standard trials), respectively. Finally, we hypothesized that hit and miss trials for each
deviant type should elicit distinct power and/or phase modulations in the neural frequency band(s) that correspond to the timescales of each deviation in spectrotemporal regularity.

To test for activity related to the spectrotemporal regularities in the standard sequence, we tested how ECoG power and phase activity aligned to the standard sequence differed from randomly selected portions of the ECoG signal. To test for activity related to deviant detection, we performed three separate sets of analyses to test for power and phase modulations as a function of trial type and/or behavioral report, which are described in detail below. The specific analyses for each hypothesis are described in detail below.

(1) Testing the sensitivity of ECoG signals to concurrent spectrotemporal regularities in the standard pattern

First, we tested whether the three spectrotemporal regularities in the standard sequence elicited power or phase-alignment enhancements in the neural frequency bands reflecting each regularity’s corresponding time scale (see Auditory stimuli and task design and Fig. 3.1). To test for this sensitivity, we first extracted the ECoG signals corresponding to the second presentation of the standard pattern in every trial and then performed a wavelet decomposition to extract the instantaneous power and phase from ~0.4-200 Hz (Manning et al., 2009; Burke et al., 2014). Next, we computed the mean log power and pairwise phase consistency (PPC; and unbiased measure of phase alignment) (Vinck et al., 2010) across the standard pattern. We ignored the first presentation of the standard pattern to remove any influence of potential ERP responses to stimulus onsets on our measures spectrotemporal modulation.

A randomization procedure tested the significance of these power and PPC measures. First, we randomized the relationship between the ECoG signal and the timing of the task and calculated its random mean power and PPC values at each wavelet frequency. This randomization was repeated 1000 times to create distributions of power and PPC values. The significance levels of the actual power and PPC measures (2-tail comparisons) were calculated relative to their
respective random distributions. Reliable electrodes had lengths of significant activity (either power or PPC) in contiguous frequency bands that exceeded the false-positive rate (i.e., corrected $p<0.05$).

To test whether the number of spectrotemporal-regularity-modulated electrodes in a particular brain region was significantly greater than chance, we used a “counts t-test” analysis (Ramayya et al., 2015). We first converted the number of significant electrodes that were modulated by regularity into z-scores using a binomial null distribution based on the total number of electrodes and a false-positive rate of 0.05 (0.025 each for positive and negative modulations). We then tested whether the population of z-scores across subjects differed significantly from zero using a one-sampled t-test. We corrected for multiple comparisons across brain regions using false-discovery-rate (FDR) correction. In an analogous manner, we tested whether the number of behavior-modulated electrodes in a particular brain region was significantly greater than chance ((Storey and Tibshirani, 2003); Q=0.05).

For this and all subsequent brain-region-specific analyses, electrodes were categorized into separate brain regions based on their associated anatomical labels (Table 3.2).

(2) Testing the stimulus- and behavioral-sensitivity of ECoG signals in relation to deviant detection

We conducted the deviant-detection analyses in two ways. First, we used the libsvm package (freely available resource; (Chang and Lin, 2011)) to build linear classifiers and performed two-class learning for ECoG modulations as a function of stimulus type (i.e., standard or deviant) and/or behavioral report (e.g., hits versus misses). Second, we used the results from the classifier analyses to specify and constrain direct tests on wideband power and phase alignment.

One subject did not understand the task. As a result, behavioral responses were not recorded and this subject was not included in the following analyses. Two other subjects reported difficulty hearing the stimuli, and, therefore, were also not included in these analyses. All of the remaining subjects were included in the analyses that compared the responses in standard trials
to the combined responses across all deviant trials (N=14). Only subjects that completed ≥3 sessions were included in the analyses testing responses to the individual deviant types (N=9).

Table 3.2: Regions of interest. Anatomical labels used to define regions of interest.

<table>
<thead>
<tr>
<th>Lobe</th>
<th>Region of interest</th>
<th>Desikan-Killiany atlas labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal</td>
<td>Orbitofrontal</td>
<td>medialorbitofrontal, lateralorbitofrontal</td>
</tr>
<tr>
<td></td>
<td>DorsolateralPrefrontal</td>
<td>rostralmiddlefrontal, caudalmiddlefrontal</td>
</tr>
<tr>
<td></td>
<td>VentrolateralPrefrontal</td>
<td>parstriangularis, parsopercularis, parsorbitalis</td>
</tr>
<tr>
<td></td>
<td>AnteriorMedialFrontal</td>
<td>superiorfrontal, rostralanteriorcingulate, caudalanteriorcingulate</td>
</tr>
<tr>
<td></td>
<td>PosteriorMedialFrontal</td>
<td>paracentral, posteriorcingulate, isthmuscingulate</td>
</tr>
<tr>
<td>Frontal/Parietal</td>
<td>Sensorimotor</td>
<td>precentral, postcentral</td>
</tr>
<tr>
<td>Parietal</td>
<td>SuperiorParietal</td>
<td>superiorparietal</td>
</tr>
<tr>
<td></td>
<td>InferiorParietal</td>
<td>inferiorparietal</td>
</tr>
<tr>
<td></td>
<td>Supramarginal</td>
<td>supramarginal</td>
</tr>
<tr>
<td>Occipital</td>
<td>Occipital</td>
<td>cuneus, lateralloccipital, lingual, pericalcarine</td>
</tr>
<tr>
<td>Temporal</td>
<td>SuperiorTemporal</td>
<td>superiortemporal</td>
</tr>
<tr>
<td></td>
<td>OtherTemporal</td>
<td>banksts, middletemporal, inferiortemporal, fusiform</td>
</tr>
<tr>
<td></td>
<td>MedialTemporalLobe</td>
<td>entorhinal, parahippocampal; depth contacts labeled as hippocampal; entorhinal, perirhinal, or parahippocampal by neuroradiologist</td>
</tr>
</tbody>
</table>

Linear classifier analyses. To assess the extent and time course of ECoG modulations with respect to deviant detection in spectrotemporal regularity, we first built an ‘H v. CR’ classifier that compared responses to all hits (i.e., correctly reported deviant trials) to those in all correct-rejections (i.e., correctly reported standard trials). First, we bandpass filtered ECoG signals into 7 distinct frequency bands (see Data acquisition and preprocessing) and extracting the instantaneous power and phase responses using the Hilbert transform. Next, we extracted 1.6-s
epochs of ECoG activity (in each band) centered on the onset of the final tone triplet of each stimulus in the analysis, corresponding to an 800-ms (4-tone) 'baseline' period that was the same across all stimuli and an 800-ms 'assessment' period that included the final 3 tones in each stimulus (i.e., where a deviant might occur) and the first 200 ms of the response period following stimulus offset. ECoG signals were then sorted into two groups based on trial and outcome identity (i.e., H or CR). For each electrode, we built linear classifiers with 10-fold cross-validation for each time point to evaluate whether instantaneous power and/or phase activity could distinguish between hits and correct rejections, using each frequency band as a feature in the classifier. To account for uneven sample sizes, we employed a multi-downsizing technique to equalize the sizes of each class (Blagus and Lusa, 2010). For each cross-validation iteration, we built 101 linear classifiers using random downsized samples of the majority class in the training data and recorded the class predictions of the test data from each classifier. The class predictions of the test data for each cross-validation iteration were determined by majority vote from the 101 classifiers.

To identify when ECoG modulations reliably encoded hits versus correct-rejections, we first computed the grand-mean prediction-accuracy measurements across electrodes for each subject. Significant time points (i.e., classification accuracy greater than 0.5) following onset of the final tone triplet were determined across subjects using signed-rank tests. The baseline period was used to determine an appropriate significance threshold to set a false-positive rate (0.05) for the time points in the assessment period.

Significant prediction accuracy in the H v. CR analysis would suggest that ECoG activity differentially encodes spectrotemporal regularities and detected deviations. However, it did not differentiate between the dual effects associated with trial type (i.e., different stimulus characteristics) and behavioral report (i.e., reflecting different perceptual characteristics). Consequently, we conducted three additional linear-classifier analyses in an analogous manner to differentiate between stimulus- and perceptual-related modulations. To identify modulations to stimulus type, we built a 'stimulus' classifier that compared ECoG activity in all deviant trials to all
standard trials (i.e., independent of behavioral report). To identify modulations specific to behavior, independent of stimulus type, we built a 'behavioral' that compared ECoG activity in all reported deviant trials (H and FA trials) to all reported standard trials (M and CR trials). Finally, to identify modulations specific to the interaction between stimulus and behavior, we built a 'behavioral x stimulus' classifier in which we first sorted ECoG activity by trial type (i.e., standard or deviant) and then, separately, compared responses by behavioral report (i.e., FA versus CR trials and H versus M trials). We also built each of these three classifiers after first segregating the deviant trials by deviant type (i.e., local, global, or local+global).

For each of these classifiers, we tested the across-subject classification accuracy by time period with signed-rank tests after computing the grand-mean prediction-accuracy measurements for each subject. For these analyses, we used a coarser temporal analysis by dividing the 1.6-s epochs into 16 100-ms time periods. We also tested the classification accuracy of the raw ECoG signals, without temporal averaging. Finally, we repeated each of these classifier analyses after aligning ECoG activity to a 1.6-s time epoch around each behavioral report, starting 1.4 s prior to report and ending 0.2 s after.

**Direct tests on neural modulations.** To determine how the modulations in neural power and phase differentiated between stimulus and perceptual conditions, we performed tests on power and phase-alignment responses as a function of either stimulus or behavioral conditions. We used the linear-classifier analyses to identify the significant time periods and electrodes to include in these analyses.

For the direct stimulus analysis, ECoG signals were sorted into two groups based upon trial type (i.e., standard and deviant). Next, a wavelet decomposition was performed to extract the instantaneous power and phase responses. Mean log-power and PPC estimates were calculated for each frequency during the time period 400-600 ms following final-triplet onset. An analogous procedure was performed to assess deviant-type-specific modulations by first sorting the deviant
trials into separate groups by the type of deviant. For each electrode, we computed the difference in power or PPC between groups (e.g., all local deviant versus standard trials, all global deviants versus standard trials, etc.). For each subject, we then computed grand-mean power and PPC differences across electrodes. Finally, the significance of the grand-mean power and PPC differences across subjects were tested using signed-rank tests, with FDR correction across neural frequencies.

In a similar manner, we computed the mean power and PPC differences as a function of frequency for the direct behavioral analysis after by sorting ECoG signals by behavioral report, independent of stimulus. Finally, in the direct behavioral x stimulus analysis we computed the mean power and PPC differences as a function of behavioral report, separately for standard trials (FA versus CR) and deviant trials (H versus M). These analyses were then repeated after separating deviant trials by the type of deviant.

RESULTS

We recorded ECoG activity from subdural surface and depth electrodes across the cortex while human subjects participated in a deviant-detection task. During this task, they reported whether or not they detected a change in a stimulus' spectrotemporal regularity. The stimulus was a sequence of tone-burst triplets that had spectrotemporal regularities on three different time scales (see METHODS). The 'local' regularity defined the relationship between each tone within a triplet (periodicity $T_L=200$ ms) and the 'global' regularity defined the relationship between triplets in a sequence (periodicity $T_G=600$ ms). Because the sequence of tone-burst triplets repeated within each stimulus, there was also a sequence-level regularity (periodicity $T_{SS}=1800$ ms). Three types of deviant sequences were tested: a 'local' deviant ($D_L$) consisted of a frequency increase in the middle tone of a sequence's final triplet; a 'global' deviant ($D_G$) consisted of a frequency increase
in all three tones of a sequence’s final triplet; and a ‘local+global’ deviant ($D_{LG}$) consisted of both a local and global deviant.

A pilot study with 4 healthy subjects, which used the same task design, tested performance as a function of 8 different levels of acoustic-frequency increases for each type of deviant sequence. The results of the pilot study were integrated into the stimulus parameters that were employed during the electrophysiological study.

We performed a series of analyses that tested two main questions regarding spectrotemporal regularity representation. First, how does an acoustic stimulus with multiple concurrent spectrotemporal regularities modulate the power and phase of electrocorticographic (ECoG)? We addressed this question by testing whether ECoG power and phase activity aligned to the standard sequence differed from randomly selected portions of the ECoG signal. Second, how are behavioral reports correlated with changes in power and phase? This question was tested using separate linear-classifier analyses and direct tests on ECoG activity to test for power and phase modulations as a function of behavioral report and/or trial type.

**Figure 3.3: Average performance in pilot study.** Across-subject mean hit rates as a function of frequency increase for $D_L$ trials (left), $D_G$ trials (middle), and $D_{LG}$ trials (right). Black dashed lines in each plot indicate the level of frequency deviation employed in version 1 of the electrophysiological experiment. Gray shaded regions depict the range of frequency-deviations levels employed in version 2 of the electrophysiological experiment. Note: $\Delta F/F$ values for $D_{LG}$ are plotted as the sum of the individual $D_L$ and $D_G$ deviation levels.
For healthy subjects, deviant detection improves as frequency deviations increase for all deviant types

For healthy subjects, hit rate increased as the frequency value of the deviant stimulus (present in the final triplet of tones) increased (Fig. 3.3). Performance was similar across deviant types. Based on these findings, we estimated the value of acoustic-frequency increase for each deviant type that would likely elicit chance performance in the neurophysiological study. In the first version of the neurophysiological experiment, we chose a deviant magnitude of $\Delta F/F = 0.045$ for each deviant type (black vertical dashed lines in Fig. 3.3). We chose this value, which corresponded to average hit rate of $\sim 0.9$ for each deviant type, in case the neurophysiological subjects’ detection thresholds were worse than the healthy subjects to ensure that the task was not too difficult. In the second version of the neurophysiological experiment, we titrated the frequency values for each deviant type (within gray shaded regions of Fig. 3.3) to ensure that the patients were operating near chance levels.

Only phase alignment is modulated in a frequency-specific manner corresponding to the timescales of spectrotemporal regularities

We first tested whether oscillatory activity was modulated by the spectrotemporal regularities in the repeating standard sequence. To test this, we measured the mean power and phase alignment of ECoG activity aligned to the presentation of the second standard sequence in every stimulus and compared these values to a distribution of null mean values taken from randomly selected segments of ECoG activity.
Table 3.3: Summary statistics for power modulations to the standard sequence.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of electrodes</th>
<th>Number of subjects</th>
<th>Proportion of electrodes with positive power modulations; counts t-test results</th>
<th>Proportion of electrodes with negative power modulations; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AnteriorMedialFrontal</td>
<td>107</td>
<td>9</td>
<td>0.379; t(8) = 2.91, p = 0.00991</td>
<td>0.322; t(8) = 2.66, p = 0.0144</td>
</tr>
<tr>
<td>DorsolateralPrefrontal</td>
<td>256</td>
<td>16</td>
<td>0.331; t(15) = 3.59, p = 0.00133</td>
<td>0.306; t(15) = 4.03, p = 5.46 × 10⁻⁴</td>
</tr>
<tr>
<td>Orbitofrontal</td>
<td>109</td>
<td>14</td>
<td>0.219; t(13) = 2.37, p = 0.0171</td>
<td>0.425; t(13) = 3.18, p = 0.00360</td>
</tr>
<tr>
<td>PosteriorMedialFrontal</td>
<td>26</td>
<td>7</td>
<td>0.302; t(6) = 2.34, p = 0.0288</td>
<td>0.351; t(6) = 2.46, p = 0.0244</td>
</tr>
<tr>
<td>VentrolateralPrefrontal</td>
<td>93</td>
<td>12</td>
<td>0.242; t(11) = 2.23, p = 0.0239</td>
<td>0.509; t(11) = 6.12, p = 3.76 × 10⁻²</td>
</tr>
<tr>
<td>Frontal/Parietal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensorimotor</td>
<td>221</td>
<td>11</td>
<td>0.491; t(10) = 13.5, p = 4.86 × 10⁻⁵</td>
<td>0.394; t(10) = 5.24, p = 1.88 × 10⁻⁴</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InferiorParietal</td>
<td>146</td>
<td>17</td>
<td>0.508; t(16) = 6.40, p = 4.41 × 10⁻⁵</td>
<td>0.333; t(16) = 4.88, p = 8.37 × 10⁻⁵</td>
</tr>
<tr>
<td>SuperiorParietal</td>
<td>55</td>
<td>9</td>
<td>0.429; t(8) = 3.34, p = 0.00509</td>
<td>0.450; t(8) = 2.60, p = 0.0158</td>
</tr>
<tr>
<td>Supramarginal</td>
<td>145</td>
<td>14</td>
<td>0.493; t(13) = 4.88, p = 1.50 × 10⁻⁴</td>
<td>0.401; t(13) = 4.57, p = 2.63 × 10⁻⁴</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital</td>
<td>44</td>
<td>8</td>
<td>0.394; t(7) = 3.42, p = 0.00553</td>
<td>0.505; t(7) = 4.96, p = 8.17 × 10⁻⁴</td>
</tr>
<tr>
<td>Temporal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MedialTemporalLobe</td>
<td>94</td>
<td>13</td>
<td>0.292; t(12) = 3.19, p = 0.00391</td>
<td>0.302; t(12) = 3.79, p = 0.00129</td>
</tr>
<tr>
<td>SuperiorTemporal</td>
<td>160</td>
<td>15</td>
<td>0.356; t(14) = 4.24, p = 4.13 × 10⁻⁴</td>
<td>0.459; t(14) = 4.32, p = 3.66 × 10⁻⁴</td>
</tr>
<tr>
<td>OtherTemporal</td>
<td>324</td>
<td>15</td>
<td>0.291; t(14) = 3.97, p = 7.04 × 10⁻⁴</td>
<td>0.472; t(14) = 5.65, p = 2.99 × 10⁻⁴</td>
</tr>
</tbody>
</table>

For each brain region (column 1), we list the number of electrodes (column 2), number of subjects (column 3), proportion of electrodes with positive modulations in power (column 4), and proportion electrodes with negative modulations in power (column 5). Positive t-statistics indicate proportions that are greater than expected, whereas negative t-statistics indicate proportions that are lower than expected. Bold text in columns 3 and 4 indicate regions that showed modulated electrodes more frequently than expected by chance (FDR-corrected p < 0.05).

Across all brain regions tested (see METHODS), the standard sequence elicited reliable modulations in power across subjects compared to random (i.e., noise) segments of ECoG (Table 3.2; counts t-test, FDR-corrected ps < 0.05). Moreover, in each brain region, we could identify both reliable increases and decreases in the wideband power spectrum. In contrast, reliable phase modulations were evident only in a subset of the tested brain regions (see Table 3.3; counts t-test, FDR-corrected ps < 0.05). These regions included all areas in the temporal cortex, the inferior and supramarginal areas of the parietal cortex, and the dorso- and ventro-lateral prefrontal cortex. Additionally, whereas modulations in power could be either positive or negative, modulations in phase alignment were exclusively positive except for the superior temporal cortex.
Table 3.4: Summary statistics for phase-alignment modulations to the standard sequence.

<table>
<thead>
<tr>
<th>Region of interest</th>
<th>Number of electrodes</th>
<th>Number of subjects</th>
<th>Frequency of electrodes with positive phase modulations; counts t-test results</th>
<th>Frequency of electrodes with negative phase modulations; counts t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frontal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior/Medial Frontal</td>
<td>107</td>
<td>9</td>
<td>0.0546; t(8) = 1.47, p = 0.0903</td>
<td>0.0243; t(8) = 0.310, p = 0.382</td>
</tr>
<tr>
<td>Dorsolateral Prefrontal</td>
<td>256</td>
<td>16</td>
<td>0.114; t(15) = 3.58, p = 0.00135</td>
<td>0.0366; t(15) = 0.780, p = 0.218</td>
</tr>
<tr>
<td>Orbitofrontal</td>
<td>109</td>
<td>14</td>
<td>0.103; t(13) = 1.49, p = 0.0805</td>
<td>0.0410; t(13) = 0.738, p = 0.237</td>
</tr>
<tr>
<td>Posterior/Medial Frontal</td>
<td>26</td>
<td>7</td>
<td>0.0390; t(6) = 0.686, p = 0.259</td>
<td>0; t(6) = -6.25, p = 1.00</td>
</tr>
<tr>
<td>Ventrolateral Prefrontal</td>
<td>93</td>
<td>12</td>
<td>0.264; t(11) = 3.14, p = 0.00467</td>
<td>0.0333; t(11) = 0.0911, p = 0.465</td>
</tr>
<tr>
<td>Frontal/Parietal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensorimotor</td>
<td>221</td>
<td>11</td>
<td>0.324; t(10) = 3.38, p = 0.00349</td>
<td>0.120; t(10) = 2.23, p = 0.0247</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inferior Parietal</td>
<td>146</td>
<td>17</td>
<td>0.125; t(16) = 2.78, p = 0.00662</td>
<td>0.0441; t(16) = 1.12, p = 0.140</td>
</tr>
<tr>
<td>Superior Parietal</td>
<td>55</td>
<td>9</td>
<td>0.181; t(8) = 1.80, p = 0.0547</td>
<td>0.0455; t(8) = 1.19, p = 0.134</td>
</tr>
<tr>
<td>Supramarginal</td>
<td>145</td>
<td>14</td>
<td>0.247; t(13) = 4.91, p = 1.42 x 10^-4</td>
<td>0.0503; t(13) = 1.90, p = 0.0401</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital</td>
<td>44</td>
<td>8</td>
<td>0.326; t(7) = 2.96, p = 0.0105</td>
<td>0.0469; t(7) = 0.718, p = 0.248</td>
</tr>
<tr>
<td>Temporal Lobe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial Temporal Lobe</td>
<td>94</td>
<td>13</td>
<td>0.154; t(12) = 2.44, p = 0.0156</td>
<td>0.0850; t(12) = 0.963, p = 0.117</td>
</tr>
<tr>
<td>Superior Temporal</td>
<td>160</td>
<td>15</td>
<td>0.482; t(14) = 6.14, p = 1.28 x 10^-5</td>
<td>0.150; t(14) = 3.25, p = 0.00259</td>
</tr>
<tr>
<td>Other Temporal</td>
<td>324</td>
<td>15</td>
<td>0.171; t(14) = 2.52, p = 0.0122</td>
<td>0.0489; t(14) = 1.59, p = 0.0671</td>
</tr>
</tbody>
</table>

Data follows the same organization as in Table 3.3.

To get a better sense of the nature of power and phase modulations across cortex, we computed the across-subject, noise-subtracted mean spectra (Figs. 3.4 and 3.5). Modulations in power across cortex primarily consisted of increases in middle (~1-16 Hz) frequencies and power decreases in low (<1 Hz) and higher (>32 Hz) frequencies (Fig. 3.4a). Power increases were significant around 5 Hz (red arrow in Fig. 3.4a; sign-rank tests, FDR-corrected ps < 0.05), which corresponds to the time scale of the local regularity (i.e., the tone repetition rate). Power decreases were significant in gamma (~32–70 Hz) and high-gamma (~70–200 Hz) activity (sign-rank tests, FDR-corrected ps < 0.05).

These patterns were generally consistent across each tested brain region (Fig. 3.4b-n). Anterior-medial frontal, inferior parietal, sensorimotor, superior temporal, and supramarginal regions exhibited significant power increases at 5 Hz, suggesting that these regions were sensitive to the local regularity. The superior temporal cortex also exhibited a significant power increase at ~1.67 Hz, which suggests sensitivity to the global regularity as well. The fact that some regions (e.g., medial temporal lobe) did not exhibit significant modulations in the across-subject power spectra, even though a reliable proportion of electrodes in these regions were significantly
modulated, suggests that electrodes in these regions had equally large power increases and decreases.

Figure 3.4: Grand-mean power modulations to standard sequence across subjects. (a) Whole-brain averages of power modulations to standard sequence. Thin gray traces reflect grand-mean, noise-subtracted power spectra across all electrodes for each subject. Thick black trace depicts across-subject mean power spectrum. Colored arrows indicate frequencies corresponding to the time scales of the sequence-level (green), global (blue), and local (red) regularities. Black bars above traces denote frequencies with significant modulations across subjects (signed-rank tests, FDR-corrected $p < 0.05$). (b)-(n) Power modulations for each brain region determined to exhibit reliable modulations in power to the standard sequence in the counts t-test analysis. The identity of each brain region is depicted above each plot. Each follows the same conventions as in (a).
Figure 3.5: Grand-mean phase-alignment modulations to standard sequence across subjects. (a) Whole-brain averages of modulations in pairwise phase consistency (PPC) in response to the standard sequence. Data follow the same conventions as in Fig. 3.4a. (b)-(j) PPC modulations for regions with reliable modulations in phase alignment determined in the co-unts t-test analysis. Data follow same conventions as in Fig. 3.4 (b)-(n). For all plots, brown arrows depict peaks at harmonics relative to the frequencies corresponding to the local and/or global regularities.

In contrast to power, modulations in phase alignment were a more direct reflection of the time scales of the spectrotemporal regularities (Fig. 3.5). In the whole-brain averaged spectra (Fig. 3.5a), we could identify peaks at frequencies corresponding to the time scales of each regularity: ~0.55 Hz (sequence-level regularity; green arrow), ~1.67 Hz (global regularity; blue arrow), and 5 Hz (local regularity; red arrow). Moreover, we could identify a smaller peak at 10 Hz, corresponding to the first harmonic of the local regularity (brown arrow). In all regions identified as having reliable
phase modulations (see Table 3.3), we could identify peaks at ~0.55 Hz, suggesting that all of these regions were sensitive to the sequence-level regularity. Additionally, in superior temporal, supramarginal, and sensorimotor regions, we could readily identify peaks corresponding ~1.67, 5 Hz, and 10 Hz. In general, the PPC estimates across neural frequencies were significantly greater than the mean of the noise distributions, likely due to spectral leakage across frequencies.

![Figure 3.6](image)

**Figure 3.6: Across-subject mean power and PPC modulations by brain region.** For comparison, the across-subject mean power modulations (left) and PPC modulations (right) are plotted for each reliably-modulated brain region in the respective counts t-test analysis. Color of traces denote brain region described in legend above plots.

Figure 3.6 depicts the grand-mean power and phase spectra across subjects for each brain region for a more direct comparison of the range of modulations with respect to noise. For power, modulations were similar across brain regions (Fig. 3.6, left). For phase, the level of phase alignment modulations seemed to be more brain-region dependent, with the larger values evident in superior temporal, sensorimotor, and ventrolateral prefrontal cortices and the smaller values evident in the medial temporal lobe and dorsolateral prefrontal cortex (Fig. 3.6, right).
Figure 3.7: Analysis of behavioral performance in neurophysiological experiment. (a) Performance as a function of trial type across subjects. Overall performance across trials is depicted first ("All trials"), followed by performance on standard trials only (1 – false-alarm rate: "1-FAR"), performance on deviant trials only (hit rate: "HR"), and subsequently performance on deviant trials of each type (local-deviant hit rate: "HR<sub>L</sub>"; global-deviant hit rate: "HR<sub>G</sub>"; local+global-deviant hit rate: "HR<sub>LG</sub>"). Colored data points depict performance for each subject, with color distinguishing among subjects. Filled data points depict performance significantly better than (or worse than) chance=0.5 (one-sample z-tests, p<0.05). Black data points and error bars depict across subject means ± standard errors. Filled data point depicts across-subject performance significantly better than chance (signed-rank test, p<0.05). (b) Individual and across-subject d' estimates. Estimates of d' combining across all deviant types is depicted first ("S vs. D"), followed by d' estimates using only performance on each type of deviant trial ("S vs. D<sub>L</sub>" using only local-deviant trials, S vs. D<sub>G</sub>" using only global-deviant trials, S vs. D<sub>LG</sub>" using only local+global-deviant trials). Color of data points follows from (a). Filled colored data points denote individual subjects with significant d' values (two-sample z-tests, ps<0.05). Filled black data points depict significant average d' values across subjects (one-sample t-tests, ps<0.05).

Subjects exhibit highly variable performance in deviant detection

Overall correct performance as a function trial type is depicted in Fig. 3.7a. Across all trial types, mean correct performance across subjects was significantly greater than chance (signed-rank test, p=0.049). Mean correct performance on deviant trials, as designed (see METHODS), was at chance levels (signed-rank test, p>0.05).
However, we found that performance was highly variable across subjects (Fig. 3.7a). The CR rate (equal to 1 – FA rate) was not significantly above chance across subjects (signed-rank test, p>0.05). Whereas some subjects were significantly better than chance on standard trials (e.g., UP041 and R1065J; one-sample z-tests, ps<0.05), other subjects were not better than chance (e.g., R1054J and TJ073). One subject was even significantly worse than chance (R1049J; one-sample z-test, p<0.001). Additionally, hit rates, both across and within deviant types, were also highly variable across subjects (Fig. 3.7a).

These results were mirrored in the d’ analysis (Fig. 3.7b). Though overall d’ measures were significant across subjects (one-sample t-test, p=0.025), only d’ measures computed with the local-deviant trials remained significant (one-sample t-test, p=0.013). Additionally, only 8 subjects had significant d’ measures, either across all deviant types or for a specific deviant (two-sample z-tests, ps<0.05; filled colored circles in Fig. 3.7b). And again, one subject had d’ values significantly less than 0. Thus, whereas the calibrated frequency deviations kept overall hit rates near chance, they were less successful at maintaining each subject’s hit rates near chance.

The variability in hit rates seemed to be largely due to the level of frequency deviation for each deviant type (Fig. 3.8). Again, for about half of the subjects, the initial level of frequency deviations (either 0.045 or 0.015 for task versions 1 and 2, respectively) for each deviant type was successful in maintaining near-chance performance. However, a few subjects performed well at these levels and required smaller frequency deviations to reach chance performance. One subject (R1065J) performed above chance for even the smallest level of deviation for the global and local+global deviants. Performance was relatively similar across sessions for subjects participating in more than one session (Fig. 3.9), suggesting that learning was generally not a factor in the task.
Figure 3.8: Hit rates by level of frequency deviation. For each subject, we plot the hit rate as a function of the proportion of frequency increase for $D_L$ trials (top), $D_G$ trials (middle), and $D_{LG}$ trials (bottom). Colors of traces and data points follow the color conventions from Fig. 3.7 to identify individual subjects.

Figure 3.9: Performance as a function of session number. For each subject, we plot the performance on standard trials ($1 - \text{false-alarm rate}$; top) and the performance on deviant trials (hit rate; bottom) as a function of test session. Colors of traces and data points follow color conventions from Fig. 3.7 and 3.8.
Power and phase activity reliably reflects deviant detection

To assess whether ECoG activity differentially detected deviations in the spectrotemporal regularities of the standard sequence, we tested whether a linear H v. CR classifier could differentiate ECoG activity during hit trials (i.e., correctly reported deviants) from activity during correct-rejection trials (i.e., correctly reported standard trials). Specifically, we measured single-electrode classification accuracies in the wideband power and phase responses after we filtered each electrode’s ECoG activity into 7 major frequency bands (see METHODS). We found that significant grand-mean classification accuracy occurred at around ~400 ms following final-triplet onset (Fig. 3.10, top; signed-rank test, FDR-corrected ps<0.05). Classification accuracy in phase generally did not become significant until stimulus offset (>600 ms; Fig. 3.10, bottom; signed-rank test, FDR-corrected ps<0.05). We also found that we could achieve similar classification results if we averaged power and phase responses into 100-ms bins (compare gray data points and black traces in Fig. 3.10). Because of the similarity in these results, subsequent classification analyses in power and phase averaged responses in 100-ms bins.

Figure 3.10: Results of the H vs. CR classifier analysis. Single-electrode classification accuracies across subjects for ECoG power (top) and phase (bottom) for discriminating hits versus correct rejections as a function of time relative to final-triplet onset. Black traces and gray shading depict grand-mean (± standard error) classification accuracy across subjects using all electrodes. Black bars above traces depict time points for which classification accuracy was reliably better than chance (signed-rank tests, FDR-corrected ps<0.05). Gray data points and traces depict classifier results when power or phase responses were binned into 100-ms time bins.
**ECoG activity discriminates stimulus- and behavioral-report-related characteristics**

Although this first linear-classifier analysis showed that ECoG activity reliably discriminated between deviant and standard stimuli, it confounded activity differences due to stimulus characteristics (i.e., standard vs. deviant stimuli) and behavioral report (i.e., reporting hearing standard vs. deviant stimuli). Because both behavioral report and stimulus characteristics could account for the observed differences, we conducted three sets linear-classifier analyses to differentiate these dual effects. In the first analysis (the ‘stimulus’ classifier), we classified single-electrode ECoG activity by trial type (e.g., standard vs. deviant), independent of a subjects’ behavioral reports. In the second analysis (the ‘behavioral’ classifier), we classified single-electrode ECoG activity by report, independent of trial type. And in the third analysis, we classified single-electrode ECoG activity by report, separately for each trial type (‘behavioral × stimulus’ classifier).

The results from the *stimulus* classifier were similar to those from the *H v. CR* classifier (Fig. 3.11). The raw ECoG signal began to reliably discriminate between standard versus deviant stimuli at ~400 ms after final-triplet offset, which was still within the stimulus period (Fig. 3.11a, left; sign-rank tests, FDR-corrected ps<0.05). However, reliable classification was not evident at the time of behavioral report (Fig. 3.11a, right). Similar results were found in the power and phase responses (Fig. 3.11b and 3.11c). In power (Fig. 3.11b, left), classification accuracy reliably increased from before final-triplet onset to after (signed-rank test, p<0.05). This was also the case in phase (Fig. 3.11c, left), with better-than-chance classification accuracy beginning at around 400 ms (signed rank test, ps<0.01, uncorrected). Once again, power and phase responses did not differentiate stimulus type at the time of behavioral report (Fig. 3.11b and 3.11c, right).
Figure 3.11: Results of the stimulus classifier analysis. (a) Single-electrode classification accuracies across subjects for the raw ECoG signal, aligned relative to final-triplet onset (left) or response (right). Traces and shaded regions depict grand-mean (± standard error) classification accuracy across subjects using all electrodes. Black bars above traces depict time points for which classification accuracy was reliably better than chance (signed-rank tests, FDR-corrected ps<0.05). (b)-(c) Single-electrode classification accuracies for power (b) or phase (c), aligned relative to final-triplet onset (left) or response (right). Data and error bars depict grand-mean (± standard error) classification accuracy across subjects using all electrodes. Individual time points with classification accuracies reliably different from chance are indicated by the symbol ‡ (individual signed-rank tests, ps<0.01, uncorrected). For all plots, brackets denote results of a comparison either between mean classification accuracy before and after final-triplet onset (left panels) or between the time periods -1.2 – -0.4 s and -0.4 – +0.2 s relative to response. Significant comparisons are indicated by the symbol † (signed-rank tests, ps<0.05). Abbreviations: H, hit; M, miss; FA, false alarm; CR, correct rejection.
Figure 3.12: Results of the stimulus classifier after first separating deviants by type. (a)-(c) Single-electrode classification accuracies across subjects for the raw ECoG signal, aligned relative to final-triplet onset (left panels) or response (right panels), using $D_L$ trials only ((a)), $D_G$ trials only ((b)), and $D_{LG}$ trials only ((c)). Traces and shading follow conventions from Fig. 3.11a. (d)-(f) Respective single-electrode classification accuracies for each discrimination in power. (g)-(i) Respective single-electrode classification accuracies for each discrimination in phase. For all plots, statistical testing follows same conventions as in Fig. 3.11. Abbreviations: $H_L$, hits on $D_L$ trials; $M_L$, misses on $D_L$ trials; $H_G$, hits on $D_G$ trials; $M_G$, misses on $D_G$ trials; $H_{LG}$, hits on $D_{LG}$ trials; $M_{LG}$, misses on $D_{LG}$ trials.

Classification performance was not as reliable when we trained a different stimulus classifier to discriminate between standard stimuli and each type of deviant stimulus separately (Fig. 3.12). We could not identify reliable classification performance in the raw ECoG signal for any deviant type (Fig. 3.12a-c). In power and phase, there was a general trend of increasing classification accuracy with time when aligned to final-triplet onset for each deviant type (Fig. 3.12d-i, left). These trends were significant in power for global deviants and in phase for local and global deviants, where classification accuracies after final-triplet onset were greater than accuracies preceding onset (signed rank tests, $p<0.05$). Again, classification accuracy was at chance when neural activity was aligned to behavioral report (Fig. 3.12d-i, right).
Next, we assessed whether neural activity could reliably differentiate between subjects’ behavioral reports, independent of stimulus type. The behavioral classifier’s performance on the ECoG signals was at chance levels when trained to discriminate between reports of standards and deviants, using the ECoG signals. (Fig. 3.13a). In power and phase, however, classification
accuracy increased following final-triplet onset (signed rank tests, $p_{S}<0.05$); Fig. 3.13b and c, left). Classification accuracy also became significant around 400 ms prior to report (signed rank tests, $p_{S}<0.01$, uncorrected; Fig. 3.13b, right).

This general pattern of results was less reliable when we trained separate classifiers to differentiate between standard reports and deviant reports using only deviant trials of a particular deviant type (Fig. 3.14). Raw ECoG activity did not discriminate between reported standards or reported deviants, for any deviant type (Fig. 3.14a-c). When aligned to final-triplet onset, classification results for power and phase was generally not better than chance (Fig. 3.14d-i, left). However, when aligned to report, we found that classification accuracy of power responses became significant up to 400 ms prior to report for the local and global deviant types (signed rank tests, $p_{S}<0.01$, uncorrected; Fig. 3.14d and e, right). We also found that classification accuracy of phase responses became significant around 200–300 ms prior to behavioral report for the local and local+global deviant types (signed rank tests, $p_{S}<0.01$, uncorrected; Fig. 3.14g and i, right).

**Figure 3.14:** Results of the **behavioral classifier** after first separating by deviant type. All conventions follow from Fig. 3.12.
Figure 3.15: Results of behavioral x stimulus classifier. Plots (a), (c) and (e) depict classification accuracies of behavioral report for raw ECoG, power, and phase on standard trials only, respectively. Plots (b), (d) and (f) depict classification accuracies by behavioral report for raw ECoG, power, and phase on deviant trials only, respectively. Otherwise, plots follow same conventions as in Fig. 3.11.

These results were generally consistent after first separating deviant trials by the type of deviant (Fig. 3.16). The raw ECoG signal generally did not differentiate between hits and misses (Fig. 3.16a-c), although there was a significant increase in classification accuracy on global-deviant trials following final-triplet onset compared to prior to onset (signed-rank test, \( p < 0.05 \); Fig. 3.16b, right). In the global and local+global deviant trials, classification accuracy in power also increased following deviant onset (signed-rank tests, \( p s < 0.05 \); Fig. 3.16d and 3.16e, left). Classification accuracy also increased for phase modulations following deviant onset in local and global deviant trials (signed-rank tests, \( p s < 0.05 \); Fig. 3.16f and 3.16g, left). These differential modulations were not consistently evident at the time of behavioral report, but occasional time points did reach significance (signed-rank tests, \( p s < 0.01 \), uncorrected; Fig. 3.16d-i, right).
Figure 3.16: Results of behavioral x stimulus classifier after first separating by deviant type. Plots (a), (d), and (g) depict classification accuracies of hits versus misses on $D_t$ trials only for raw ECoG, power, and phase, respectively. Plots (b), (e), and (h) plot the respective classification accuracies of hits versus misses on $D_G$ trials only. Plots (c), (f), and (i) plot the respective classification accuracies of hits versus misses on $D_{LG}$ trials only. Otherwise, plots follow same conventions as in Fig. 3.11. Note: classifier results on standard trials are not included because they do not depend on deviant type and, as such, are the same as in Fig. 3.15.

Multiple brain regions exhibit reliable modulations with respect to stimulus and report

To test for brain-specific differences in classification accuracy, we segregated electrodes by brain region and computed each region’s single-electrode classification accuracy. For each brain region, we calculated the across-subject mean classification accuracy for the periods 400-600 ms after final-triplet onset and -400-0 ms preceding behavioral report. Because classification accuracy was more robust in the analyses that combined across deviant types, we focused this regional analysis on those results.
Figure 3.17: Time-averaged results from *stimulus classifier* by brain region. Plots depict time-averaged mean (± standard error) classification accuracies as a function of brain region for raw ECoG (a), power (b), and phase (c) activity. Left panels in each plot depict accuracies from activity aligned relative to the stimulus, right panels depict accuracies from activity aligned relative to response. Significance of classification accuracies are as follows: *, signed-rank tests with FDR-corrected $p < 0.05$ across brain regions; ‡, singed-rank tests with $p < 0.01$, uncorrected; †, signed-rank tests with $p < 0.05$, uncorrected.

When aligned to the stimulus, we found that classification accuracies from number of regions could reliably discriminate between standard and deviant stimuli (Fig. 3.17). Both regions in the temporal cortex (in raw ECoG activity) and the medial temporal love (in phase) exhibited significant classification accuracies (signed-rank tests, FDR-corrected $p < 0.05$; Fig. 3.17a and c,
left). Additionally, assuming a slightly less stringent significance criterion, we found that a number of other brain regions discriminated between stimulus type. These include: middle and inferior temporal cortices (“OtherTemporal” in phase; signed-rank test, \( p<0.01 \), uncorrected), medial temporal lobe and occipital cortex (ECoG; signed-rank tests, \( p<0.05 \), uncorrected), superior temporal cortex (in power; signed-rank test, \( p<0.05 \), uncorrected), and the supramarginal gyrus (in power and phase; signed-rank test, \( p<0.05 \), uncorrected) (see Fig. 3.17a-c).

When aligned to behavioral report, we generally could not identify any region in which we obtained significant classification accuracy (Fig. 3.17a-c, right). Though, using a less stringent significance criterion, the sensorimotor and orbitofrontal cortices reliably exhibited significant classification accuracy in the raw ECoG signal and phase, respectively (signed-rank tests, \( p<0.05 \), uncorrected; Fig. 3.17a and 17c, right).

In contrast to the results from the stimulus classifier, with the exception of the dorsolateral prefrontal cortex (phase; signed-rank test, FDR-corrected \( p<0.05 \); Fig. 3.18a-c, left), we could not identify any brain regions that had significant classification accuracy for behavioral report, when data were aligned relative to the stimulus (Fig. 3.18a-c, left). Using a less stringent significance criterion, a classifier could decode behavioral reports in the temporal cortex (ECoG; signed-rank test, \( p<0.05 \), uncorrected; Fig. 3.18a, left) and in the dorsolateral prefrontal and superior parietal cortices (power; signed-rank tests, \( p<0.05 \), uncorrected; Fig. 17b, left).

When aligned to behavioral report, however, we found significant classification accuracy in the sensorimotor cortex, inferior parietal cortex, and the supramarginal gyrus (power; signed-rank tests, FDR-corrected \( p<0.05 \); Fig. 3.18b, right). Using a less stringent significance threshold, classification accuracy was significant in the superior parietal cortex and “OtherTemporal” (power; signed-rank tests, \( p<0.05 \), uncorrected; Fig. 3.18b, right) and in the dorso- and ventro-lateral prefrontal and inferior and superior parietal cortices (phase; signed-rank tests, \( p<0.05 \), uncorrected; Fig. 3.18c, right).
Finally, the results from the **behavioral x stimulus** classifier indicate chance performance discriminating false alarms and correct rejections during standard trials for all brain regions when aligned to the stimulus (Fig. 3.19a-c). In contrast, significant classification accuracy when aligned to behavioral report was found in sensorimotor cortex, supramarginal gyrus, inferior parietal cortex, and ventrolateral prefrontal cortex (see Fig. 3.19b and c, left). During deviant trials, a similar set of
brain regions exhibited significant classification accuracies discriminating hits and misses, but this time both when aligned relative to the stimulus and when aligned relative to report (see Fig. 20).

Figure 3.19: Time-averaged results from *behavioral x stimulus* classifier by brain region on standard trials. All plotting conventions follow from Fig. 3.17.
Behavioral performance on detection task correlates with linear-classifier performance

Given the large variability in individual performance on the detection test (see Fig. 3.7), we tested whether individual performance affected classification accuracy. To test whether behavioral performance affected the linear-classifier analyses, we first correlated false-alarm rates with grand-mean classification accuracies on standard trials. We found that classification accuracy in power (aligned relative to behavioral report) was negatively correlated with false-alarm rate ($r^2 = -0.52$, $p < 0.05$).
\( p < 0.05; \text{ Fig. 3.21}). \) In other words, poor performance on the task was associated with poor classification performance.

To further test this hypothesis, we sorted subjects into two groups based on performance on standard trials: good performers, who performed significantly above chance on standard trials; and poor performers, who performed at chance or worse on standard trials. We then computed the across-subject mean classification accuracies in \textit{behavioral x stimulus} analysis for each group to see how task performance affected classification accuracy.

We found that, in general, significant classification accuracies occurred only for the good performers (Fig. 3.22). For example, for good performers, classification accuracy increased in both power and phase following final-triplet onset on deviant trials (Fig. 3.22b and 3.22d, left). In contrast, we could not identify any reliable changes in classification accuracy in the poor performers. Although classification accuracies in power seemed to increase on standard trials when aligned to both stimulus and response (see Fig. 3.22a), we could not identify any individual time point for which classification accuracy was significantly better than chance. These results suggest that poor performance hindered our ability to detect reliable modulations with respect to behavior.

\textbf{Figure 3.21: Correlation between behavioral performance and classification accuracy.} Plot depicts false-alarm rate versus whole-brain classification accuracy on standard trials for each subject. Black data points correspond to classification accuracies from power activity, gray data points correspond to accuracies from phase activity. Black and gray lines depict linear best-fit line to power and phase data, respectively. The \( r^2 \) and significance of the correlations are depicted in the figure using the same color shading.
Figure 3.22: Comparison of behavioral x stimulus classifier results between good and poor performers. For all plots, good performers are depicted in black and poor performers are depicted in gray. Otherwise, plots follow same conventions as Fig. 3.15. Results of statistical tests also follow conventions from Fig. 3.15, except that black versus gray shading denotes significance tests for good versus poor performers, respectively. Note: significance tests are conducted within subject groups against chance performance, not across groups.

Direct tests on power and phase spectra only reliably distinguish stimulus characteristics

Finally, given that the classifier analyses suggested that both power and phase responses reliably distinguished between both stimulus characteristics and behavioral report, we computed direct tests on the wideband power and phase-alignment spectra to determine which neural frequencies may have contributed to the classification results. We conducted these analyses in three ways, similar to the classification analyses: (1) a stimulus analysis, in which we tested how wideband power and phase alignment was modulated by stimulus type (i.e., standard versus deviant); (2) a behavioral analysis, in which we tested how power and phase were modulated by behavioral report (i.e., reported standards versus reported deviants); and (3) a behavioral x stimulus analysis, in which we tested for modulations as a function of behavioral report separately for each stimulus type (i.e., FA versus CR and H versus M). For each of these analyses, we used a wavelet decomposition to compute the instantaneous power and phase as a function of frequency.
for each electrode (see METHODS) and then aligned neural activity relative to final-triplet onset. We then computed the average power and PPC spectra for each frequency during the final 200 ms of the stimulus period, corresponding to the time period that exhibited significant classification accuracies in the linear classifier analyses. Finally, for each electrode, we computed the difference in power or PPC spectra between testing conditions for each analysis. For example, for the stimulus analysis, we computed the difference in power and PPC spectra between deviant trials and standard trials. We also repeated these three analyses after first separating deviants by type (i.e., local, global, local+global), as we did for the classifier analyses. The results shown below reflect the whole-brain averages of these difference spectra for each subject.

Across subjects, deviant trials eliciting higher power in low-frequency (~0.5-1 Hz) and lower power in mid-frequency (~16-20 Hz) compared to standard trials (Fig. 3.23a; signed-rank tests, FDR-corrected ps<0.05). In contrast, there were generally not any differences in phase alignment between deviant and standard trials (Fig. 3.23c). With respect to behavioral report, we also could not identify any reliable modulations in either power or phase alignment (Fig. 3.23b and 23d). Finally, we could not identify any reliable modulations in the behavioral x stimulus analysis (Fig. 3.24).
Figure 3.23: Wideband differences in power and PPC as a function of stimulus and behavior. Plots (a) and (c) depict grand-mean difference spectra between deviant and standard (i.e., deviant minus standard) stimuli in power and PPC, respectively. Plots (b) and (d) depict grand-mean difference spectra between reported deviants and reported standards (i.e., reported deviants minus reported standards) in power and PPC, respectively. For all plots, thin gray traces reflect grand-mean difference spectra for each subject. Thick black traces reflect across-subject mean difference spectra. Significance of difference values are as follows: *, signed-rank tests with FDR-correct $p<0.05$ across frequencies; ‡, singed-rank tests with $p<0.01$, uncorrected; †, signed-rank tests with $p<0.05$, uncorrected.
Figure 3.24: Wideband differences as a function of behavior on standard and deviant trials. Plots depict grand-mean difference spectra between reported deviants and reported standards either on standard trials only ((a) and (c), respectively) or on deviant trials only ((b) and (d), respectively). Otherwise, plotting conventions follow from Fig. 3.23.

After first separating deviant trials by type, we repeated the above analyses to test whether the different deviants elicited distinct modulations with respect to stimulus or behavioral characteristics. We found that, for each analysis, each deviant type elicited a similar pattern of modulations (Fig. 3.25). We could not identify any differences in the modulations among deviant types.
Figure 3.25: Results of wideband-difference analyses after first separating by the type of deviant stimulus. Plots (a) and (b) depict difference spectra between standard trials and each type of deviant trial in power and PPC, respectively. Color of traces denote type of deviant: red for \(D_L\), blue for \(D_G\), and purple for \(D_{LG}\). Plots (c) and (d) depict difference spectra in power and phase, respectively, by behavioral report only (i.e., independent of stimulus type) using all standard trials and all deviant trials of the same deviant type. Color of traces denote type of deviant included in each analysis: red for \(D_L\) trials and all standard trials, blue for \(D_G\) trials and all standard trials, and purple for \(D_{LG}\) trials and all standard trials. Plots (e) and (f) depict difference spectra between reported standards and reported deviants, separately for each stimulus type. Color of traces denote type of deviant: red for \(D_L\), blue for \(D_G\), purple for \(D_{LG}\), and black for standard trials. For all plots, traces and shading depict the mean (± standard error) across-subject, grand-mean difference spectra. Significance-testing conventions follow from Fig. 3.24, with the color of the symbols corresponding to the type of deviant tested. Note: significance tests are conducted within trial-type groups against zero, not across trial-type groups.
DISCUSSION

This study had three main goals. First, we tested if the standard sequence of tone triplets modulated the power and phase of neural oscillations. Second, we tested if stimulus differences (i.e., standard versus deviant stimuli) and/or choice differences (i.e., reported standards versus reported deviants) modulated the power and phase of these oscillations. Third, we tested if a listener’s ability to detect a particular type of deviant was correlated with changes in the frequency band corresponding to time scale of the deviants. We found that the standard sequence modulated ECoG power across cortex but were generally not in the frequency bands that corresponded to the regularities’ time scales. In contrast, changes in the ECoG phase corresponded to the time scale of the regularities. Further, these changes were only evident in specific regions of cortex. Finally, a series of linear classifiers were able to discriminate both stimulus- and choice-related information that was contained in ECoG power and phase.

Power and phase-alignment are differentially modulated by the standard sequence in cortex

Although power modulations were visible across cortex (see Fig. 3.4), modulations in phase alignment were restricted to a subset of regions (see Fig. 3.5), including dorso- and ventro-lateral prefrontal cortex, sensorimotor cortex, the supramarginal gyrus and inferior parietal cortex, and multiple regions in the temporal lobe. Moreover, whereas the power modulations generally did not relate to the time scales of the spectrotemporal regularities, modulations in phase alignment were regularity specific. This widespread activation of the cortex in response to acoustic stimulation is similar to previous findings (Edwards et al., 2005; 2009; Hsiao et al., 2009; Ishii et al., 2009; Besle et al., 2010; Hsiao et al., 2010; Dykstra et al., 2011; Chennu et al., 2013; Golumbic et al., 2013; Eliades et al., 2014). These widespread neural modulations may reflect a mechanism to facilitate the coupling neural processes across cortex to acoustic stimulation, such as for use in
multisensory representations; see (Foxe and Schroeder, 2005; Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007).

Our finding that phase codes the stimulus regularities is consistent with previous findings (Patel and Balaban, 2000; Luo et al., 2006; Lakatos et al., 2008; Besle et al., 2010; Lakatos et al., 2013; Henry et al., 2014). Our results extend these findings by determining that these frequency-specific phase modulations occur not only in auditory cortex, but also across regions along the ventral and dorsal auditory pathways, which are thought mediate auditory perception and audiomotor behaviors, respectively (Rauschecker, 2011; Bizley and Cohen, 2013; Christison-Lagay et al., 2015). We also found that regions outside of these two pathways reliably exhibited frequency-specific phase alignment, particularly the sensorimotor cortex and regions in the medial temporal lobe.

Multiple lines of evidence lend support to the notion that our identified regions with reliable phase modulations are involved in regularity representation. First, damage to the dorsolateral prefrontal (dIPFC) and parietal cortices reduces the amplitude of the mismatch negativity, an automatic brain response reflecting a detected stimulus change from a commonly presented stimulus (Alho et al., 1994; Alain et al., 1998). Because the commonly presented stimulus creates a spectrotemporal regularity, it is unsurprising that those regions involved in change detection also reflect the presence of spectrotemporal regularity itself. Second, a growing body of literature suggests that the regions of the inferior parietal cortex, in particular, are critically associated with spectrotemporal processing and perceptual organization (Giraud et al., 2000; Cusack, 2005; Dykstra et al., 2011; Teki et al., 2011; 2016), which would necessarily require them to process information related to the regularities in acoustic stimuli. Third, both the dIPFC and motor cortices are important for speech perception and production (Kotz and Schwartze, 2010; Morillon et al., 2015): in the dIPFC, temporal information is integrated with memory representations to optimize comprehension; and in motor cortex, temporal-structure information from the supplementary and pre-motor areas guides articulation. Also, there is evidence to suggest that the motor system is
important for perceptual processes requiring temporal predictions, such as in beat and rhythm processing (Schubotz, 2007; Schroeder et al., 2010; Arnal and Giraud, 2012; Grahn and Rowe, 2013; Morillon et al., 2015). Finally, the medial temporal lobe may be involved in the integration of complex temporal patterns and statistical learning (Turk-Browne et al., 2009; Aly et al., 2013; Geiser et al., 2014; Schapiro et al., 2014; Garrido et al., 2015; Barascud et al., 2016), which could be useful in spectrotemporal regularity processing.

**Preliminary findings in detection task suggest power and phase modulations distinguish stimulus and behavioral characteristics in detection task**

Our finding that listeners’ behavioral reports modulated the power and phase responses prior to the time of stimulus offset is consistent with the idea that these responses reflect perceptual differences rather than preparatory motor activity. Preparatory-motor-related activity should occur around the time of the behavioral report (i.e., the time of the keyboard press). Indeed, when we aligned neural data relative the behavioral report, we found modulations in sensorimotor (i.e., primary motor and sensory cortices) cortex. These modulations are likely reflecting the different motor actions required for each key press (see Figs. 3.18–3.20).

Choice-related modulations in power were evident across a number of brain regions, including temporal cortex, inferior parietal cortex, orbitofrontal cortex, and ventrolateral prefrontal cortex (vIPFC) (see Fig. 3.20). Temporal, inferior parietal, and vIPFC prefrontal cortices have each been implicated in auditory perceptual organization (Romanski et al., 1999; Kaas and Hackett, 2000; Romanski and Goldman-Rakic, 2002; Hickok and Poeppel, 2007; Bizley and Cohen, 2013; Garell et al., 2013; Fukushima et al., 2014; Christison-Lagay et al., 2015), suggesting that these modulations may reflect feedforward or feedback signals that underlie auditory perceptual decision-making. Also, our finding that the orbitofrontal cortex was modulated is also consistent with previous work: this cortical region has been implicated in a variety of decision-making tasks (Bechara et al.,
2000; Rolls, 2004; Schoenbaum and Roesch, 2005) and may related to choice-related outcome expectancies (Rolls, 2004).

The differences in classification accuracy between power and phase responses may relate to how we temporally averaged the signals. Because power modulations occur more slowly over time, averaging power estimates over relatively long time bins may not differentially affect responses across frequency bands. In contrast, because phase responses vary over the course of a single oscillatory cycle, it is possible that phasic temporal averaging may systematically reduce information as a function of increasing frequency band. Thus, future analyses should perform a more fine-grained temporal analysis of the power and phase responses.

Similarly, it is important to note that subject-to-subject variability in task performance made it difficult to assess the extent to which we could identify reliable choice-related modulations. Only 7 of the 17 subjects performed significantly better than chance on standard trials, suggesting that the majority of the subjects either had extreme difficulty with the task or simply did not understand it. These differences in task performance directly correlated with the classification performance in the linear-classifier analyses (see Figs. 3.21 and 3.22). And of the 7 subjects who performed well, only 5 of them completed more than 1 session of the task. This lack of subjects with multiple sessions likely also made it difficult to identify power and phase modulations that distinguish between each type of deviant.

Finally, although we have interpreted the results of the choice-related classifier analyses as evidence that neural oscillations encode behavioral choice, it is possible that these findings were partly influence by stimulus differences as well. As a reminder, the classification analyses either combined across deviant types (for the combined analysis) or collapsed across levels of frequency deviation (for the separate analyses for each deviant type). The fact that different brain areas seemed to be modulated by stimulus and report characteristics (see Figs. 3.17-3.20) suggest that the choice-related classifiers at least partially reflected choice rather than pure stimulus differences. Nonetheless, fully distinguishing between purely sensory versus purely perceptual representations
will likely require a modification in the stimulus design to more accurately determine deviant-detection thresholds for each subject such that we can choose a single level of deviation for each deviant type.

**Future directions for further study**

Despite the qualifications described above, the preliminary analyses strongly suggest that we can identify a number of distinct brain regions that reflect the spectrotemporal regularities in the standard sequence and reflect the stimulus versus perceptual characteristics of detected deviations. Future work on this task should focus on a few improvements to the task design, as well as acquiring more subjects who perform well in the task and focusing specific regions of interest.

Because one of the main issues with the current design was the overall low number of trials for each deviant type, it will be beneficial in future work to reduce the number of deviant types from 3 to 2 by removing the local deviant variant. This would still allow us to distinguish between neural modulations in response to deviations on the global time scale (i.e., the global deviants) and those in response to deviations on the local time scale. Moreover, it will allow for ~50% more trials of each remaining deviant types per session, thus substantially increasing the statistical power for each deviant type. Finally, subjects should run enough sessions to reach at least 150 trials of each deviant type to increase the likelihood of detecting differential effects of each deviant type. This can be achieved either with the current session duration, meaning subjects should perform a minimum of ~10 sessions, or by increasing the number of trials per session.

Another issue with the current design was the titration of deviation levels across sessions, which potentially confounded the ability to distinguish choice-related activity. To overcome this issue, future subjects should perform an initial calibration session in which we measure each subject’s frequency deviation thresholds for each deviant type and determine an appropriate level of frequency increase for each deviant type that will be used on subsequent testing sessions. This
calibration can be achieved by the staircase method (Levitt, 1971; Gifford et al., 2014), which is a popular approach to testing sensory discrimination thresholds. Additionally, future sessions should remove trial-by-trial feedback in case it may improve discrimination thresholds over time (Campbell and Small, 1963).

A third issue relates to assessing task performance. Behavioral performance on standard trials should be assessed for each subject to ensure that they understand the task and are performing reliably better than chance. This could be simultaneously achieved during the calibration session by including standard trials and ensuring that subjects only report hearing deviants reliably when presented with deviant stimuli. Only subjects that perform well on standard trials should be included in subsequent test sessions. Based on the current results, we expect to need an additional 5-10 subjects that perform well on the task in order to test region specificity of neural responses and reliability across subjects.

Finally, the preliminary results suggest that, if possible, future work should focus on subjects with electrodes in inferior parietal cortex, vIPFC and dIPFC, orbitofrontal cortex, and the temporal cortex. Focusing on these regions that showed reliable modulations to the standard pattern and to behavior will increase the likelihood of further determining the extent to which these regions are responsible for the sensory and behavioral correlates of spectrotemporal regularity representation.
CHAPTER 4

4. Characterizing the impact of category uncertainty on human auditory categorization behavior


ABSTRACT

Categorization is an important cognitive process. However, the correct categorization of a stimulus is often challenging because categories can have overlapping boundaries. Whereas perceptual categorization has been extensively studied in vision, the analogous phenomenon in audition has yet to be systematically explored. Here, we test whether and how human subjects learn to use category distributions and prior probabilities, as well as whether subjects employ an
optimal decision strategy when making auditory-category decisions. We asked subjects to classify the frequency of a tone burst into one of two overlapping, uniform categories according to the perceived tone frequency. We systematically varied the prior probability of presenting a tone burst with a frequency originating from one versus the other category. Most subjects learned these changes in prior probabilities early in testing and used this information to influence categorization. We also measured each subject's frequency-discrimination thresholds (i.e., their sensory uncertainty levels). We tested each subject's average behavior against variations of a Bayesian model that either led to optimal or sub-optimal decision behavior (i.e., probability matching). In both predicting and fitting each subject's average behavior, we found that probability matching provided a better account of human decision behavior. The model fits confirmed that subjects were able to learn category prior probabilities and approximate forms of the category distributions. Finally, we systematically explored the potential ways that additional noise sources could influence categorization behavior. We found that an optimal decision strategy can produce probability-matching behavior if it utilized non-stationary category distributions and prior probabilities formed over a short stimulus history. Our work extends previous findings into the auditory domain and reformulates the issue of categorization in a manner that can help to interpret the results of previous research within a generative framework.

**AUTHOR SUMMARY**

Categorization is an important cognitive process that allows us to simplify, extract meaning from, and respond to objects in the sensory environment. However, categorization is complicated because an object can belong to multiple categories. Thus, to inform our categorical judgments, we must make use of prior information. Given the importance of categorization, we hypothesized that humans utilize optimal strategies for making categorical judgments that allow us to minimize
categorization errors. We found, though, that whereas subjects used prior information (i.e., category prior probability), they were sub-optimal in their categorization behavior. This seems to be common in other perceptual and cognitive tasks as well. We then explored the bases for this sub-optimal behavior and found that it can be consistent with an optimal strategy if we assume that subjects have trial-by-trial noise in components of the judgment process. This work extends previous similar findings into the field of auditory categorization and provides a means to reinterpret previous results.

INTRODUCTION

Categorization is a natural and adaptive process that allows the brain to organize the typically high-dimensional and continuous sensory information into robust hierarchical and discrete representations. These discrete representations, or categories, are a means to mentally manipulate, reason about, and respond to objects in our environment (Grinband et al., 2006). For instance, in auditory perception, humans and other animals can ignore the natural acoustic variability that exists between different utterances of the same vocalization in order to differentiate one type of vocalization (e.g., a howl) from a second type (e.g., a bark). In other situations, listeners can use this variability to identify one caller (e.g., Lassie) from another (e.g., Benji).

The perceptual ease with which we can categorize sound belies the complex computations underlying this ability. One reason categorization is complex is that a sensory property may be ambiguous with respect to the stimulus’ category membership. For example, because both dogs and wolves can produce howls, the acoustic structure of the howl by itself may not provide enough information to the listener for proper identification of the caller. In such cases, and in the absence of other sensory information, the listener needs to rely on other sources of information to correctly categorize a sound and identify whether the howl came from a dog or a wolf. This information can
be prior knowledge such as knowing that the probability of encountering a wolf is low. Since prior information is subjective, it is of fundamental interest to understand the degree to which an observer acquires this information and then uses it to perform categorical judgments.

The utility of prior information in visual categorization has been well studied (Lee, 1963; Lee and Janke, 1964; 1965; Ulehla, 1966; Healy and Kubovy, 1981; Bohil and Maddox, 2001; Hansen et al., 2012a; 2012b). In comparison, our understanding of how prior information informs categorical judgments in audition is relatively limited and has only more recently become an active area of research (Sullivan et al., 2005a; 2005b; Holt and Lotto, 2006; Ley et al., 2012; Scharinger et al., 2013). More importantly, auditory categorization has not been tested or modeled in situations in which the auditory stimulus is ambiguous with regard to its category membership. Understanding auditory-categorization behavior is important for differentiating between modality-specific versus modality-general computational strategies, which can provide insights into the underlying neural computations.

In particular, categorization can be understood as the result of a probabilistic inference process in which the observer combines sensory and prior information according to their relative levels of uncertainty (noise) (Knill and Richards, 1996). Bayesian statistics is a useful mathematical framework to formulate generative models for such categorical inference processes. However, it requires a precise quantification of the different levels of uncertainty in order to provide behavioral predictions that allow for unique model interpretations. For example, different decision strategies can lead to very similar model predictions if the sensory noise levels are allowed to be free parameters.

The purpose of this study was two-fold: (1) to test whether human subjects can learn and use category-prior information when making auditory categorical judgments and (2) to carefully constrain and validate a generative Bayesian model of auditory categorization against experimental data. To this end, we developed a novel auditory categorization task that required subjects to categorize the frequency of a tone burst into one of two overlapping categories (“A” or “B”). We
systematically varied the prior probability of choosing a frequency from category “A” or “B” in different blocks of the experiment. Furthermore, we determined each subject’s sensory uncertainty by measuring individual frequency-discrimination thresholds. Based on these uncertainty measurements, we formulated a Bayesian model to individually quantify how well each subject learned the categorical priors (i.e., the category distributions and prior probabilities) and to test whether subject’s employed an optimal decision strategy. We found that most subjects appropriately learned the different category prior probabilities, yet showed some variability and uncertainty in the shape of the learned category distributions. Furthermore, given the measured sensory uncertainty during the experiment, subjects’ overall behavior was more consistent with probability matching rather than an optimal decision strategy for category choice. Further analyses indicated that overall probability-matching behavior could emerge if, trial-by-trial, subjects employed an optimal decision strategy and assumed non-stationary categorical priors.

METHODS

Ethics statement

All subjects participated in a purely voluntary manner, after providing informed written consent, under the protocols approved by the Institutional Review Board of the University of Pennsylvania.

Experimental setup

Six subjects (two female) participated in two tasks: (1) a discrimination task that estimated each subject’s frequency-discrimination thresholds and (2) an auditory-categorization task that tested how each subject used category-prior information. Both tasks were conducted in a darkened anechoic chamber (2 m × 1.5 m, Industrial Acoustics Company, Inc.), which housed a chair for the
subject, a gamepad, a table mounted with an LCD computer screen (P190S, Dell, Inc.), a speaker (MSP7, Yamaha, Inc.), and a chin rest. The speaker was positioned ~0.1 m below a subject’s ears when his/her head was placed on the chin rest. The gamepad registered the subject’s responses during each task. Both the discrimination and categorization tasks were designed and implemented in MATLAB (version R2010b) with the Tower-of-Psych and Snow-Dots packages (freely available resources (Goldstone, 1998; Heasley and Gold, 2009)). For both tasks, the stimuli were 750-ms tone bursts (10-ms $\cos^2$ ramp; frequency range: 500 – 5550 Hz). The tone frequencies were distributed uniformly in $\log_{10}$ units. Stimuli were synthesized with an RX6 Multifunction Processor (Tucker-Davis Technologies, Inc.) with a sampling rate of 25 kHz and were presented at 65 (± 3) dB SPL.

**Discrimination task and analysis**

Each subject participated in a two-interval, two-alternative forced choice frequency-discrimination task. This task measured each subject’s frequency-discrimination threshold at eight different “standard” frequencies, which were distributed between 500-5550 Hz: 794, 1260, 2297, 2639, 3031, 3482, 4462, and 4976 Hz. A trial began with a visual “GO” cue on the computer screen, followed by the presentation of the first tone burst. After a 1000-ms delay, the second tone burst was presented. Following offset of this second tone burst, the subject had 2000 ms to report which tone burst had the higher frequency. Subjects only received feedback (in the form of a yellow circle on the computer screen) when a response was not made within the allotted response window.

In each trial, one tone burst was one of the standard frequencies, whereas the other “comparison” tone burst had a different frequency. We used a 2-up-1-down adaptive staircase procedure (Levitt, 1971) to adjust the frequency of the comparison tone across trials. On a trial-by-trial basis, the order of the standard and comparison tone bursts was randomized, as well as the choice of the standard tone burst. Each subject participated in 2-4 experimental sessions. Each
session consisted of two blocks of trials; each block contained 30 or 40 trials per standard tone frequency (320 or 480 total trials).

The data for each subject were collapsed across sessions and only trials in which a response was made within the allotted response window were included in subsequent analyses. We computed a psychometric function representing the probability that the subject reported the comparison tone ($v_{comp}$) as higher than the standard tone ($v_{stand}$). Since the values of $v_{comp}$ varied across subject and session, $v_{comp}$ values were binned into five equidistant bins (in $\log_{10}$ units) for each $v_{stand}$ and subject. Each subject’s psychometric functions (i.e., one function for each standard tone frequency) were fit with a cumulative Gaussian with free parameters $\mu$ and $\sigma$ using a maximum-likelihood fitting procedure to the raw data.

We assumed that a subject’s discrimination process was the result of a comparison between the frequencies of the standard and comparison tone bursts. We also assumed that the subject’s sensory measurements of the comparison and standard tone bursts followed Gaussian distributions, each with the same standard deviation, $\sigma_v$, that we defined as the frequency-discrimination threshold of that standard tone frequency $v_{stand}$ (Green and Swets, 1966; Macmillan et al., 1977; Creelman and Macmillan, 1979). Consequently, $\sigma_v$ was calculated directly from the $\sigma$ derived from the cumulative Gaussian fit: $\sigma_v = \sqrt{\sigma^2 / 2}$. We then computed each subject’s frequency-discrimination threshold as the average of the values measured at each of the eight standard tone frequencies (in $\log_{10}$ units). We used this average value for the predictions of our Bayesian model (see Bayesian model).

**Categorization task and analysis**

Each subject then participated in a two-alternative, forced-choice categorization task. The subject reported whether the frequency of a tone burst was a member of one of two different frequency categories (“A” or “B”).
The frequency range between 550–5550 Hz was divided into two equal (in \( \log_{10} \) units), but overlapping, piecewise-uniform category distributions (Fig. 4.1a). Category “A” contained frequency values between 500 to 2488 Hz. Category “B” contained frequency values between 1115 to 5550 Hz. These two categories were designed so that category “A” comprised the lower two-thirds of the frequency range, whereas category “B” comprised the upper two-thirds of the frequency range (again in \( \log_{10} \) units). As a consequence of this design, one part of each category’s distribution was exclusive to that category (i.e., the extreme thirds of the entire frequency range), whereas the other part was shared with the other category (i.e., the middle third of the range).

Our critical experimental manipulation was to vary the category prior probabilities, \( P(C) \), where \( C \) was either category “A” or category “B”. We varied the prior probabilities, on a block-by-block basis, by appropriately selecting the proportion of trials originating from a particular category. We tested the influence of three different category prior probabilities (Fig. 4.1b). In two of the manipulations, it was more likely that the frequency of a tone burst originated from one category than the other. In the third manipulation, it was equally likely that the frequency of a tone burst originated from either category.

Before the first session, the category prior probabilities were explained to each subject. A trial began with a brief 1500-ms countdown, followed by a visual ‘GO’ cue indicating the imminent presentation of a tone burst. After tone-burst offset, the subject had 1000 ms to report a choice. Subjects received visual feedback on every trial: a green circle for correct responses, a red circle for incorrect responses, and a yellow circle for no response within the allotted 1000-ms response window. In separate blocks of trials, the prior probability for category “A” was one of three values: \( P(C=“A”) = 0.25, 0.5, \) or 0.75. On a trial-by-trial basis, we randomly selected the category according to its prior probability. Once a category was selected, we randomly selected a frequency from that category. As noted above, because the category distributions were piecewise uniform, any stimulus within the category was equally likely: \( P(\nu|C) = k \) for all frequencies \( \nu \) within the category.
distribution \((C=“A” \text{ or } C=“B”)\) and \(P (ν|C) = 0\) outside of the distribution. The value of \(k\), where \(k>0\), is defined by the width of the category distributions.

![Figure 4.1: Schematic diagram of the categorical priors employed in the categorization task. (a) The category distributions over tone-burst frequency are piecewise uniform, such that all frequencies for a particular category are equally likely. (b) Three category prior probabilities were employed in separate blocks of trials by varying the proportion of trials that presented a tone belonging to each category. Here, \(P (C)\) represents the category prior probability, where \(C = “A” \text{ or } C = “B”\).](image)

Each subject participated in 3-5 sessions of the categorization task; each session included one block of each of the three category prior probabilities. In total, each subject completed between 600-1000 trials for each category prior probability.

For each subject, we computed the psychometric function \(P (\hat{C} = “A”|ν)\) (where \(\hat{C}\) represents the subject’s category choice) for each of the three category prior probabilities across all sessions. Tone frequencies were binned into nine equidistant bins that spanned the entire frequency range: three frequency bins in each of the two unambiguous frequency regions and three bins in the ambiguous frequency region. We fit each psychometric function with a cumulative Gaussian using a maximum-likelihood procedure and identified the frequency at which a subject was equally likely to choose \(\hat{C} = “A” \text{ or } \hat{C} = “B”\): that is, the point of subjective equality (PSE). We
also fit cumulative Gaussians to each subject’s categorization performance separately for each session to test for any potential learning effects throughout the course of the experiment.

**Bayesian model**

We developed a Bayesian model that tested three key aspects of each subject’s categorization behavior. First, we tested whether subjects used the category-prior information for their categorical decisions. Second, we tested the degree to which subjects were able to learn category distributions. Finally, we tested the degree to which subjects employed an optimal decision strategy given the characteristics of the categorization experiment.

![Figure 4.2: Graph of the Bayesian model.](image)

(a) The category identity $C$ of the frequency of a tone burst (top level) constrains the values of the tone frequency $\nu$ (middle level). The auditory sensory signal $m$ represents a noisy measurement of the true tone frequency $\nu$. The black arrows define the generative conditional probability densities $P(\nu|C)$ and $P(m|\nu)$, respectively. The task of the observer is to infer the category membership of the tone’s frequency from this noisy sensory measurement $m$ (red line from bottom to top level). (b) The category identity is modeled probabilistically using three $P(C = "A")$ conditions in the categorization task (top panel). Given a particular category, the probability of a certain tone frequency is governed by the respective conditional distribution for frequency $P(\nu|C)$ (middle panel). The sensory process of the Bayesian observer is modeled as a Gaussian process centered at the true stimulus frequency (bottom level). The width $\sigma_\nu$ reflects the degree of uncertainty in the sensory process due to noise and determines an observer’s ability to discriminate tones of different frequencies. Thus, we constrained this width with data from an additional discrimination experiment.

Categorization can be considered an inference process over the generative graphical model shown in Figure 4.2a. The true category $C$ of a stimulus is governed probabilistically
according to the prior probability $P(C)$ (Fig. 4.2b, top panel). The category distribution, $P(\nu | C)$, indicates the probability that a stimulus from a category $C$ has a certain tone frequency $\nu$. We assumed that each tone with frequency $\nu$ generated a sensory signal $m$ according to the probability density $P(m | \nu)$, which characterized the sensory uncertainty and noise in the auditory pathway. We assumed $P(m | \nu)$ to be Gaussian with a mean centered on the true tone frequency $\nu$ and a standard deviation $\sigma_\nu$ that reflected the level of sensory uncertainty (Fig. 4.2b, bottom panel). We measured $\sigma_\nu$ for each subject as his or her frequency-discrimination threshold (see **Discrimination task and analysis**).

We assumed that subjects performed Bayesian inference over this generative model when solving the categorization task: given the sensory evidence $m$, subjects computed the posterior probability $P(C | m) = \frac{P(m | C) P(C)}{P(m)}$. In this equation, $P(m | C)$ is the likelihood that the measured frequency belonged to a particular category $C$ = “A” or $C$ = “B”. The likelihood $P(m | C)$ was calculated by marginalizing over the tone frequency as $\int P(m | \nu) P(\nu | C) d\nu$. We assumed that subjects either (1) learned the experiment’s stimulus distributions (“objective priors”; Fig. 4.2b, middle-left) or (2) only learned an approximation of these distributions (“subjective priors”). For the latter case, we parameterized $P(\nu | C)$ using two piecewise-uniform distributions, each convolved with a Gaussian (Fig. 4.2b, middle-right). The subjective category distributions can be thought of as noisy estimates of the objective distributions. Each subjective distribution had its own mean ($\mu_\alpha$ and $\mu_\beta$) but had the same distribution width ($w$) and the same Gaussian standard deviation ($\sigma_C$). Finally, similar to the category distributions, the values of the category prior probability $P(C)$ were assumed either to be (1) the experimental prior probabilities (objective priors) or (2) the free parameters $\pi_{25}$, $\pi_{50}$, and $\pi_{75}$, representing each category prior probability (subjective priors).

Based upon the posterior $P(C | m)$, we tested whether subjects employed an optimal decision strategy to make a category choice (either $\hat{C} = "A"$ or $\hat{C} = "B"$). This strategy is a maximum a posteriori (MAP) strategy, in which subjects chose the most probable category given $m$. In other
words: $P(\hat{C}|m) = \begin{cases} 1 & \text{for } P > 0.5 \\ 0 & \text{otherwise} \end{cases}$. Thus, the subjects chose $\hat{C} = "A"$ if $P(C = A|m) > P(C = "B"|m)$, and chose $\hat{C} = "B"$ otherwise.

We also tested whether subjects’ decisions reflected probability matching (MATCH) as a general index of sub-optimal categorization behavior (Gaissmaier and Schooler, 2008; Koehler and James, 2009; Otto et al., 2011). Probability matching is equivalent to a decision strategy that results in subjects choosing a category probabilistically according to the posterior probability $P(C|m)$. In other words: $P(\hat{C}|m) = P(C|m)$.

Finally, to directly compare and fit the model’s predictions to each subject’s behavioral data, we computed the psychometric function as a function of the true frequency $\nu$ as $P(\hat{C}|\nu) = \int m P(\hat{C}|m)P(m|\nu)dm$.

**Model predictions and fits**

Assuming objective priors, we used the Bayesian model to quantitatively predict each subject’s categorization performance. We assumed the likelihood function $P(m|\nu)$ was a Gaussian distribution with a standard deviation $\sigma_{\nu}$, which was measured and fixed separately for each subject ($\sigma_{\nu,\text{mean}}$; see Discrimination task and analysis). Under these assumptions, the model has no free parameters. Therefore, we could predict each subject’s psychometric function for each category prior probability and for both optimal (MAP) and sub-optimal (MATCH) categorization. We calculated the quality of the MAP and MATCH predictions by computing their respective log-likelihood values across all $P(C="A")$ conditions. We rescaled these log-likelihood values relative to the predictions of two reference models: (1) an empirical model, which represents how well the observed data explains itself (i.e., a binomial model that employs the empirical choice probabilities), and (2) a random-guessing model (Stocker and Simoncelli, 2006).

Assuming that subjects only learned noisy estimates of the categorical priors (i.e., subjective priors), we also computed maximum-likelihood fits of the model for both MAP and MATCH behavior
to each subject’s categorization performance. The sensory uncertainty $\sigma_y$ was again fixed for each subject based on the results of the discrimination experiment. Thus, the model fit with the subjective priors had seven free parameters, namely $\mu_A$, $\mu_B$, $\omega$, $\sigma_C$, $\pi_{25}$, $\pi_{50}$, and $\pi_{75}$ (see Fig. 4.2b and previous section). We tested the goodness of fits by again comparing the normalized total log likelihoods for both MAP and MATCH.

Finally, to assess the full potential of either type of decision behavior to explain each subject’s categorization performance, we computed maximum-likelihood fits of the model using subjective priors, this time including $\sigma_y$ as an additional free parameter (for a total of eight free parameters). Once again, we tested the goodness of fits by comparing the normalized total log likelihoods.

RESULTS

Individual subject’s frequency-discrimination thresholds

We measured each subject’s frequency-discrimination threshold to determine individual sensory uncertainty. The frequency-discrimination experiment required subjects to indicate the interval that contained the higher-frequency tone burst.

For each subject, we calculated discrimination thresholds $\sigma_y$ for each standard frequency, which is summarized in Figure 4.3a. As expected (Fechner, 1966; Moore, 1973), we found that the thresholds were approximately constant across the tested frequency range. Consequently, for each subject, we computed the mean of the thresholds ($\sigma_{y,mean}$) across the eight standard frequencies (Fig. 4.3b). We used $\sigma_{y,mean}$ as the measure of each subject’s sensory uncertainty in our Bayesian model.
Human subjects can quickly learn category priors

Because the subjects were initially unaware of the categorical priors, subjects had to learn both the category distributions and the category prior probabilities to make informed category decisions. To test whether subjects learned this information, we first compared each subject’s psychometric functions (i.e., $P(C = "A"|\nu)$) across the three different values of the category prior probability $P(C="A"). We fit these psychometric functions with a cumulative Gaussian and extracted
the point of subjective equality (PSE) for each curve. The psychometric functions and Gaussian fits for an example subject (S3) are depicted in Figure 4.4a. Two main points can be taken from this figure. First, as the tone frequency increased, the probability that the subject chose $\hat{C} = "A"$ decreased. Second, as $P(C="A")$ increased, the psychometric functions shifted toward higher tone frequencies. However, the slopes of the psychometric functions remained consistent across category prior probability. These effects were comparable across individual subjects, with all but subject S2 exhibiting clear effects of the different category prior probabilities. These findings are summarized in Figure 4.4b and 4.4c.

These effects of the different category prior probabilities were evident as early as the first session. Generally, additional experience with the categorical priors had little differential effect on PSE and slope (Fig. 4.5). Thus, for subsequent analyses we grouped each subject’s data across sessions.
Figure 4.4: Effects of category priors on psychometric data for individual subjects. (a) Psychometric functions depicting the probability of choosing $C = "A"$, given the true tone frequency, for an example subject. Data points denote observed performance calculated by binning stimulus frequencies into nine equidistant bins. Lines depict cumulative Gaussian fits to raw data. Shading of lines and data points denote $P(C = A)$ condition. Error bars and shaded regions represent bootstrapped 95% CIs. (b) Medians and bootstrapped 95% CIs of the PSE of the fitted psychometric functions for each prior probability and subject. (c) Medians and bootstrapped 95% CIs of the $\sigma$ values of the fitted psychometric functions for each prior probability and subject. The $\sigma$ values are plotted in log₁₀ units. For (b) and (c), shading of the data points denotes the different $P(C = A)$ conditions.
Figure 4.5: Effects of learning. The extracted PSEs (a) and slopes (b) for each subject as a function of session. For both sets of plots, the data points represent the median and 95% CIs based on bootstrapped behavioral data. Shading denotes the different prior probabilities.
Figure 4.6: Predictions of the Bayesian model with different categorization behaviors. (a) Predicted psychometric functions for the model with objective priors during each of the three prior-probability conditions. The predictions assuming probability-matching (MATCH) behavior are on the left, whereas those of the MAP decision strategy are on the right. (b) Predicted psychometric functions for the model with subjective priors. Example predictions are plotted for three selected values of $\sigma_C$ for MATCH (left) and MAP (right). Line colors distinguish MAP versus MATCH and color shade denotes the three prior probabilities. For all model predictions, $\sigma_U$ was fixed to the mean discrimination threshold across all subjects.

Under the subjective-priors assumption, the predicted characteristics of the psychometric functions change distinctly for MAP and MATCH (Fig. 4.6b). With MATCH, the psychometric functions become smoother overall with increasing values of $\sigma_C$ (Fig. 4.6b, left column). However, the vertical shifts with increasing $P(C=\text{"A"})$ are still evident. The predictions for the MAP decision
strategy are similar to those under the objective-priors assumption (compare Figs. 4.6a and 4.6b, right column). Contrary to what is seen in the predictions for MATCH behavior, here $\sigma_c$ does not affect the slopes but, instead, affects the relative lateral shifts of the psychometric functions.

Figure 4.7: Model comparisons using objective priors and individually measured sensory noise $\sigma_w$. Rows distinguish the responses from and predictions for each subject. Columns distinguish the three prior probabilities. For all plots, the data points represent mean performance and bootstrapped 95% CIs in the categorization task. Line colors distinguish MAP versus MATCH and color shade denotes the three prior probabilities.
Data versus model predictions for objective priors

We compared the predictions of the Bayesian observer with each subject's behavior assuming the objective priors (see METHODS). In general, the model predictions for both types of decision behavior did not accurately reflect subjects’ behavior (Fig. 4.7). MATCH behavior predicted step-like psychometric functions (see Fig. 4.6) that were reflected only in some subjects' performance (e.g. S4). The predictions of the model with the MAP decision strategy were even less accurate: this decision strategy predicted slopes of the psychometric functions that were substantially and consistently steeper than those observed in each subject.

We quantified the quality of the two model predictions by calculating the total likelihood of the models given each subject's behavior. MATCH was significantly more predictive of each subject's performance, as exemplified by the likelihoods for each type of decision behavior across subjects (Fig. 4.8). In fact, the MAP strategy was significantly worse than a random guess for all subjects, whereas MATCH was better than random guessing for half of the subjects (i.e., S1, S4, and S5).

![Figure 4.8: Normalized likelihoods for the Bayesian model predictions.](image)

Likelihoods are normalized between that of a random-guessing model and empirical performance, defined as how likely the measured performance explains itself (see METHODS). Color denotes MAP versus MATCH.
Figure 4.9: Model comparisons using subjective prior distributions with observed individual responses. The format of the data is the same as that in Fig. 4.7. For all plots, shaded regions denote bootstrapped 95% CIs for subjective-prior model fits.

Data versus model fits with subjective priors

Because the objective category distributions did not fully predict the subjects' performances, we used subjective categorical priors and fit the Bayesian model (see Fig. 4.2 and METHODS). However, as before, we fixed $\sigma_y$ to reflect each subject's measured frequency-discrimination threshold.
Fits assuming MATCH behavior almost perfectly accounted for the data, with an accuracy that approached empirical performance (Figs. 4.9 and 4.10). However, the fits under the MAP strategy were still poor: the MAP strategy failed to account for the slopes of the psychometric functions (Fig. 4.9). Except for subject S1, the MAP strategy yielded fits that were significantly worse than random guessing. In fact, the MAP-strategy fits to the data did not provide any better account of the data than its predictions based on the objective priors (compare Figs. 4.8 and 4.10).

Figure 4.10: Normalized likelihoods for the Bayesian-model fits assuming subjective priors. The format of the data is the same as that in Fig 4.8.
Finally, we were interested in reconstructing the subjective category distributions for the subjects and comparing them to the objective distributions; because the MAP decision strategy provided a poor description of subjects’ performances, we focused only on the fits assuming MATCH behavior.

The reconstructed category distributions tended to more closely resemble Gaussian distributions rather than boxes (Fig. 4.11). Both the modeled category means and category widths either were close to or overlapping with the actual means and widths of the objective distributions (Fig. 4.12a-c). However, the category edges were much less defined as compared to the edges of the objective distributions, exemplified by large $\sigma_c$ values (Fig. 4.12d). Overall, the fitted category prior probabilities $\pi_{25}$, $\pi_{50}$, and $\pi_{75}$ for individual subjects were remarkably similar to the actual values 0.25, 0.5, and 0.75, respectively (Fig. 4.12e-g).
Figure 4.12: Fitted model parameters for MATCH behavior of the model with subjective priors. (a-d) Boxplots depicting the range of (a) the fitted means for the category-“A” distribution; (b) the fitted means for the category-“B” distribution; (c) the fitted widths that were shared between both category distributions; and (d) the widths of the fitted Gaussian functions that were convolved with the fitted uniform distributions. For plots (a-d), the thin dashed lines denote depict the values of $\mu_A$ (a), $\mu_B$ (b), $\omega$ (c), and $\sigma_C$ (d) that reflect the objective priors. (e-g) Boxplots depicting ranges of the fitted prior probability parameters $\pi_{25}$, $\pi_{50}$, and $\pi_{75}$. Thin dashed lines denote experimental prior probability values. For all plots, the stars denote values of parameters fit to the measured data, whereas the boxplots denote the median, 50%, and 95% CIs of the parameter values estimated from bootstrapped empirical responses. Note that subject S2’s categorization performance was not influenced by the category prior probabilities.
Figure 4.13: Likelihood comparisons for model fits. (a) Normalized log-likelihoods (see Fig. 4.8) for MAP and MATCH. Data points denote median and bootstrapped 95% CIs. Dashed line depicts the unity line. (b) Boxplots depicting the range of the fitted sensory uncertainties ($\sigma_U$) for MAP (red) and MATCH (blue). Stars denote fitted values to the measured data. Boxplots denote the median, 50%, and 95% CIs of the bootstrapped data. Black points denote measured discrimination thresholds for each subject and their 95% CIs.

**Analysis of categorization behavior with subjective priors and all free parameters**

The previous model analyses revealed that probability matching (MATCH) is much better than the optimal (MAP) strategy in both predicting each subject’s categorization behavior as well as explaining behavior after fitting the model with subjective priors. However, this comparison assumes that we have accurately measured each subject’s sensory uncertainty. It is possible that, with additional sources of sensory uncertainty (e.g., memory noise (Harris, 1952)), the MAP
strategy could be equally as descriptive as MATCH behavior. Indeed, under certain noise conditions, MAP and MATCH are mathematically equivalent (Ashby and Maddox, 1993). To address this possibility, we performed an additional analysis in which all of the parameters were fit, including $\sigma_v$ (for a total of eight free parameters).

When we included $\sigma_v$ as a free parameter, both strategies accurately reflected individual subject’s categorization behavior (fits not shown). However, we found that, without exception, MATCH behavior was still a better explanation of each subject’s performance (Fig. 4.13a). Moreover, in order for the MAP strategy to achieve this improvement in explanatory power, the sensory noise $\sigma_v$ had to be $10–100$ times larger than the measured values for each subject. In comparison, the fitted levels of $\sigma_v$ obtained from the MATCH fits were quite close to the individually measured discrimination thresholds for each subject (Fig. 4.13b).

**Effects of noise on the categorical priors**

Up to now, the model formulations assumed that subjects’ estimates of the categorical priors were constant. However, this may not be true. Thus, we were interested in determining how trial-by-trial noise on the categorical priors may affect categorization performance. In particular, we wanted to test whether this additional noise could cause performance under an optimal decision strategy (MAP) to appear sub-optimal (MATCH).

We conducted a series of simulations in which we added noise to both the means of the category distributions and the prior probabilities (Fig. 4.14a). Increasing category-distribution noise ($\sigma_{cd}$) led to decreases in the slope of the psychometric function (Fig. 4.14b). Note, even though the net effect of this noise is similar to having constant Gaussian-shaped distributions (Fig. 4.14b, inset), the predicted categorization performance is different from the MAP predictions with constant Gaussian-shaped distributions (see Fig. 4.6). In the latter case, there is no effect on the slopes of the psychometric function.
Figure 4.14: Simulations of behavior under the assumption of additional sources of categorical-prior noise. (a) Illustration of the two types of added noise: noise in the means of the category distributions ($\sigma_{CD}$, top) and noise in the category prior probabilities ($\sigma_i$, bottom). For the simulations, we computed the net effect on the psychometric function from 600 iterations of varying either the category means (b) or the category prior probabilities (c) assuming one of eight different levels of Gaussian noise. For (b) and (c), we also note the net effect on the corresponding estimates of the category distributions and category prior probabilities, respectively. (b) Net effects of noise in the category means ($\sigma_{CD}$) on the psychometric function for $P(C=\text{"A"}) = 0.25$. Colors denote the level of added noise. The effects were similar for each prior probability. (c) Effects of prior-probability noise ($\sigma_i$) on the psychometric function. Panels depict effects for $P(C=\text{"A"}) = 0.25$, 0.5, and 0.75, respectively. Insets for each panel depict mean and 95% CIs for $P(C=\text{"A"})$. Colors denote level of $\sigma_i$ noise. Note, because probabilities range from 0 to 1, samples were fixed to remain within the range 0–1.

Increasing prior-probability noise ($\sigma_i$) exhibited qualitatively different effects on performance as a function of $P(C=\text{"A"})$ (Fig. 4.14c). First, under asymmetric prior-probability conditions (i.e., $P(C=\text{"A"}) = 0.25$ or 0.75), sufficiently small levels of $\sigma_i$ (e.g., below ~0.08) did not substantially influence the psychometric function (Fig. 4.14c, left and right panels). However, larger levels of $\sigma_i$ caused the function to exhibit plateaus. Moreover, depending on the level of $\sigma_i$, we could observe over-, under-, or true probability matching; compare the bright and dark red traces in the left and right panels of Figure 4.14c. Interestingly, when the prior probabilities were symmetric (i.e., $P(C=\text{"A"}) = 0.5$), any level of $\sigma_i$ led to psychometric functions with a characteristic plateau.
One potential interpretation of this noise is that subjects’ categorical priors are non-stationary. Specifically, we hypothesized that subjects estimated the categorical priors only over recent trial history. To investigate this hypothesis, we computed running estimates of $P(C=\text{“A”})$ over different bin lengths of consecutive trials and compared the variability in these estimates with the levels of $\sigma_\pi$ that yielded step-like psychometric functions. We found that the variability in $P(C=\text{“A”})$ over relatively short bin lengths (i.e., generally <16 trials) was generally consistent with these $\sigma_\pi$ levels (Fig. 4.15).

Figure 4.15: Variability in category prior probabilities computed as running estimates of $P(C=\text{“A”})$ over different lengths of stimulus history (i.e., number of trials). Panels depict mean ± 1 SD of running averages of the stimulus history in the experiment for $P(C=\text{“A”}) = 0.25, 0.5,$ and $0.75$, respectively. Shading of data points denote different bin lengths. For comparison, the true $P(C=\text{“A”})$ estimates ± 1 SD of the largest $\sigma_\pi$ noise levels from Fig. 4.14 are depicted with solid and dashed lines. Colors of lines are the same as in Fig. 4.14c.
DISCUSSION

We found that subjects learned the categorization task to varying degrees. All but one subject could use the category-prior information to solve the task. Subjects learned general characteristics of the category distributions (i.e., high versus low frequencies) and the category prior probabilities as early as the first session. This is consistent with previous work showing that the largest effects of category learning occur early in training and then are fine-tuned with further experience (Edgell and Morrissey, 1987; Kruschke and Johansen, 1999). Our finding that subjects learned the category prior probabilities is consistent with previous visual categorization tasks (Lee and Janke, 1965; Healy and Kubovy, 1978; Estes et al., 1989; Bohil and Maddox, 2001; Hansen et al., 2011; 2012a). However, the systematic evaluation of prior probabilities and category learning in this study is novel for audition.

One goal of this study was to test whether subjects employed an optimal decision strategy to perform auditory categorization under categorical ambiguity. In order to do this, we developed a single generative Bayesian model that allowed us to both predict and fit each subject’s psychometric curve for all tested conditions under instances of either optimal or sub-optimal categorization behavior. A critical component of this approach was that we separately estimated each subject’s perceptual noise by measuring frequency-discrimination thresholds.

One finding of our model predictions was that subjects’ performances were not accurately predicted assuming the objective priors (i.e., box-shaped distributions). This suggests that subjects were limited in their ability to learn the objective priors. Indeed, our model fits were consistent with the hypothesis that subjects learned smooth approximations of the box-shaped distributions. This finding may not be surprising: previous work has demonstrated that subjects often assume approximate versions of experimental distributions when learning new behavioral tasks (Fried and Holyoak, 1984; Maddox, 2002; Berniker et al., 2010; Acerbi et al., 2012). It is possible that the large degree of uniform overlap between the categories contributed to subjects’ difficulties in estimating the category distributions. However, other evidence suggests that subjects can, to an extent, learn
category distributions that are non-Gaussian (Neumann, 1977; Jazayeri and Shadlen, 2010; Acerbi et al., 2012). Therefore, with extensive training, subjects might have been able to learn the objective priors.

Another important finding was that subjects’ performances were more consistent with probability matching. This was the case after both predicting and fitting performance with our Bayesian model. Because this type of behavior reflects sub-optimal categorization, we conducted further analyses to investigate whether subjects actually implemented an optimal decision strategy but performed sub-optimally due to additional uncertainties (Ashby and Maddox, 1993; Maddox and Ashby, 1993; Green et al., 2010).

Additional memory noise was unlikely to account for this possibility for two reasons. First, when sensory noise was a free parameter and could account for additional memory noise, probability matching still outperformed the optimal decision strategy. Second, the fitted values of the sensory noise for the optimal strategy were 10–100 times larger than our measured estimates (Fig. 4.13). This difference between the measured and fitted values seems unreasonable given previous work on the effects of memory noise on frequency discrimination (Harris, 1952).

We also simulated the effects of additional noise on the category distributions and prior probabilities. The results of the simulations suggested that a combination of category-distribution and prior-probability noise could lead to psychometric functions that mimic probability-matching behavior (i.e., shallow psychometric functions with a plateau), even though the decision strategy was optimal (see Fig. 4.14).

Categorical-prior noise could reflect true uncertainty or subjects’ tendencies to search for patterns in sequences of random events (Ayton et al., 1989; Wolford et al., 2004; Gaissermaier and Schooler, 2008; Koehler and James, 2009). One interpretation is that our subjects assumed that the categorical priors changed over time (i.e., they were non-stationary). Under this assumption, our analyses suggested that subjects’ estimates of the categorical priors were reflections of the short-term stimulus history (see Fig. 4.15). Future work is necessary to determine more
quantitatively whether subjects whose performance is most sensitive to the local trial history are more likely to exhibit psychometric functions that mimic probability-matching behavior and how this effect changes after extensive training.

Together, our results suggest that the prevalence of probability matching in perceptual tasks might reflect model assumptions of stationarity that are not correct (Thomas and Legge, 1970; Healy and Kubovy, 1981; Vulkan, 2000; Wozny et al., 2010; Summerfield et al., 2011). In other words, the interpretation of subjects’ categorical behavior should not focus on sub-optimal versus optimal decision strategies but, rather, should focus on the degree to which subjects assume the environment is stationary and which factors can impact these assumptions. For example, changes in cost-reward structures may not change subjects’ decision strategy, but may influence their view of environmental stationarity (Healy and Kubovy, 1978; 1981; Bohil and Maddox, 2001; Wozny et al., 2010).
CHAPTER 5

5. GENERAL DISCUSSION

In the following chapter, I describe the strengths and weaknesses of the current work, along with avenues for future research. I focus first on Chapters 2 and 3, which describe the oscillatory correlates of regularity representation and deviant detection, and then segue into the results of Chapter 4 on the psychophysical nature of auditory categorization under categorical uncertainty.

Summary of the oscillatory correlates of regularity representation

Chapters 2 and 3 tested the contribution of neural oscillations to representing the spectrotemporal regularities in an auditory stimulus. Specifically, we aimed to identify stimulus- and choice-related modulations in these oscillations. We found that spectrotemporal regularities induced a complex set of modulations in oscillatory power and phase across a wide expanse of cortex. Modulations in phase alignment occurred at neural frequencies that were generally directly correlated with the timescales of the regularities in the stimuli. In contrast, power modulations were generally uncorrelated with the timescales of the regularities. Further, modulations in phase
alignment exhibited reliable correlations with the degree of spectrotemporal regularity, whereas modulations in power, on average, were uncorrelated.

With respect to deviant detection, we found that both power and phase were modulated by both stimulus and report characteristics, with power modulations generally providing more information with respect to differences in stimulus or report. However, we were unable to identify modulations in either power or phase between detected and undetected deviants that were specific to each deviant type. This could be because either modulations in power or phase are truly not dependent on the type of deviant detected, or because we lacked sufficient statistical power to distinguish these modulations due to small sample sizes in both the number of subjects and the number of trials each subject performed for each deviant type. More subjects and deviant trials will ultimately be required to answer this question. Considering our analyses that collapsed across deviant types found reliable modulations with respect to behavioral report in temporal, inferior parietal, and ventrolateral prefrontal cortices, future work should focus on the contributions of these regions to the perception of spectrotemporal regularity.

**Strengths of the current investigation on regularity representation and perception**

The strengths of the current work on the nature of spectrotemporal-regularity representation arise from the design of both the experiments and the analyses employed to test the nature of the neural representation. These include: (1) designing a spectrotemporally dynamic stimulus with the ability to quantifying the degree of spectral regularity; (2) analyzing the wideband power and phase responses to produce a more complete picture of neural representation; (3) analyzing electrodes across cortex to more fully probe the extent to which cortical responses may contribute to auditory processing; and (4) designing a detection task that could potentially distinguish between the sensory versus putative perceptual representations of deviant detection.

First, by employing a metric to quantify regularity, we were able to test directly how neural activity reflects the degree of regularity in a stimulus. Previous work had begun to assess stimulus
regularity parametrically (Barascud et al., 2016; Teki et al., 2016), but these studies fell short of true parameterization because they only tested a subset of stimuli that differed in their inferred regularities. Consequently, it is difficult to say how the results of this previous work might generalize to stimuli with other kinds of spectrotemporal regularities. However, the advantage of our approach was our systematic quantification of regularity can be easily expanded to include more complex stimuli with multiple frequency components. For example, the Kolmogorov complexity can easily be computed for stimuli consisting of pseudo-random repetitions of an arbitrary number of tone-burst frequencies. Because of the generalizability of the Kolmogorov complexity, we believe that our approach provides a better means to further study regularity representation in the future.

Second, we employed an assumption-free approach to study the wideband power and phase correlates of spectrotemporal regularities. This approach allowed us to reveal that the nature of spectrotemporal regularity representation is more complex than previously described. Although previous studies had shown that a stimulus with a spectrotemporal regularity induces oscillatory stimulus-phase alignment in the frequency band correlating to the time scale of the regularity (Patel and Balaban, 2000; Lakatos et al., 2005; Bidet-Caulet et al., 2007; Besle et al., 2010; Henry and Obleser, 2012; Lakatos et al., 2013), we found that additional frequency bands, both harmonically related and unrelated, were also modulated by stimulus regularity. Thus, it is possible that previous studies may have missed reliable modulations in other neural frequency bands. We believe that a complete understanding of regularity representation (and of any neural representation in general) requires a comprehensive investigation of the underlying neural mechanisms.

Third, by analyzing neural activity across cortex, we were able to show the extent to which neural activity along and even beyond the classically-defined auditory pathways may contribute to auditory processing. Many studies have suggested that regions beyond the classical ventral auditory pathway play a role in auditory perception, including anterior regions of the temporal cortex near temporal poles (Belin et al., 2000; Poremba et al., 2004; Belin, 2006; Perrodin et al., 2011), the intraparietal lobule (Cusack, 2005; Teki et al., 2011; Barascud et al., 2016; Teki et al., 2016),
and regions in the medial temporal lobe (Barascud et al., 2016). Particularly, the intraparietal lobule is thought to play a role in the perceptual organization, such as binding stimulus features both within and across modalities (Kitada et al., 2003; Cusack, 2005; Miller and D'Esposito, 2005; Buelte et al., 2008; Xu and Chun, 2009; Yokoi and Komatsu, 2009; Werner and Noppeney, 2010; Teki et al., 2011). Additionally, the medial temporal lobe may be involved the integration of complex temporal patterns (Turk-Browne et al., 2009; Aly et al., 2013; Geiser et al., 2014; Schapiro et al., 2014; Garrido et al., 2015). Our findings that these same regions reflect spectrotemporal regularities and deviant detection are consistent with these results and suggest that future work should focus on elucidating how each of these reasons may contribute to regularity representation in auditory perception.

Finally, by attempting to design the deviant-detection task to be near chance performance, we had the opportunity to identify choice-related modulations in neural activity. Until recently, the majority of work relating to regularity representation and deviant detection has neglected to analyze neural correlates of behavioral choice (for reviews, see (Näätänen et al., 2007; Winkler, 2007; May and Tiitinen, 2010)). However, to ultimately understand how sensory information is transformed into perceptual representations, it is critical to determine the underlying cortical mechanisms that reflect stimulus versus perceptual characteristics. More recent work has begun to address these questions in the context of regularity representations (Henry and Obleser, 2012; Henry et al., 2014), showing that the phase of an entrained neural oscillation is predictive of a listener’s reports of gap detection. For example, in a visual-discrimination task, the phase of a low-frequency oscillation that is entrained to the stimulus-presentation rate correlated with error rates (van den Brink et al., 2014). These results are consistent with our finding that phase information reflects detected deviations in spectrotemporal regularity (e.g., Fig. 3.15), in support of a role for neural-phase encoding of both spectrotemporal regularities and deviant detection. However, we also found that power information was generally more indicative of reported deviants. Thus, future work should determine the frequency specificity of the power modulations that reflect the deviant-detection process to better
understand the underlying cortical mechanisms. For instance, whereas narrowband power modulations might reflect a true oscillatory rhythm, more wideband shifts of the power spectrum may reflect underlying neural spiking activity (Manning et al., 2009).

**Pitfalls of the current investigation on regularity representation and perception**

One major concern that affects all human ECoG work is related to the applicability of the findings to healthy individuals. However, several important factors lend support to the broader applicability of our findings. First, none of our subjects were diagnosed with any hearing impairments, suggesting that the neural responses we characterized are not influenced by auditory-perceptual deficits (although poor performance on the deviant detection task suggest that at least some subjects may have had underlying cognitive deficits). Second, only electrodes that were deemed free of epileptic activity by the clinical staff were included in the analyses, and subjects were generally not tested within 12 hours of a recorded seizure. Thus, it is unlikely that our results are confounded by long-lasting effects of epileptic activity or interictal events. Third, similar results regarding the representation of spectrotemporal regularity in oscillatory phase have been found in healthy individuals (Patel and Balaban, 2000; Henry and Obleser, 2012; Henry et al., 2014).

Another general issue was a lack of statistical power due to the small number of subjects. Our method for determining localization of significant effects tested the reliability of significant modulations across subjects. Given our whole-brain approach to studying the neural correlates of auditory perception, a fine-grained analysis of each brain region in the passive-listening task (Chapter 2) would have required more subjects to withstand multiple-comparisons corrections.

This lack of statistical power was particularly apparent in the deviant-detection task (Chapter 3). Although it was possible to use the data from all 17 subjects to localize reliable modulations to the standard sequence, only 14 of the subjects could perform the task (see Fig. 3.7), and only 9 of those 14 performed more than 2 sessions of the task. Further still, only 7 subjects performed significantly better than chance on standard trials, and only 5 of those performed more
than 1 session of the task. As a result, it was difficult to assess the reliability of each brain region across subjects. Further, we were limited in our ability to distinguish between the neural correlates of each type of deviant due to the low number of trials of each deviant type.

Finally, it is important to note that the process for maintaining near-chance performance in the deviant-detection task could have been better optimized on an individual-subject basis. To maximize testing time for each subject, we opted to simultaneously titrate task difficulty across sessions by modifying the level of frequency increases while subjects performed the task. As a result, it is possible that the classifier analyses that distinguished modulations as a function of behavioral report were at least partially reflecting stimulus differences based on the magnitude of acoustic-frequency change in the deviant stimuli, as hit rates generally correlated with the magnitude of the acoustic-frequency change for each deviant (see Fig. 3.8). We attempted to control for this confound in the wideband analyses by computing the average power and phase spectra separately for each level of frequency increase, but we were unable to identify the specific nature of the modulations identified in the classifier analyses. In any case, to convincingly disentangle the neural correlates of perceptual versus stimulus differences, we would likely need to redesign the task to first measure each subject’s discrimination thresholds using a staircase method (Levitt, 1971; Gifford et al., 2014). Then, subsequent testing would use ideal levels of frequency deviations for each deviant type that is suited for each subject. Additionally, removing feedback during testing might minimize any improvements in deviant-detection thresholds over multiple testing sessions (Campbell and Small, 1963), which would eliminate the need to change the task difficulty across sessions to maintain near-chance performance.

**Future directions on the study of regularity representation and perception**

The results and preliminary findings of the passive-listening and deviant-detection tasks generate a number of further questions regarding regularity representation and perception that should be considered. For instance, are the observed oscillatory effects true oscillatory components
in the ECoG signals or do they reflect temporally regular, evoked-type activity? How does oscillatory activity reflect more complex kinds of spectrotemporal regularities that can vary dynamically on a number of different timescales? Are neural oscillations causally related to spectrotemporal-regularity representation and perception? And what are the potential causal roles of regions beyond the cortical auditory pathways in spectrotemporal-regularity representation and perception?

Although we have interpreted our results that phase-based representations of spectrotemporal regularities reflect true oscillatory components, it is possible that they, instead, simply reflect repetitive evoked-like activity. Indeed, the oscillatory versus evoked nature of population-level neural activity remains an outstanding question in the literature, with evidence in support of both interpretations (Fell et al., 2004; Mäkinen et al., 2005; Klimesch et al., 2007; Schroeder and Lakatos, 2009; Lakatos et al., 2013). There is even debate as to the extent to which certain analysis techniques can distinguish between oscillations and evoked activity (Sauseng et al., 2007). Ultimately, further studies will be required to fully elucidate the differential contributions of each type of activity to spectrotemporal-regularity representation, which will likely require careful design of stimuli as well as special models and analytical techniques to distinguish between the two alternatives (Truccolo et al., 2002; Luzhou Xu et al., 2009).

Assuming our interpretation that regularities are reflected in neural oscillations, it will be critical to systematically explore how oscillatory phase is modulated by more complex stimuli with multiple degrees of regularity across multiple time scales. By designing stimuli that vary in their degree of spectrotemporal regularities across multiple time scales, we can determine how distinct neural oscillations interact when multiple regularities are detected and how these interactions affect the perceptual qualities of the stimuli. For example, music perception is characterized by the grouping of individual notes across multiple time scales that form the representation of the beat and meter (Gordon et al., 2011; Nozaradan et al., 2012; Grahn and Rowe, 2013). By designing stimuli that vary in their spectrotemporal regularity across multiple time scales, we can correlate
neural oscillatory activity with behavioral measures related to how the individual acoustic events in the stimuli are grouped perceptually.

Despite our findings that oscillatory phase reflects spectrotemporal regularities, we could not determine the causal versus correlative role that it plays in auditory perception. Consequently, future work is necessary to distinguish between these two possibilities. One potential way to do this would require the ability to manipulate neural oscillatory activity via electrical stimulation and determine the behavioral consequences of spectrotemporal-regularity detection. If by manipulating neural oscillatory activity we can affect a listener’s behavioral reports of detected spectrotemporal regularities, we could more confidently say that the oscillatory activity itself plays a causal role in regularity representation. If, on the other hand, modulating oscillatory activity does not affect a listener’s perception, it might suggest that oscillatory activity is simply epiphenomenal of the underlying causal neural mechanisms. Based on the current work and other related work (Lakatos et al., 2005; 2008; Schroeder and Lakatos, 2009; Schroeder et al., 2010; Lakatos et al., 2013; Henry et al., 2014), we hypothesize that neural oscillatory activity is causally related the perception of spectrotemporal regularities.

Finally, based upon the current results, it is clear that multiple regions across cortex exhibit activity that is reliably modulated by spectrotemporal regularity and behavioral reports of deviant detection. Future work should focus on region-specific analyses of each of the identified brain regions in order to more systematically assess the specific role(s) that each region may play in auditory perception. This includes determining the causality of neural activity in influencing perception as well as whether and how information is communicated among brain regions.

**Summary of the computational strategies in auditory categorization**

Chapter 4 explored the computational strategies that human listeners employ to categorize stimuli given categorical uncertainty. Specifically, we tested the subjects’ ability to learn category priors to inform categorical decisions. First, we found that whereas most subjects could learn the
experimental category-prior information, none of the subjects could learn the shape of the objective-(boxed-shaped-) prior distributions. Second, we found that each subjects’ performance more closely resembled probability matching rather than an optimal performance based on a maximum a posteriori decision strategy. Finally, we found that a combination of category-distribution and prior-probability noise could lead to psychometric functions that mimic probability-matching behavior, even if a subject’s decision strategy was optimal.

**Strengths of the current investigation of auditory categorization**

The auditory-categorization task was able to uniquely probe the computational strategies human listeners employ under categorical uncertainty because we first estimated each subject’s sensory noise (i.e., acoustic frequency-discrimination thresholds), which provided a strong constraint in the Bayesian model. This allowed us to probe the uncertainties related specifically to the categorical priors in order to determine how those properties affected categorization performance. By quantifying each subject’s sensory noise, we could determine that categorization performance was primarily determined by the subject’s ability to learn the categorical priors. Moreover, using the measured sensory noise as a constraint, we were able to model how a listener’s categorization performance could appear to mimic probability-matching even if the listener employed an optimal decision strategy for category choice.

Additionally, by analyzing the subjects’ behavior using a Bayesian model, we could interpret our understanding of auditory categorization within a generative framework that can be used to further test the neural underpinnings of auditory categorization. Much of human cognition seems to be a process of probabilistic inference (Kersten and Yuille, 2003; Kersten et al., 2004; Ma et al., 2006; Vilares and Körding, 2011; Ma, 2012), whereby the brain deals with inherent uncertainties in the sensory information it receives to generate perceptual representations and guide behavior. This has been demonstrated in a variety of perceptual and cognitive tasks (Vilares and Körding, 2011): for example, human subjects are capable of combining prior information with
sensory information in arm-reaching and pointing tasks in a manner that approaches ideal performance (Körding and Wolpert, 2004; Brouwer and Knill, 2009). Additionally, previous work has demonstrated how neural activity can encode probabilistic inference variables in a variety of perceptual and cognitive tasks (Gold and Shadlen, 2001; Ma et al., 2006; Gold and Shadlen, 2007; Wei and Stocker, 2012). For example, neurons in the lateral intraparietal area exhibit activity that correlates with the likelihood that the motion of random dots favors one direction over another (Gold and Shadlen, 2001). Thus, by reframing the problem of auditory categorization within a generative framework, our work can bridge the gap between our understanding of the extensive psychophysical literature on categorization and the underlying neural mechanisms responsible for it.

Pitfalls of the current investigation of auditory categorization

One of the primary limitations of this auditory-categorization study is its simplistic design: two arbitrary categories were assigned based upon a single stimulus dimension. As such, it is difficult to determine how the current findings would generalize to more realistic situations with multiple categories defined across a combination of stimulus dimensions. For example, a listener might categorize the emotional content of human speech (i.e., happy, sad, angry, etc.) based on a number of stimulus dimensions, including tone, sound level, speech production rate, and types of words used. It is important to note that this is a common concern in most categorization studies, which have focused on distinguishing between only two relevant categories (Maddox et al., 2004). Understanding how category learning is affected by the complexity of the task could provide insight into the location and nature of the neural underpinnings of auditory categorization. It is believed that there are at least two different kinds of category learning: explicit, rule-based learning, for which category boundaries can be easily described verbally; and implicit, information-integration learning that relies on the combination of information over multiple stimulus dimensions (Maddox and Ashby, 2004; Maddox et al., 2004; Ashby and Maddox, 2011; Reetzke et al., 2016). These two kinds of
learning are thought to be governed by different neural systems (Maddox et al., 2004; Ashby and O'Brien, 2005; Seger and Miller, 2010; Reetzke et al., 2016), and they are differentially affected by the number of categories and the nature of the rules governing category boundaries (Maddox et al., 2004). Therefore, whereas our categorization task may reflect rule-based category learning, the categorization of emotional content might reflect information-integration category learning.

Another limitation was that, with our current task design and the Bayesian model, we could not directly determine whether subjects’ performances reflected a truly sub-optimal decision strategy or whether they simply reflected uncertainty in the stationarity of the categorical priors. In order to distinguish between these alternatives, we would have to measure categorization performance over time to determine whether or not experienced listeners still exhibited sub-optimal performance. If we could determine that performance approaches optimality as a listener gains experience, it would suggest that our results simply reflected uncertainties in the categorical priors. Alternatively, if performance remains sub-optimal even after extensive training, then it might suggest that listeners were actually performing sub-optimal in the task.

**Future directions in the study of auditory categorization**

One obvious next step in the study and modeling of auditory categorization is a more detailed assessment of category learning. As stated above, this would allow us to potentially distinguish between the two alternatives regarding categorical uncertainties versus sub-optimal decision making. Additionally, it would allow us to determine the extent to which listeners could learn exact representations of the box-shaped category distributions versus approximations of them.

In addition to studying category learning, a critical next step would be to determine the underlying neural correlates of the categorization process. Specifically, the Bayesian framework suggests that specific pieces of information should be represented in neural activity to perform this task optimally. First, it would be important to determine where and how the category distributions
are represented, as well as the category prior probabilities. Second, we could determine where how the categorical priors are combined with the sensory information to form an estimate of the posterior probabilities for each category. Finally, we could determine where neural activity correlates with category decisions.

Conclusions

In conclusion, in the neurophysiological studies, we have shown how oscillatory phase alignment seems to be a general mechanism by which spectrotemporal regularities across multiple time scales are represented. Additionally, neural power and phase in multiple distinct brain regions reflect a listener’s reports of detected deviations in spectrotemporal regularity. These results suggest further avenues of study to determine the causal role of oscillatory activity and the contributions of multisensory cortical regions to auditory perception. In the auditory-categorization task, we found that listeners learned noisy estimates of the categorical priors and that their performance was more consistent with a sub-optimal decision strategy leading to probability matching. Future work should focus on distinguishing categorical uncertainty from sub-optimal decision strategies and determining the neural underpinnings of auditory categorization.
REFERENCES


Hansen KA, Hillenbrand SF, Ungerleider LG (2012b) Effects of prior knowledge on decisions made under perceptual vs. categorical uncertainty. Front Neurosci 6:1–10.


Maddox WT, Filoteo JV, Hejl KD, Ing AD (2004) Category Number Impacts Rule-Based but Not


Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of Object Representations in the


Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation


